Article

Accumulation Dynamics of Defective Genomes during Experimental Evolution of Two Betacoronaviruses

Julia Hillung [1,†] , **María J. Olmo-Uceda** [1,†] , **Juan C. Muñoz-Sánchez** [1] and **Santiago F. Elena** [1,2,*]

1   Instituto de Biología Integrativa de Sistemas (I2SysBio), CSIC-UV, Catedrático Agustín Escardino Benlloch 9, 46980 Paterna, Valencia, Spain; julia.hillung@csic.es (J.H.); mariajose.olmo@csic.es (M.J.O.-U.); jc.munoz@csic.es (J.C.M.-S.)
2   Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
*   Correspondence: santiago.elena@csic.es; Tel.: +34-963-544-779
†   These authors contributed equally to this work.

Abstract: Virus-encoded replicases often generate aberrant RNA genomes, known as defective viral genomes (DVGs). When co-infected with a helper virus providing necessary proteins, DVGs can multiply and spread. While DVGs depend on the helper virus for propagation, they can in some cases disrupt infectious virus replication, impact immune responses, and affect viral persistence or evolution. Understanding the dynamics of DVGs alongside standard viral genomes during infection remains unclear. To address this, we conducted a long-term experimental evolution of two betacoronaviruses, the human coronavirus OC43 (HCoV-OC43) and the murine hepatitis virus (MHV), in cell culture at both high and low multiplicities of infection (MOI). We then performed RNA-seq at regular time intervals, reconstructed DVGs, and analyzed their accumulation dynamics. Our findings indicate that DVGs evolved to exhibit greater diversity and abundance, with deletions and insertions being the most common types. Notably, some high MOI deletions showed very limited temporary existence, while others became prevalent over time. We observed differences in DVG abundance between high and low MOI conditions in HCoV-OC43 samples. The size distribution of HCoV-OC43 genomes with deletions differed between high and low MOI passages. In low MOI lineages, short and long DVGs were the most common, with an additional cluster in high MOI lineages which became more prevalent along evolutionary time. MHV also showed variations in DVG size distribution at different MOI conditions, though they were less pronounced compared to HCoV-OC43, suggesting a more random distribution of DVG sizes. We identified hotspot regions for deletions that evolved at a high MOI, primarily within cistrons encoding structural and accessory proteins. In conclusion, our study illustrates the widespread formation of DVGs during betacoronavirus evolution, influenced by MOI and cell- and virus-specific factors.

Keywords: defective genomes; evolutionary dynamics; experimental evolution; mutant swarm; virus evolution

1. Introduction

Owing to their high mutation rates, rapid replication, and large population sizes, RNA viruses quickly adapt to novel hosts and trigger epidemics by crossing species barriers [1–3]. However, many mutations during infection lead to aberrant genomes that cannot complete the infectious cycle. In vitro, these incomplete genomes, known as defective viral genomes (DVGs), accumulate in virus populations when the virus is passaged repeatedly at a high multiplicity of infection (MOI) [4–6]. Some DVGs can be encapsidated and transmitted. When co-infecting with a wild-type virus, in some cases, DVGs may interfere with virus replication [7–9]. In this case, the term defective interfering particle (DIP) is used to refer to encapsidated DVGs that negatively impact wild-type virus replication. This interference occurs on multiple levels, including resource competition, immune response activation, and sometimes, the survival of infected cells [5,8–10]. Although specifics of DIP generation

and regulation in RNA viruses are still not fully understood, their role in modulating viral infections is well-established [11,12].

Deletions are the most common type of DVGs among RNA viruses [13,14]. They are thought to form through homologous recombination due to observed homology in specific regions and RNA structures [15–17]. Copy-back (cb) and snap-back (sb) genomes are also prevalent DVG classes in negative-sense RNA viruses. These genomes have loops with complementary ends [18,19].

Coronaviruses have the largest known RNA virus genomes, ~30 kb in length [20]. They encode nonstructural proteins involved in viral RNA synthesis and interact with host cell functions. Replication generates a set of subgenomic mRNAs, all sharing common 5′ leader and a 3′ terminal sequences [21]. Specific regions, such as the leader sequence and a portion of the replicase gene, contribute to efficient replication and transmission. Additionally, *cis*-acting elements like the transcriptional regulatory sequences (TRS) are crucial for transcription [22]. In most described coronavirus DIPs, the combination of both internal and terminal genome sequences enables the formation of secondary and higher-order structures essential for their function as replication signals [23]. Deletion mapping of murine hepatitis virus (MHV) DIPs reveals the requirement of an internal and discontinuous sequence for replication [24]. The 3′ proximal coding regions, including the *N* cistron, can tolerate substantial alterations, suggesting they are not part of the 3′ *cis*-acting elements [25–28]. Efficient genome packaging into virions requires specific RNA signals, primarily located at the terminal sequences [29–33].

To investigate the dynamics of DVG accumulation in betacoronaviruses infecting highly susceptible cells, we conducted an evolution experiment by performing undiluted passages of (*i*) the human coronavirus OC43 (HCoV-OC43) in baby hamster kidney cells (BHK-21) and human large intestine carcinoma cells (HCT-8), and (*ii*) of MHV in murine liver cells (CCL-9.1). We anticipated that a high MOI would promote the accumulation of DVGs over passages due to frequent co-infection with wild-type viruses. As a control, we also carried out parallel evolution experiments at a low MOI. We replicated both MOI conditions and virus–cell type combinations three times. After experimental evolution, we employed high-throughput RNA-seq and bioinformatic tools to reconstruct DVG populations in the ancestral virus and in samples at four equidistant evolutionary passages. We did not specifically evaluate whether the accumulated DVGs acted as DIP, but focused on describing their temporal dynamics and potential mechanisms of generation.

2. Materials and Methods

2.1. Viruses and Cells

Human large intestine carcinoma cells (HCT-8; ATCC® CCL-244TM) and murine liver cells NCTC 1469 (CCL-9.1; ATCC® CCL-9.1TM) were obtained from the American Type Culture Collection (ATCC). Baby hamster kidney cells (BHK-21) were kindly provided by Dr. R. Sanjuán (I²SysBio, Valencia, Spain). BHK-21 cells and HCT-8 cells were cultured at 37 °C in DMEM (Gibco-Invitrogen, Grand Island, NY, USA) supplemented with 0.22% (*w/v*) sodium bicarbonate (Sigma Aldrich, Saint Louis, MO, USA), sodium pyruvate (Sigma Aldrich), 10% fetal bovine serum (FBS), 1× penicillin-streptomycin (Gibco-Invitrogen), 1× amphotericin B (Gibco-Invitrogen), and non-essential amino acids using standard laboratory procedures. CCL-9.1 cells were cultured at 37 °C in DMEM media supplemented with 0.22% (*w/v*) sodium bicarbonate (Sigma Aldrich), sodium pyruvate (Sigma Aldrich), 10% horse serum, 1× penicillin-streptomycin, 1× amphotericin B, and non-essential amino acids. All cell lines were routinely tested for mycoplasma contamination.

The viruses used in this study were HCoV-OC43 (ATCC® VR-1558) and MHV (ATCC® VR-766), both obtained from ATCC. The maintenance medium for HCoV-OC43 was DMEM supplemented with 0.22% (*w/v*) sodium bicarbonate (Sigma Aldrich), sodium pyruvate (Sigma Aldrich), 2% FBS, 1× penicillin-streptomycin (Gibco-Invitrogen), and 1× amphotericin B (Gibco-Invitrogen). Cell culture media of CCL-9.1 was also used as maintenance medium by infection with MHV.

2.2. Experimental Evolution

To generate the initial virus inoculum, BHK-21 and HCT-8 cells were plated one day before and grew to 70–80% confluency. Afterward, the cell layers were infected with the corresponding virus at a MOI of 0.01. HCoV-OC43 was grown for 4 d at 33 °C, and MHV was grown for 20 h at 37 °C. The supernatants from infected cells were then collected and used as the infectious stock. For successive passages, all cell types were grown in 6-well plates to 70–80% confluency and infected with 250 μL of undiluted viral inoculum from the previous passage. HCoV-OC43 was passaged 47 times in BHK-21 cells and 31 times in HCT-8 cells, while MHV was passaged 19 times in CCL-9.1 cells.

The MOI was not maintained the same throughout all passages; instead, at high MOI, we infected every passage without dilution to get the highest MOI possible. For low MOI samples, the inocula were diluted to avoid infecting each cell with more than one infectious particle. Table S1 shows the range of MOIs for each virus and cell type.

2.3. Viral Load Quantification

The quantity of infectious virus in the supernatant was determined after each passage. To determine the infectious viral particles for HCoV-OC43, plaque assays were performed in BHK-21. The procedure involved seeding the cells into 6-well plates and incubating them for 24 h until reaching 80% confluency. Serial dilutions of the virus were prepared in DMEM, and 250 μL of each dilution was used to infect the wells of the 6-well plate for 90 min at 33 °C and 5% CO_2 with swinging every 15 min. A negative control consisting of only DMEM media without virus was included for comparison. After the infection period, the cells were covered with 2 mL of media, which consisted of a 50:50 mix of 2× DMEM supplemented with 2% FBS and 2% agar. The agar–DMEM layer was overlaid with liquid DMEM supplemented with 1% FBS as it solidified. The plates were then incubated at 33 °C and 5% CO_2 for 5 d. After incubation, cells were fixed using 10% formaldehyde and plugs were removed. Monolayers were stained with 2% crystal violet in 10% formaldehyde, washed with tap water, and plaques counted to determine the plaque-forming units (PFU) per mL of inoculum.

The infectious viral particles for MHV were determined by $TCID_{50}$. Cells were seeded into 96-well plates and incubated for 24 h to reach 90% confluence. Viral transfer samples were diluted in complete DMEM. Once diluted, 125 μL of the dilutions for each sample were used to infect the wells of the 96-well plate. A negative control consisting of only complete DMEM was also included. The plates were then incubated at 37 °C for 2 d. After incubation, the wells were inspected under a light microscope for the presence of visual cytopathic effect, and the $TCID_{50}$/mL was calculated using the Reed and Muench method [34]. For comparison purposes with plaque assay results, $TCID_{50}$/mL values were converted into PFU/mL by multiplying by $\ln(0.5) \approx 0.693$.

2.4. Preparation of Viral RNA

Samples from four equidistant time series of each evolutionary lineage were collected to extract RNA using the NZY Viral RNA Isolation kit (NZYTech, Lisboa, Portugal) following the manufacturer's instructions. After extraction, the RNA was purified and concentrated using the RNA Clean&Concentrator-5 Kit (Zymo Research, Irvine, CA, USA).

2.5. High-Throughput RNA Sequencing

Libraries for high-throughput genome sequencing (HTS) were prepared by Novogene UK Company Ltd., Cambridge, UK (www.novogene.com; accessed on 17 April 2024) using at least 200 ng of total RNA or a larger amount if available. The ribosomal RNAs (rRNA) from both eukaryotes and prokaryotes were depleted from the total RNA samples. The remaining RNAs were fragmented into 250–300 bp fragments and reverse-transcribed into double-stranded cDNAs, followed by end repair, A tailing, and adapter ligation. After fragment size selection and PCR amplification, the metatranscriptome library was checked for quality, and sequencing was conducted on the Illumina NovaSeq6000 platform

(Illumina, San Diego, CA, USA). The raw data obtained from the samples ranged between 3 GB and 15 GB, providing ~64–124,981 × viral nucleotide mean depth, with a median of 13,362.52. The sequencing was not strand-specific.

2.6. Processing of HTS Reads and Identification of DVGs

Raw paired reads from Illumina metagenomic data from each sample were merged into a single interleaved file using the reformat.sh script from the BBMap/BBTool package (https://github.com/BioInfoTools/BBMap/blob/master/sh/reformat.sh (accessed on 17 April 2024)). The interleaved reads were then processed using the BBduk.sh tool to remove adapter sequences, clean, and trim the reads to obtain high-quality nucleotides at both ends (https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh (accessed on 17 April 2024)). Next, the remaining reads were mapped to the corresponding host genomes: the hamster genome for BHK-21 cell lineages, the human genome for HCT-8, and the mouse genome for CCL-9.1, using the BWA-MEM algorithm [35]. Reads that mapped with the host genomes were excluded from further analysis. The resulting reads were analyzed using the metasearch tool DVGfinder to identify and classify DVG events. Predicted DVG species sizes were calculated based on the breaking or start points (BP) and rejoining or end points (RP) by the program [36]. Reference genomes used were ATCC VR-1558 for HCoV-OC43 and GU593319.1 for MHV.

The raw sequencing data were deposited in GenBank (NCBI) under BioProject number PRJNA1021788.

2.7. Processing and Diversity Measures of DVG

Only the results from DVGfinder Filtered mode were used. In the case where both possible DVG directions were present in the original results, the sum of their abundance values was used, and BP and RP lost their original meaning. The canonic sgRNAs were identified and discarded from the rest of the analysis. A deletion-type DVG was identified as sgRNAs when its start coincides with the TRS-leader and its end with one of the canonical TRS-backs; both coordinates have a flexible margin of ten bases.

Richness was understood as the number of different DVGs relative to the total number of mapped viral reads of the group (sample or lineage, depending on the case) per 10^5 (reads per hundred thousand; RPHT). Abundance was used as a measure of the number of reads supporting the DVG, also expressed as RPHT.

2.8. Identification of Transmitted DVGs

To study the possible transmission of DVGs between passages, each DVG was labeled as persistent if the exact same event (DVG type with same BP and RP coordinates) was identified in a previous passage of the same lineage (including the stock).

2.9. Characterization of Deletion Type DVGs

All the DVGs with deletions bigger than three bases were analyzed from different perspectives: their length, their location in the genome, their distance to the nearest TRS, their nucleotide composition, and their secondary structure. The cistrons that were more affected by deletions were also identified. Smaller DVGs were discarded to minimize known Illumina error profiles [37].

2.9.1. Deletion Size Distribution

The distribution of the reconstructed size of deletion type DVG was calculated considering the abundance of each species in the sample. For a better visualization of the frequency coordinates, only deletions bigger than 50 bases were represented. The size of the deleted fragments was analyzed by a Lomb–Scargle periodogram method [38].

2.9.2. Distribution of Deletion Breaking (BP) and Rejoining (RP) Points in the Viral Genomes

For every cell type, a histogram of the start and end position of deletions was represented. The data was generated with the method hist from the base package graphics, setting the breaks argument to the quotient of the viral length between 50.

2.9.3. K-Mer Analysis

The *k*-mer composition of the sequences ten bases pre- and post- BP and RP of deletions was compared with the *k*-mer composition of the complete genome (*k* = 1 to 3) with a χ^2 test (chisq.test from base package stats, with rescale.p = T). The *k*-mer composition was analyzed with the oligonucleotideFrequency method from Biostrings v2.66.0 [39].

2.9.4. Minimum Free Energy (MFE) around Junctions

The MFE of the structures predicted with RNAfold v2.5.1 [40] for 50-base sliding windows were used to perform a Spearman correlation analysis with the frequencies for 50-base windows of being start or end coordinates of a deletion calculated previously. RNA2Drawer [41] was used for representation of the obtained structures.

2.9.5. Deletions per Base Calculation

To evaluate the viral cistrons more affected by deletions, the findOverlaps method from GenomicRanges v1.50.2 [42] was used in a homemade script in R version 4.3.2. The number of deletions was relativized to the cistron length.

2.10. Other Statistical Analyses

Except otherwise indicated, statistical analyses were done with SPSS version 28.0.1.0 (IBM, Armonk, NY, USA).

Viral load time series for individual lineages were fitted to the general ARIMA(*p*, *d*, *q*) model:

$$\left(1 - \sum_{i=1}^{p} \rho_i L^i\right)(1 - L)^d V_t = \left(1 + \sum_{j=1}^{q} \mu_j L^j\right)\varepsilon_t \tag{1}$$

where V_t represents the viral load at passage *t*; *L* is the lag operator such that $L^n V_t = V_{t-n}$; ρ_i are the parameters of the autoregressive part of the model with $p \geq 0$ being the order of the autoregressive model (number of time lags); μ_i are the parameters of the moving average part with $q \geq 0$ being the order of the moving-average model; $d \geq 0$ is the differencing order, representing the number of times the time series is differenced to achieve stationarity; and ε_t are the error terms assumed to be independent identically distributed N(0, σ^2). ARIMA fitting and model selection were done using the package forecast in R version 4.3.2 in RStudio version 2023.12.1+402. Estimated slope values were further analyzed by runs tests or pairwise Mann–Whitney tests.

Diversity and abundance of DVG classes were analyzed using the Scheirer–Ray–Hare (SRH) non-parametric two-way analysis of variance. In all tests done, the specific factors analyzed (i.e., virus species, cell type, or MOI) were orthogonal to the class of DVG (i.e., deletions, insertions, 3′, and 5′ cb).

Other specific statistical tests are presented in the text as needed.

2.11. Code Implementation

All the code developed for this work is accessible in https://github.com/MJmaolu/ AccumulationDynamicsDVGs/tree/main (accessed on 17 April 2024). Intensive computations were run on the HPC cluster Garnatxa at I2SysBio (CSIC-UV). The R scripts were run in Rstudio version 4.2.1 (23 June 2022).

3. Results

3.1. Experimental Evolution of Betacoronavirus DVGs Populations

Viral load was evaluated at each passage, and dilutions were adjusted to ensure that, on median, MOIs remained within two wide but disjoint intervals that can broadly be defined as higher and lower MOIs, respectively (Table S1). Notice that what is defined as high or low MOIs depends on the particular combination of virus and cell type, and is not an absolute but a relative term. What is relevant is that, in all cases, the $\log_2$-fold difference between high and low MOIs was consistently in the range of 10–20.

The time-series data for viral load are shown in Figure 1. Individual lineage series were fitted to ARIMA models and the best-fitting one was determined according to the minimum BIC criterium (Table S2). In each model, the slope parameter defines the rate of evolution per passage. The average rates of evolution for each combination of factors are also displayed in Table S2. In the case of HCoV-OC43 in BHK-21 cells, the average rate was significantly positive at a high MOI (runs test, $p < 0.001$) but not different from zero at a low MOI (runs test, $p = 1.000$). On average, rates were 13% faster at a high MOI. In the case of HC-OC43 in the HCT-8 cell, both rates of evolution were significantly negative (runs tests, $p < 0.001$ in both cases). Rates were, on average, 47% faster at a high MOI, although the difference was not significant (Mann–Whitney, $p = 0.719$). Finally, MHV in CCL-9.1 cells showed no difference from zero at a high MOI (runs test, $p = 1.000$) but were significantly negative at a low MOI (runs test, $p < 0.001$), declining 567% faster at a low MOI (Mann–Whitney, $p < 0.001$).
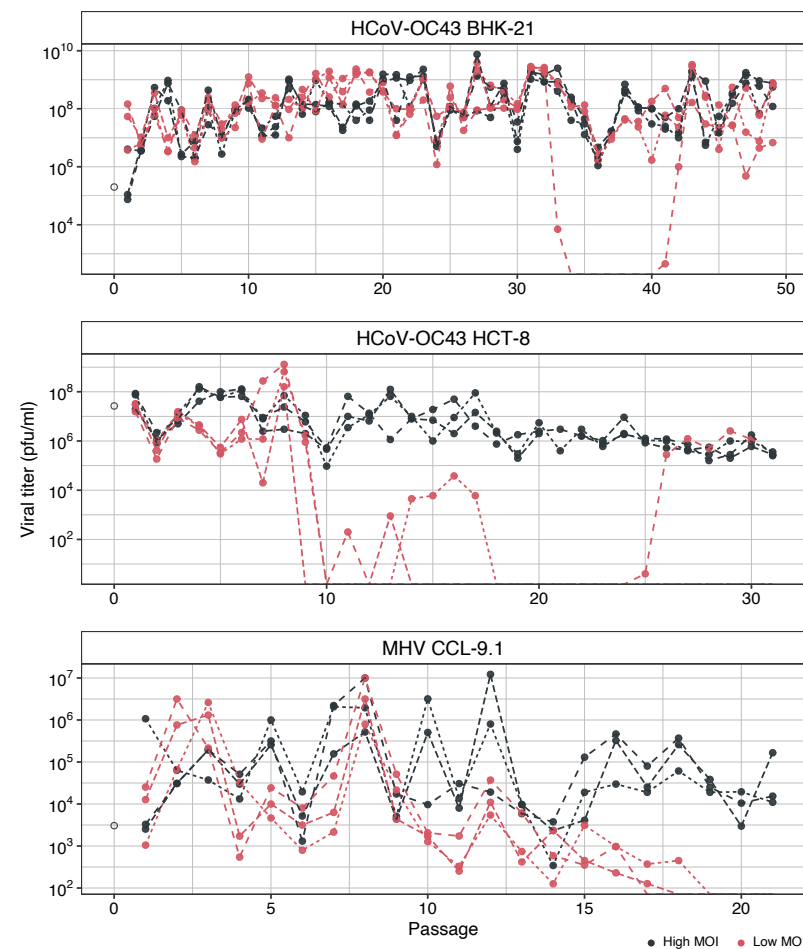

Figure 1. Evolution of viral loads at high and low MOIs. (**Upper panel**) HCoV-OC43 in BHK-21. (**Middle panel**) HCoV-OC43 in HCT-8. (**Lower panel**) MHV in CCL-9.1. Black dots and lines: high MOI; red: low MOI. Empty dot: ancestral virus (stock).

3.2. Evolution of DVG Richness and Abundance Is Cell Type and Virus Dependent

Since the sequencing was not strand-specific, in the following analyses we will not distinguish between DVGs found in the negative- or positive-sense RNA strains. In addition, prior to all the evolutionary relevant analyses, canonical subgenomic RNAs (sgRNAs) were filtered out by excluding any reconstructed DVGs whose junctions agreed with the expected coordinates of sgRNAs within a range of ± 10 nucleotides. The percentages of sgRNAs observed in each evolving lineage are shown in Figure S1. Interestingly, these percentages were affected by MOI, virus species, and host cell type. In the case of HCoV-OC43, 51.35% less sgRNAs relative to total DVGs were observed at a high MOI than at a low MOI in BHK-21. This reduction was 41.44% in HCT-8. These observations suggest that, for a given amount of sgRNAs, more DVGs were accumulating at high MOIs. In contrast, the trend was the opposite for MHV in CCL-9.1: 19.26% more sgRNAs per total DVGs were present at high MOIs than at low MOIs.

Next, we examined the dynamics of DVG generation and accumulation during the evolution experiments. Figure 2 shows two different measures of DVG diversity over time. Firstly, the number of unique DVGs was expressed as RPHT (Figure 2A), which represents diversity in terms of the different DVG types observed. Secondly, the total RPHT per DVGs (Figure 2B) was a measure of the total amount of DVGs relative to absolute viral accumulation. In all lineages, a common trend was observed: an increase in DVG richness from the initial to intermediate stages of viral evolution (Figure 2A), with no significant effect of MOI (Wilcoxon paired-samples tests, $p \geq 0.382$ in all three cases). Throughout the course of the evolution experiment, some disparities were noted between parallel replicates, but the overall pattern for HCoV-OC43 in both cell types was a reduction in richness after the initial surge, maintaining a relatively stable level until the end of the experiment, with no discernible differences observed between both cell types. For MHV at a high MOI, DVG richness increased up to passage nine and then remained stable until the end of the experiment. However, at a low MOI, MHV displayed differences between parallel lineages, with the number of DVG reads either increasing or decreasing after passage nine, depending on the lineage. The shifts in richness at different passages were more remarkable for HCoV-OC43 in BHK-21 and for MHV in CCL-9.1 cells.

Furthermore, the abundance of DVG counts increased to varying degrees at the onset of evolution in all lineages (Figure 2B). After the initial passages in low MOI lineages, the abundance of detected DVGs tended to decline, while in high MOI lineages, DVG abundance remained relatively steady or even slightly increased. Interestingly, notable disparities were noted in the accumulation of DVGs between high and low MOIs for HCoV-OC43 in both cell types (Wilcoxon paired-samples tests, $p = 0.022$ and $p = 0.041$, respectively), with DVGs consistently being more abundant in lineages evolved at a high MOI. In contrast, no significant differences in terms of the total number of accumulated DVGs were observed for MHV lineages evolved at high and low MOIs (Wilcoxon paired-samples test, $p = 1.000$).

Next, we tested whether DVG richness and abundance were correlated, that is, whether more abundance would contain more different DVG types. The alternative being abundance resulting from the accumulation of a few different DVGs. Spearman correlations were significant for HCoV-OC43 infecting BHK-21 at a low MOI ($r_S = 0.882$, 13 d.f., $p = 0.001$) as well as for MHV at both MOIs (low: $r_S = 0.900$, 10 d.f., $p = 0.002$; high: $r_S = 0.776$, 13 d.f., $p = 0.005$) but not in the other three cases. These observations suggest that viral populations containing more DVGs may also contain more diverse types.
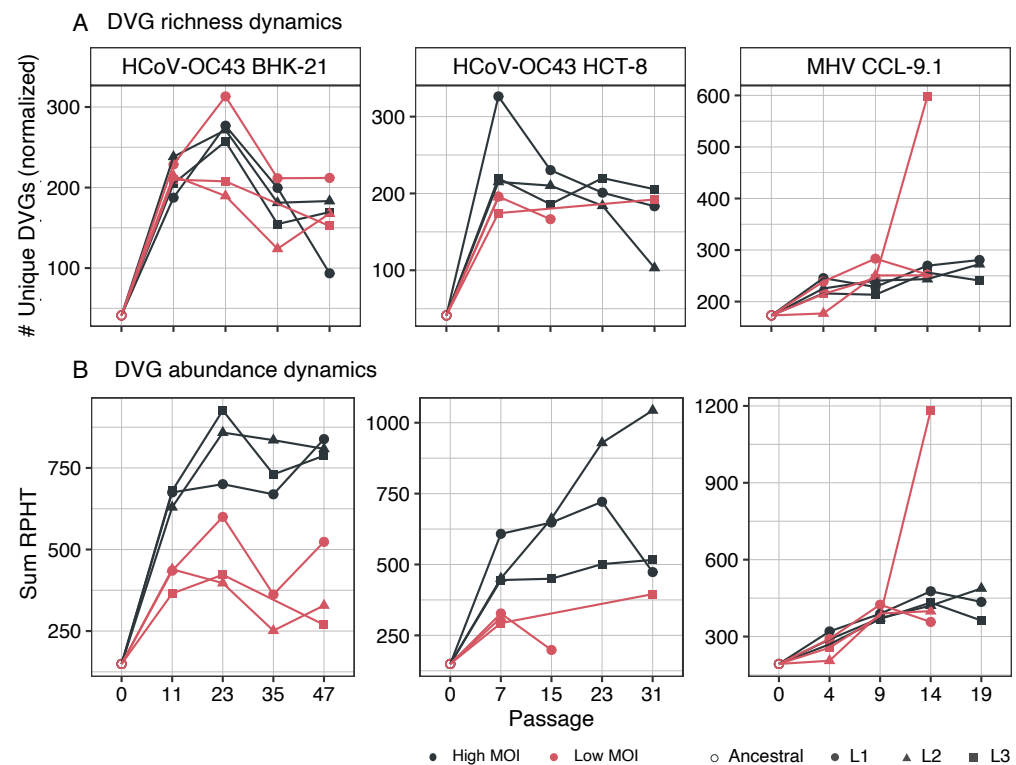
Figure 2. Evolution of richness and abundance of DVGs. (**A**) DVGs richness, as the number of unique DVGs relative to the total viral RPHT. (**B**) DVGs abundance, estimated as the total number of DVG read counts relative to the total viral reads in the sample, also RPHT. Black color represents high MOI while red represents low MOI. Different lineages are presented by different symbols. When a lineage was extinct, no point is shown.

3.3. Deletions Are the Most Pervasively Reconstructed Type of DVG

DVG classes considered in the reconstruction include deletions, insertions, and cb or sb genomes (hereafter all referred to as cb) at 5′ or 3′ ends. Richness was evaluated as in Section 3.2 but now distinguishing between the four types of DVGs; estimates are shown in Figure 3A. Firstly, we asked whether the two viruses show differences in the number of unique DVGs types. A Scheirer–Ray–Hare (SRH) two-way non-parametric ANOVA test showed highly significant differences between the two viruses ($p < 0.001$), with HCoV-OC43 generating, on average, 47% more different DVGs than MHV. Likewise, the overall distribution of different DVGs also differed among the four classes of DVGs, with deletions being the most abundant type, followed by insertions and with 3′ and 5′ cb being the least common ($p < 0.001$). Interestingly, these differences among the four classes were consistent for the two viruses ($p = 0.442$). Secondly, we interrogated the HCoV-OC43 data for differences among cell types. In this case, the SRH test found no differences in the number of different DVG types among the two cells ($p = 0.093$), although the distribution of the four DVG types was strongly biased by deletions and insertions ($p < 0.001$). Thirdly, we tested whether MOI had an overall effect on the number of different DVGs generated. In this case, the SRH test also found significant differences ($p < 0.001$) between lineages evolved at high and low MOIs, with the former containing 74% more variants than the latter. The rank-order of the four classes of DVGs was not affected by MOI ($p = 0.894$).
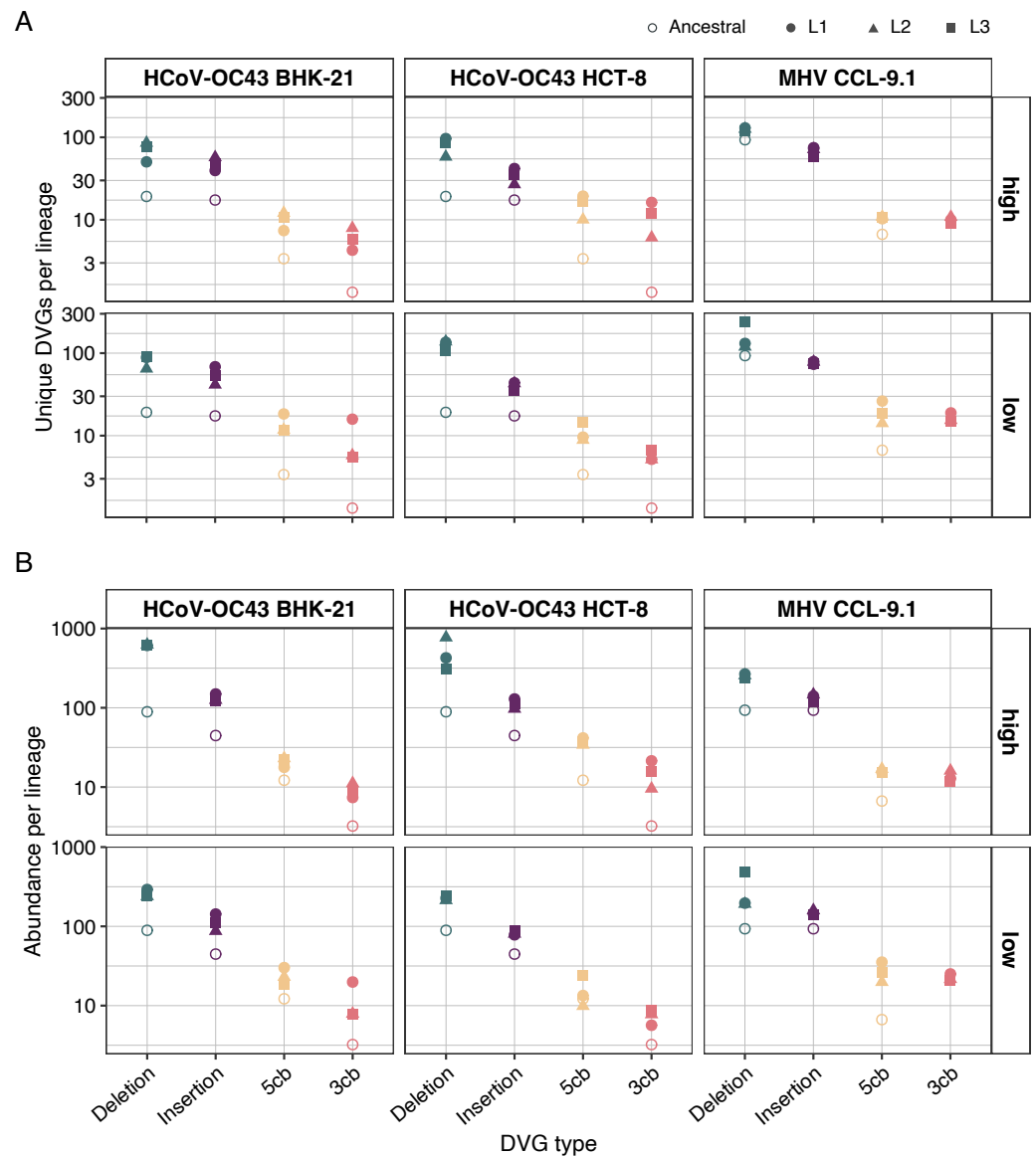
Figure 3. Diversity of DVG classes detected in each lineage of experimental evolution. (**A**) Richness represented as the normalized number of unique DVGs present in each lineage. (**B**) Abundance represented as the normalized number of reads supporting the DVG founded in each lineage. Empty points represent the stock sample (passage 0), the values for the stock in HCoV-OC43 are the same for BHK-21 and HCT-8 at both MOIs. The MHV ancestral is the same for both MOIs in CCL-9.1.

Abundance was evaluated as in Section 3.2 but now distinguishing between DVG types; estimates are shown in Figure 3B. Following the same logic as in the previous paragraph, we firstly sought for an effect of virus species on DVG abundance. In this case, the SRH test found no significant differences among the two viruses in the abundance of DVGs ($p = 0.575$). Overall, deletions were the most variable class, followed by insertions and 5' and 3' cb (these two at approximately the same variability) ($p < 0.001$); these differences were not affected by the viral species ($p = 0.093$). Secondly, we found no significant differences among the two cell types in the abundance of DVGs generated from the HCoV-OC43 genome ($p = 0.835$), nor an effect on the distribution of the four classes of DVGs ($p = 0.994$). Finally, differences in MOI had an overall effect on DVGs abundance, being 35% more abundant in lineages evolved at a high MOI ($p < 0.001$), although this effect did not translate into a change in the proportion of the four classes of DVGs ($p = 0.856$).

3.4. The Distribution of Deletion Sizes Suggests That Purifying Selection upon DVG Size Depends on Particular Virus–Cell Type Combinations

Next, we honed in on deletions for our analysis, as they were the most common type of DVGs. We examined the sizes of defective genomes with deletions at each sampled time point. One-base deletions were predominant in all the lineages. Thus, to uncover the pattern of remaining deletion sizes, we filtered out DVGs with deletions shorter than 50 nucleotides.

The evolution of the distribution of deletion sizes, accounting for their corresponding abundance, is shown in Figure 4A. The size distribution of HCoV-OC43 genomes with deletions differed between high and low MOI passages in both cell types, with bigger deletions being more common at a high MOI (Figure 4A, black bars). Notably, certain deletion sizes at a high MOI became predominant throughout the evolution (BHK-21: ~12.4 kb, ~22.5–30.7 kb; HCT-8: ~13.6 kb). HCoV-OC43 evolved to cell-specific size distributions, although some lengths were common to both cell types. At a low MOI (red bars in Figure 4A; for a more detailed presentation, see Figure S2), large deletion sizes ~30–31 kb are predominant for HCoV-OC43, followed by ~18 kb in BHK-21 and ~18–20 kb in HCT-8 cells. The 30,698-nucleotide-long DVG is widely present across all the lineages and MOIs of HCoV-OC43 infected cells but it was also present in the viral stock.

For MHV, there were also differences in the distribution of deletion sizes between high and low MOIs, although it is less obvious than what was shown for HCoV-OC43. The size of the most frequent deletions at a high MOI were ≥17 kb (Figure 4A, black bars), with varying maxima to different time points along viral evolution. The last passage showed a clear predominance of the longest deletions (~31.1 kb, 19.1 kb). However, the distribution of deletion sizes at a low MOI lacked a clear abundance pattern (Figure S2).

From these analyses, we can conclude that although the generation of short deletions is more likely, the occurrence of very long deletions is also common. The size distribution of deletions is highly dynamic and underwent changes throughout the course of the evolution experiment, with certain lengths being pervasively favored. This dynamism is influenced by the combination of viral species and cell type, with the viral species being the most influential factor.

Aguilar Rangel et al. [14] proposed the idea that the triplet periodicity of the genetic code must place a constraint on deletions in coding sequences. To test this hypothesis, we evaluated whether a characteristic multiplicity existed in the distribution of deletion size. A Lomb–Scargle periodogram (an adaptation of the classic Fourier transformation for unevenly sampled data [38]) of deletion sizes revealed a robust tri-nucleotide periodicity in the cistrons (Figure 4B). HCoV-OC43 data in both cell sizes and MOIs show a characteristic frequency multiple of three (a strong signal at period three and a weaker one at period six). No obvious periodicity has been observed for MHV at a high MOI but a weak signal at period three and a stronger one at period six appears to exist (Figure 4B). This analysis suggests that deletions more likely to disrupt the translation reading frame did not accumulate as much as those that keep it, suggesting that DVG size was evolving under a purifying selection.
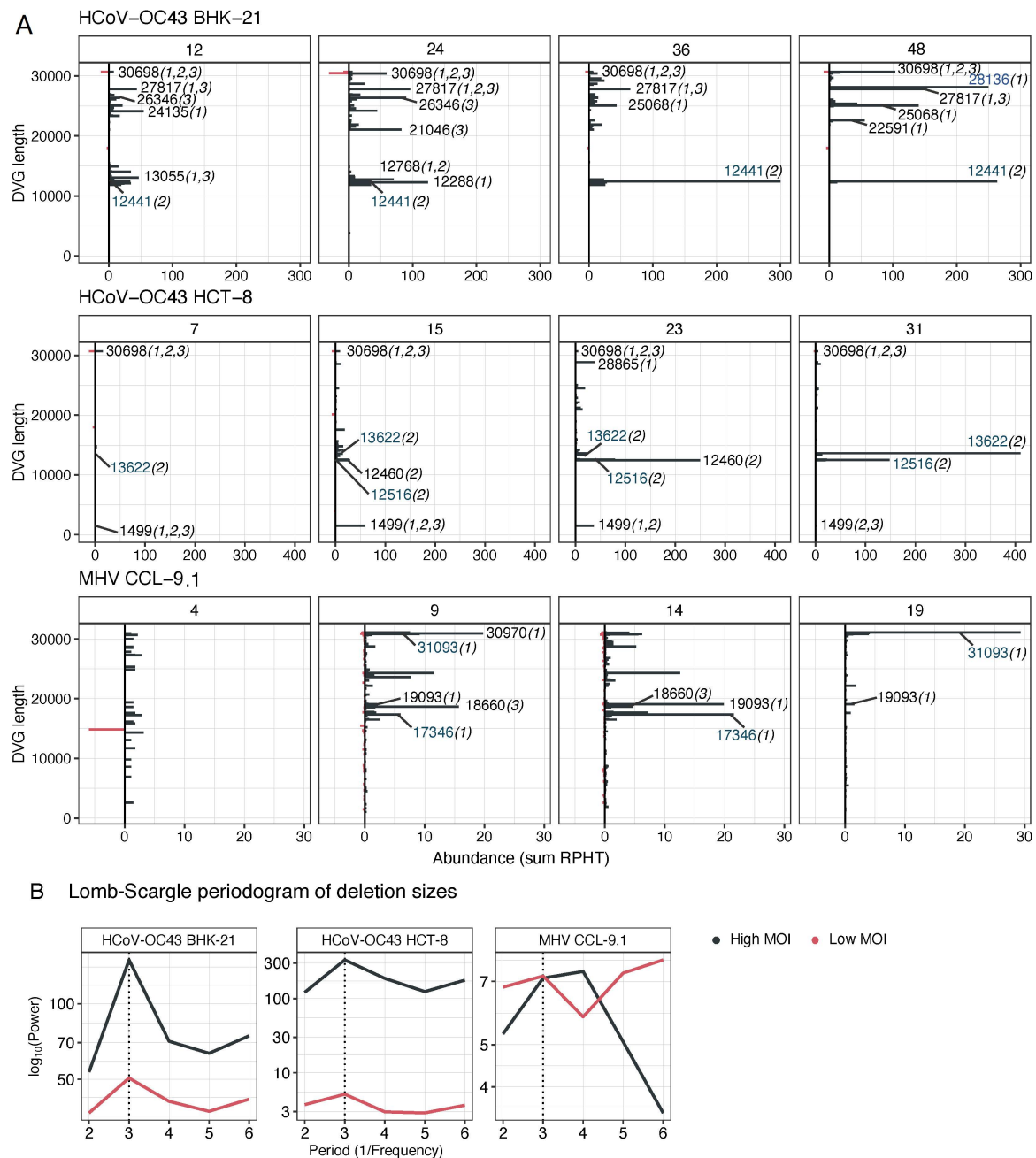
Figure 4. Distribution of deletion sizes. (**A**) Changes in the size distribution of deletions along the evolution experiments. Size distribution of genomes with deletions across different evolutionary time points (upper box of each panel, passage numbers), taking into account the variability and abundances of the reconstructed deletions. Replicates of lineages were combined into a single diagram. Numbers indicate some predominant sizes, between parenthesis the lineages in which they were found. Blue numbers highlight the lengths that gain predominance with evolution. Genomes with deletions of fewer than 50 nucleotides were excluded from the visualization. (**B**) Lomb–Scargle periodogram of deletion sizes. All deletions > 3 nucleotides were included in the frequency analysis.

3.5. Factors That Influence the Occurrence of Deletions

Next, we conducted an analysis to determine if specific domains of the viral genome were more prone to act as BP and RP for deletions owing to their sequence or involvement in structural elements. Firstly, we hypothesized that the emergence of deletions could be linked to the very same mechanisms involved in sgRNA formation. These mechanisms in the coronavirus genome involve TRS. To test this possibility, we calculated the frequency of

each genomic position as BP or RP of a deletion (Figure 5A). Deletions shorter than three nucleotides were excluded hereafter based on the consideration that very small deletions, such as sequencing errors [37], must follow a different distribution compared to larger, biologically relevant events. At this stage, we want to restress that canonical sgRNAs were also filtered out before the analysis to avoid spurious results.
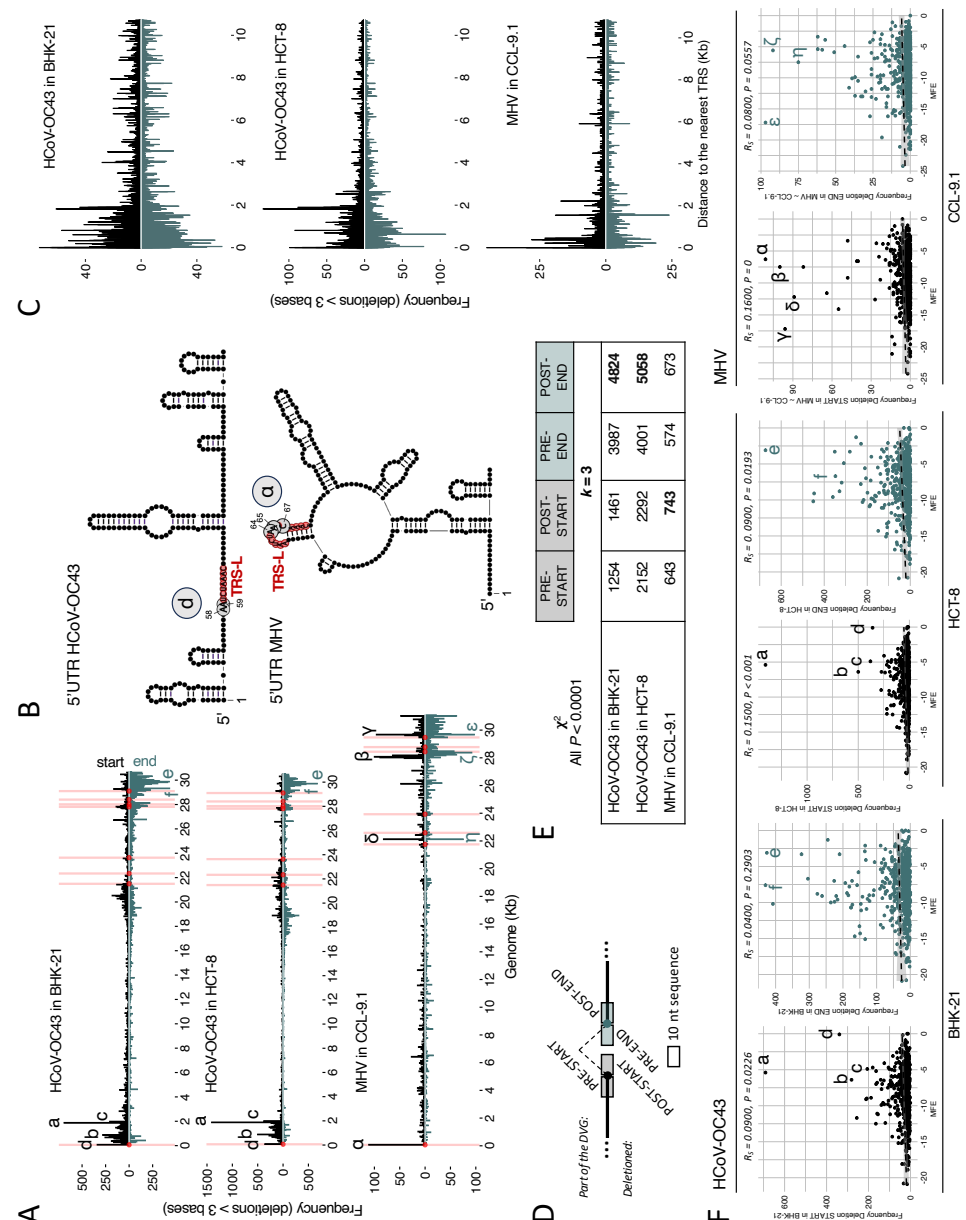

Figure 5. Genomic determinants of DVG formation. (**A**) Frequencies of BP and RP of deletions > 3 at each genome positions (in kb). Each bar of the histogram represents a 50-base window. The upper part of the diagram indicates start points, while the lower part does so for end points. Specific TRS positions are highlighted with pink bars and with circles at the base of the start and end points. The most frequently affected areas are indicated with letters a-f for HCoV-OC43 and α–η for MHV. HCoV-OC43 ranges: a = 1850–1900; b = 850–900; c = 1950–2000; d = 50–100; e = 29,850–29,900; and f = 29,800–29,850. MHV ranges: α = 50–100; β = 28,050–28,100; γ = 29,650–29,670; δ = 28,000–28,050, ε = 29,650–29,700, ζ = 28,350–28,400, and η = 2100–22,150. (**B**) 5′UTR structure prediction and localization of one of the most frequent start points (d in HCoV-OC43 and α in MHV) of deletions; TRS-L highlighted in red. (**C**) The distance in nucleotides from each deletion's BP (above) and RP (below) to the nearest TRS (in kb). (**D**) Schematic representation of DVG junction with pre- and post-

BP and RP sequences as studied in the *k*-mer analysis. (**E**) Results from the χ^2 test of 3-mer distribution between the sequences around the deletion compared to those found in the complete genome. Sequences of 10 nucleotides before the BP (pre-BP), 10 after the RP (post-RP), 10 after the BP (post-BP), and 10 before the RP (post-RP) were analyzed. All the results are significant ($p < 0.001$), with the highest χ^2 value of each case bolded. (**F**) Correlation between frequency of BP (black) or RP (green) points with the minimum free energy (MFE) predicted for the 50-base window secondary structure.

In HCoV-OC43, BP accumulated within the first 2 kb of the genome, while the more frequent RP accumulated within the last 2 kb (Figure 5A), with additional likely sites around 415 nucleotides from the 3′ end in BHK-21 and 765 and 1215 in HCT-8. Interestingly, the most likely BP d is two nucleotides upstream from the TRS-L and a few nucleotides before a large structural hairpin in 5′UTR (Figure 5B). This observation was consistent for both cell types. In the case of MHV, the most likely BP was labeled as α, and it lay in the middle of TRS-L and the loop of a hairpin structure (Figure 5B). The density of potential BP and RP in MHV was also high at the last third of the genome for BP and RP, but highly frequent sites were also found in the TRS-L sequence (67 nucleotides), at 970 and at 5975 nucleotides for BP and at 996 and 5996 nucleotides for RP (Figure 5A).

To further explore the proximity of BP and RP to canonical the breakpoints for sgRNA generation (i.e., TRS), we calculated the distance in nucleotides from each deletion's BP (above) and RP (below) to it nearest TRS (Figure 5C). The frequency distribution at various distances from TRS (~0–11 kb) appeared relatively symmetric for both BP and RP (Figure 5C). The most frequent recombination points in all cell types, but particularly in BHK-21, were concentrated within a range of 2 kb around TRS (where the a–c hotspots were also concentrated; Figure 5A), showing a consistent decrease in frequencies with increasing distance from TRS (Figure 5C).

Next, we explored whether the likelihood of being BP and RP can be attributed to specific sequences. The richness in A/U sequences around breakpoints was documented previously for the influenza virus [43] and previous studies with coronaviruses pointed to the UUG triplet as the preferred sequence for spanning junction start positions and a significant A increase in the end positions [44]. Following that, we analyzed the distribution of nucleotides (*k*-mer analysis) of the surroundings of the recombination points (Figure 5D). Figure 5E shows the results of the χ^2 test supporting distinct 3-mer distributions between sequences around the deletion coordinates and those found in the entire genome ($p < 0.001$ in all cases). Analysis of shorter *k*-mers produced identical results. In addition to all nucleotide distributions being significantly different at the recombination coordinates, we observed in HCoV-OC43 that, irrespective of the cell type, the most notable changes in the *k*-mer distribution were observed in the sequences surrounding the RP. In MHV, the highest χ^2 values were in the sequences downstream of the BP (Figure 5D).

Inspired by the results shown in Figure 5B associating highly frequent BP with RNA secondary structures, we tested whether the likelihood of being BP or RP could be associated with the complexity of secondary structures. We assumed that the lower the minimum free energy (MFE), the more complex the RNA structure [45], and the more likely the RNA polymerase will jump out of the template and rejoin somewhere else. As shown in Figure 5F, very small yet significant Spearman correlations were observed between the frequency of BP and MFE for both viruses and, in the case of HCoV-OC43, both cell types (in all cases, $p \leq 0.023$). Regarding RP, a significant correlation was only found for HCoV-OC43 in HCT-8 ($p = 0.019$). Qualitatively, regions with the highest deletion frequency tended to have intermediate MFE values. This suggests that deletions occurred more frequently in structured regions, but not necessarily in highly complex structures.

3.6. E and Nsp12, Two Viral Proteins with Viroporin Potential, Are the Most Affected by Deletions

Now our focus has moved on to exploring whether deletions evenly affected all cistrons or were preferentially found in certain cistrons. To achieve this, we calculated the density of deletions per cistron by dividing the number of deletions found in a given cistron

by its length. Data are shown in Figure 6. These data were then fitted to a generalized linear mixed model (GLMM) in which cell type, MOI, and cistron were treated as fixed effects and lineage (nested factor within the interaction of the fixed factors) and passage (within-individual repeated measures) as random ones, and a Gaussian distribution and identity link function were used. Focusing on HCoV-OC43, no differences among cell types (main or in combination with other fixed factors) were found ($\chi^2 \leq 1.971$, $p \geq 0.923$). MOI had a net effect, with density across cistrons being 178% higher at a high MOI ($\chi^2 = 55.961$, 1 d.f., $p < 0.001$). Also, significant overall differences existed among passages ($\chi^2 = 36.509$, 6 d.f., $p < 0.001$), with the density of deletions across cistrons increasing with evolutionary time. Highly significant differences were found among cistrons ($\chi^2 = 126.321$, 7 d.f., $p < 0.001$), with ORF1ab and S showing the lowest density of deletions whilst E and ns12.9 showed the highest. Most interestingly, a significant interaction between MOI and cistron was observed ($\chi^2 = 36.298$, 7 d.f., $p < 0.001$), with a reduction in the magnitude of the differences between cistrons at a low MOI.
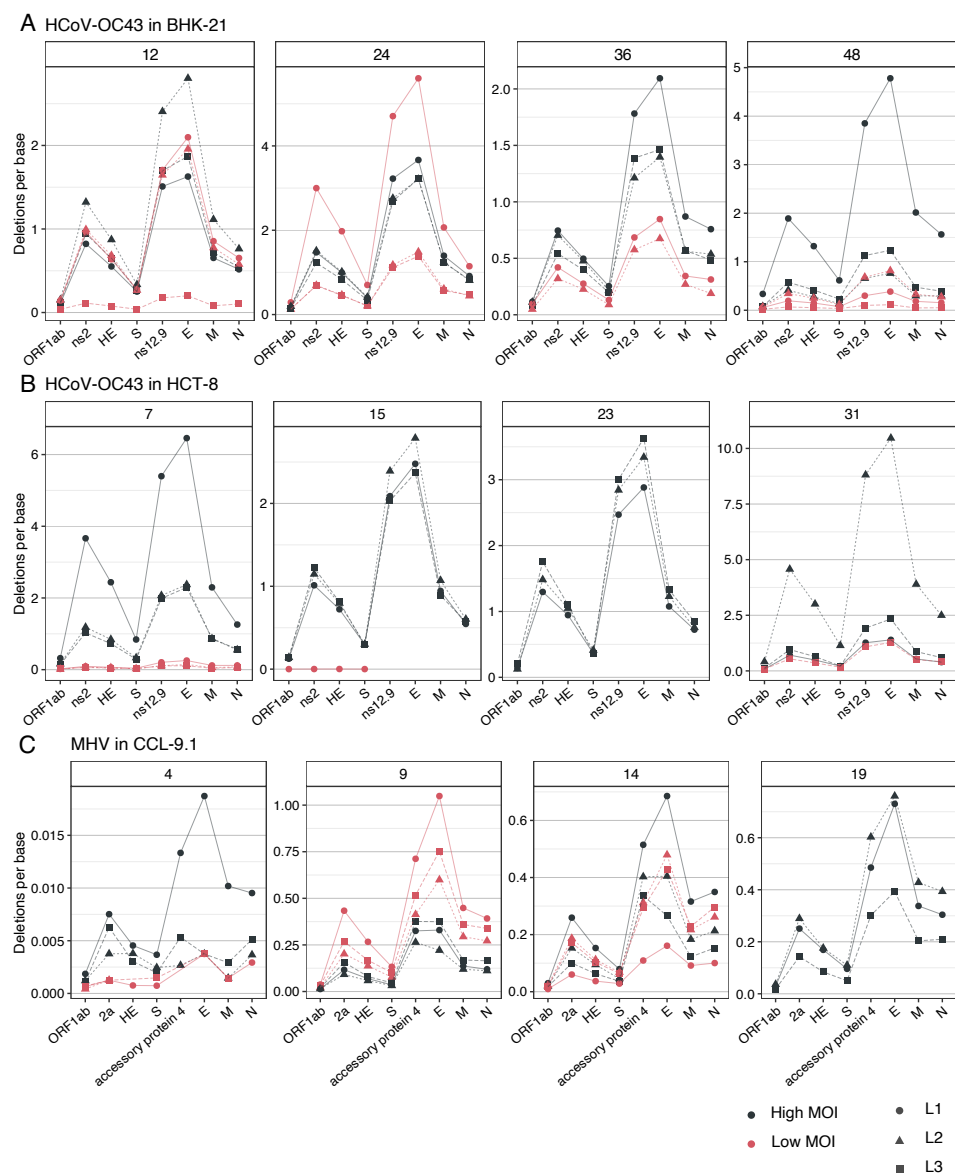

Figure 6. Observed frequency of deletions per cistron per base. Each panel corresponds to the indicated passage in the evolution experiment. (**A**) HCoV-OC43 evolved in BHK-21 cells. (**B**) HCoV-OC43 evolved in HCT-8 cells. (**C**) MHV evolved in CCL-9.1 cells.

The results for MHV were somehow different. A significant increase in the overall deletion density along evolutionary time (χ^2 = 88.730, 3 d.f., *p* < 0.001) was observed. Significant differences among cistrons existed (χ^2 = 75.268, 5 d.f., *p* < 0.001), again with ORF1ab and S showing the lower deletion density and the auxiliary protein 4 and E showing the highest. All other factors and combinations were not significant.

3.7. A Subset of DVGs Persists along the Evolution Experiment

Next, we examined the temporal changes in the proportions of persistent DVGs during the experimental evolution. DVGs qualified as de novo if they were observed only in a single passage, while they were qualified as persistent if, after the first detection, they were observed at subsequent passages of the same lineage. For each virus species and cell type independently, data in Figure 7A were fitted to a GLMM with MOI and type of DVG treated as fixed effects and lineage (nested factor within the interaction of the fixed factors) and passage (within-individual repeated measures) as random ones, and a Gamma distribution and log link function. In the case of HCoV-OC43 in BHK-21, as the number of passages increases, the percentage of persistent DVGs rose consistently in BHK-21 cells for all DVGs types (χ^2 = 18.830, 3 d.f., *p* < 0.001) regardless of the MOI (χ^2 = 4.872, 3 d.f., *p* = 0.181). Interestingly, a highly significant overall difference was found among DVG types (χ^2 = 515.509, 3 d.f., *p* < 0.001), with deletions and insertions being the most common type of persistent DVGs (consistent with the results shown in Section 3.3). Yet, the rate at which DVGs accumulated was the same for all types (χ^2 = 8.859, 9 d.f., *p* = 0.450). Another interesting result of this analysis was the significant interaction between DVG type and MOI (χ^2 = 13.705, 3 d.f., *p* = 0.003), due to a disproportionate larger reduction (57.68%) of deletions at a low MOI (5.31% for 3′ cb, 4.39% for 5′ cb, and 22.33% for insertions). However, the pattern observed for HCoV-OC43 in HCT-8 cells differed from that in BHK-21 cells. There was a net overall effect of MOI (χ^2 = 14.602, 1 d.f., *p* < 0.001), with 65.16% more DVGs accumulating at a high MOI, being that this effect also dependent on the evolutionary time (χ^2 = 6.332, 2 d.f., *p* = 0.042), with an increasing trend for all types of DVGs at a high MOI but mostly declining at a low MOI.

The case of MHV showed similarities and differences with HCoV-OC43 (Figure 7A) as all the terms in the GLMM were highly significant. Among the similarities, the proportion of persistent DVGs increased with the number of passages (χ^2 = 72.091, 3 d.f., *p* < 0.001), although this increase depended both on the type of DVG and on MOI (χ^2 = 85.477, 6 d.f., *p* < 0.001). Also, overall, the abundance of DVGs depended on their type, with deletions being the most common followed by insertion, 5′, and 3′ cbs (χ^2 = 253.615, 3 d.f., *p* < 0.001). The most remarkable difference was that the overall effect of MOI went in the opposite direction: 24.15% more accumulation at a low MOI. The increase in DVG abundance at a low MOI might seem counterintuitive, but various factors, such as the possibility of secondary infections at a low MOI (especially for MHV with a short infection cycle), could contribute to co-infection and the eventual transmission of DVGs.
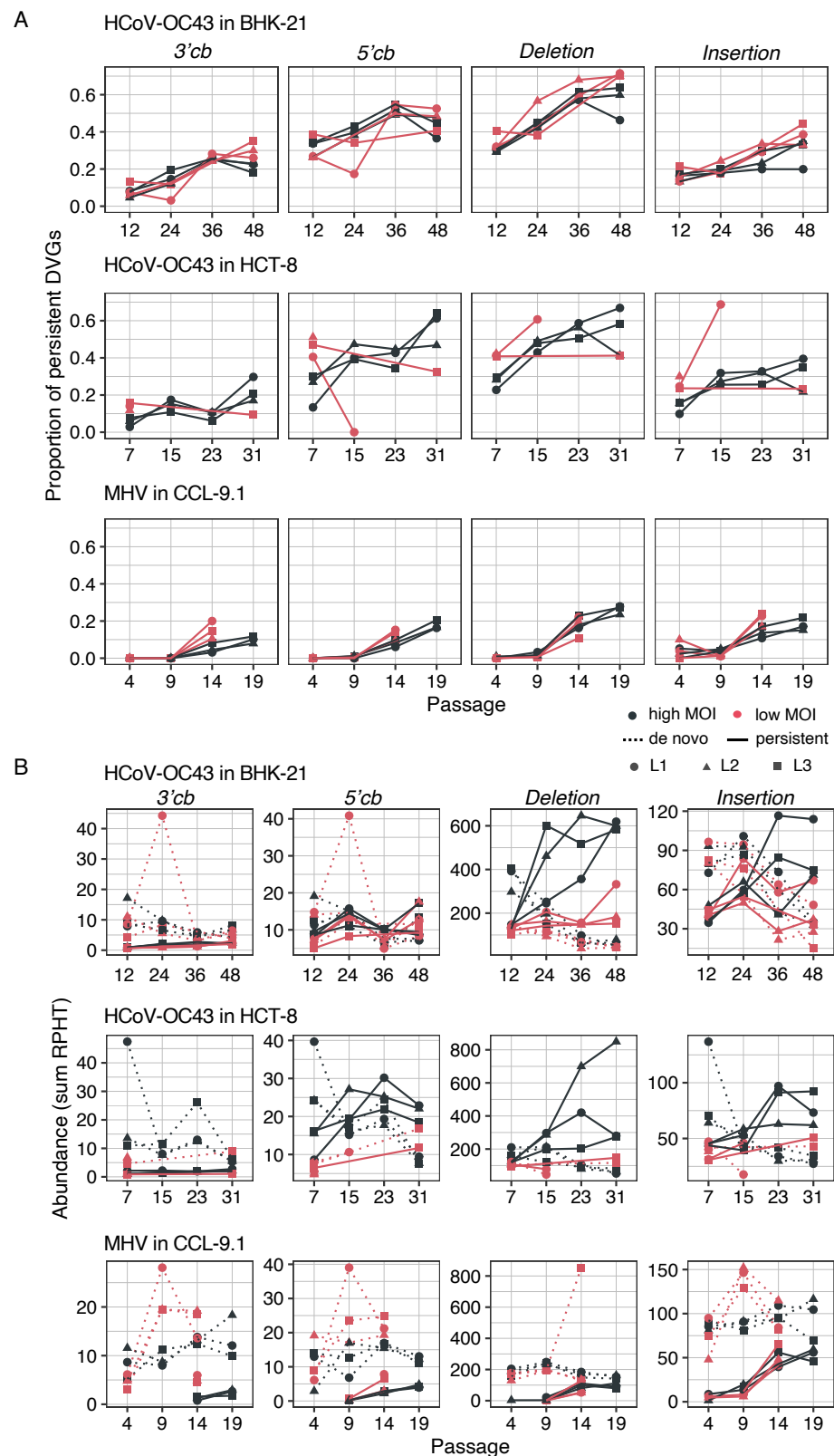
Figure 7. Transmission of persistent DVGs. (**A**) Proportion of persistent DVGs at different time points of experimental evolution, separated by DVG type. The proportion was calculated as the ratio of DVGs classified as persistent among all DVGs in the sample. (**B**) Abundance of persistent (solid lines) and de novo generated (dotted lines) DVGs at high (black) and low (red) MOIs, separated by DVG type. Evolutionary lineages are represented by shapes. RPHT: DVG reads per hundred thousand viral reads.

We considered that the most parsimonious explanation for the temporal persistence of DVGs was that they were transmitted across passages. This assumption was supported by the occurrence of deletions being not random but associated with specific regions of the viral genome, suggestive of some sort of selection at play. However, we could not rule out the possibility that some deletions might pervasively emerge de novo at the same coordinates. To test this possibility, we calculated the relative abundance of reads categorized as persistent or as de novo at the various time points, distinguishing between different DVG types and MOI condition (Figure 7B). This distinction was added as an additional fixed factor to the GLMM above and its effect on the abundance of the different types of DVGs was evaluated. In the case of HCoV-OC43 in BHK-21 cells, this factor was significant by itself (χ^2 = 26.159, 1 d.f., $p < 0.001$) as well as in interaction with all the other factors (in all cases, $\chi^2 \geq 6.613$, $p \leq 0.010$). Three particularly relevant cases were the following: (*i*) the interaction of this factor with passage number (χ^2 = 26.159, 1 d.f., $p < 0.001$); while de abundance of the de novo DVGs decreased, on average, 54.17% along the evolution experiment, transmitted DVGs had increased 114.36%. (*ii*) The interaction with MOI (χ^2 = 6.613, 1 d.f., $p = 0.010$). At a high MOI, transmitted DVGs were 46.01% more abundant than at a low MOI, while de novo ones were only 19.87% more, suggesting that a high MOI favors the transmission of already existing DVGs. (*iii*) The interaction with DVG type (χ^2 = 250.006, 3 d.f., $p < 0.001$). While deletion DVGs were 146.50% more abundant among the transmitted class than among de novo class, all other types were less abundant in the transmitted class (79.74% for 3′ cb, 5.99% for 5′ cb, and 4.86% for insertions). When analyzing the patterns in HCT-8 cells, the results were not fully congruent. For example, the interaction with MOI was not significant (χ^2 = 0.403, 1 d.f., $p = 0.525$). The situation with MHV was similar to HCoV-OC43 in HCT-8 and we are not discussing it in detail.

3.8. Hotspots of BP and RP for Persistent DVGs

In Sections 3.5 and 3.6, we demonstrated that recombination events leading to deleted genomes were not randomly distributed across the genome. Here, we explored whether specific hotspots were overrepresented among persistent DVGs, potentially linked to their viability in *trans*-complementation with the wild-type helper virus. Figure 8 illustrates the coordinates of BP and RP for persistent DVGs across all evolutionary lineages. As a reference, Figure 8A,C show schematic representations of the viral genomes. Overall, the most frequent coordinates in HCoV-OC43 infecting BHK-21 (Figure 8B top) aligned with the findings shown in Figure 5A. BPs were notably concentrated within the first 2 kb at the 5′ end of the genome, with another densely populated region between positions 20 and 22 kb. Notably, BPs exhibit minimal variability, while RPs were more diverse, spanning positions from 17–30 kb. This was particularly evident for the two vertical lines around positions 50 and 2000, corresponding to the previously characterized a and d hotspots (Figure 5A). At a low MOI, we observed fewer hotspots, with a high density within the first 2 kb, similar to a high MOI (Figure 8B top). However, the high-density points in the last third of the genome at a high MOI were absent in the low MOI samples. In lineages evolved in HCT-8 (Figure 8B bottom), persistent deletions condensed at the same hotspots as in BHK-21. A notable difference was observed: in HCT-8, the variability of BP increased at specific positions, forming a range of BPs spanning positions from 0 to 5 kb and another range around positions 20 to 24 kb (Figure 8B bottom). These hotspots seemed to be specific to HCoV-OC43 in HCT-8 cells. At a low MOI, the limited number of persistent DVGs hampered pattern recognition, but the identified spots aligned well with those seen at a high MOI.
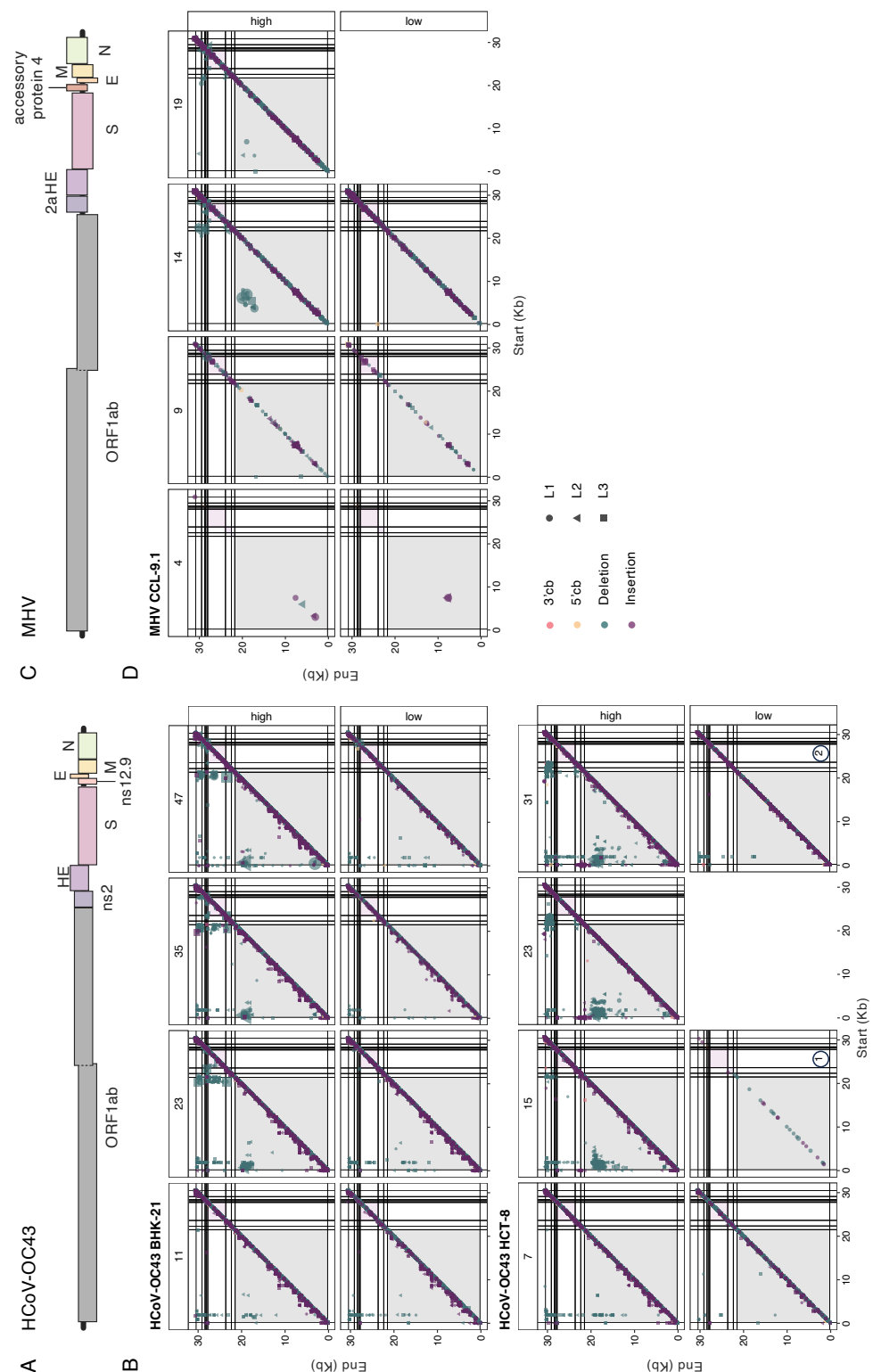
Figure 8. Distribution of persistent DVGs BP and RP. (**A**) Schematic representation of HCoV-OC43 genome. (**B**) Coordinates found for HCoV-OC43, at each passage characterized, in both cell types. (**C**) Schematic representation of MHV genome. (**D**) As in (**B**) but for MHV. Coordinates from parallel lineages were joined into the same plot and represented by different symbols. Background colors indicate the cistrons as in (**A**,**C**).

The number of spots for persistent DVGs was small for MHV at both high and low MOIs (Figure 8D). However, two regions with accumulated coordinates for DVGs were

observed at later evolutionary passages. These hotspots spanned from positions 3 to 9 kb and from 16 to 22 kb.

3.9. Independently Evolved Lineages Show Convergent Deletions, Reinforcing the Idea That Some DVGs Must Be under Positive Selection

Next, we aimed to assess the level of reproducibility in the generation of DVGs amongst lineages of the same virus evolved in the same cell host type. Figure 9 depicts the counts of unique deletions (>3 nucleotides) shared among the three independent lineages. A question that immediately arose was whether the count of shared unique DVGs among lineages differed from what would be expected based on the composition of each individual lineage. To illustrate this assessment, we focused on HCoV-OC43 in BHK-21 at a high MOI (Figure 9A). The total count in lineage 1 was 4920 + 116 + 140 + 126 = 5302, 3831 for lineage 2, and 3465 for lineage 3. The count of unique DVGs observed exclusively in lineage 1 was 4920. The expected counts for this group can be calculated as $(4920/5302) \cdot (1 - 3532/3831) \cdot (1 - 3142/3465) \cdot (5302 + 3831 + 3465) \approx 85$. Using the same rationale, the expected counts could be computed for the other six groups displayed in the panel (BHK-21 low: ≈ 53; HCT-8 high: ≈ 248; HCT-8 low: ≈ 13; CCL-9.1 high: ≈ 14; CCL-9.1 low: ≈ 2). Applying a goodness-of-fit test, we discovered that observed counts significantly deviated from the expected values ($\chi^2 = 586{,}779.739$, 6 d.f., $p < 0.001$). This deviation was primarily due to the deficit in counts in the three pairwise comparisons and especially in counts in the group shared by all three lineages. This finding suggested that most of the observed deletion DVGs were produced in a lineage-specific manner, while very few were produced in an almost deterministic manner and shared by all lineages. Similar qualitative outcomes were obtained for HCoV-OC43 in other scenarios (Figure 9A,B), with a very large excess of linage-specific unique DVGs ($\chi^2 \geq 15{,}940.544$, 6 d.f., $p < 0.001$). The same results were found for MHV at both MOIs ($\chi^2 \geq 71{,}675.329$, 6 d.f., $p < 0.001$).

Noticeably, despite their scarceness, for both viruses and regardless of cell types, the count of shared DVGs among the three lineages consistently remained higher at a high MOI (126 vs. 47 and 308 vs. 12, for HCoV-OC43 in BHK-21 and HCT-8, respectively; and 20 vs. 2 for MHV).

Subsequently, to further characterize the few common deletions shared by the three parallel lineages, we examined their coordinates. In HCoV-OC43 samples (Figure 9A), the coordinates of deletion BP and RP paralleled with the coordinates of the persistent deletions shown in Figure 8B. Again, we observed that the most frequent breakpoints were at the 5′ end of the genome for both cell types and MOIs (within the first 2 kb) and in the last third part of the 5′ genome (around 21 kb) appeared at a high MOI, here more specifically in the HCT-8 samples. Similar to the case of persistent DVGs, we observed a limited number of deletions of BP but highly variable RP.

Finally, we examined the impact of cell type on the generation of common deletions among HCoV-OC43 lineages (Figure 10). Consistent with the probabilistic computations detailed above, we observed a substantial deviation between the actual counts in each of the three categories and the expected count distribution for both MOIs ($\chi^2 \geq 28{,}669.853$, 2 d.f., $p < 0.001$). This effect primarily stemmed from a shortage of shared DVGs between the two cell types, suggesting that the spectrum of DVGs generated by HCoV-OC43 was influenced, to some extent, by the specific cell in which it was replicating. Notably, this deficit was more pronounced for lineages evolved at a high MOI. Additionally, when comparing the count numbers in the three groups for lineages evolved at high and low MOIs, we discovered a highly significant distinction (homogeneity $\chi^2 = 4228.490$, 2 d.f., $p < 0.001$). Specifically, at a high MOI, lineages evolved in HCT-8 cells contained proportionally more unique DVGs than lineages evolved in BHK-21 cells, whereas the situation was reversed at a low MOI (Figure 10A).
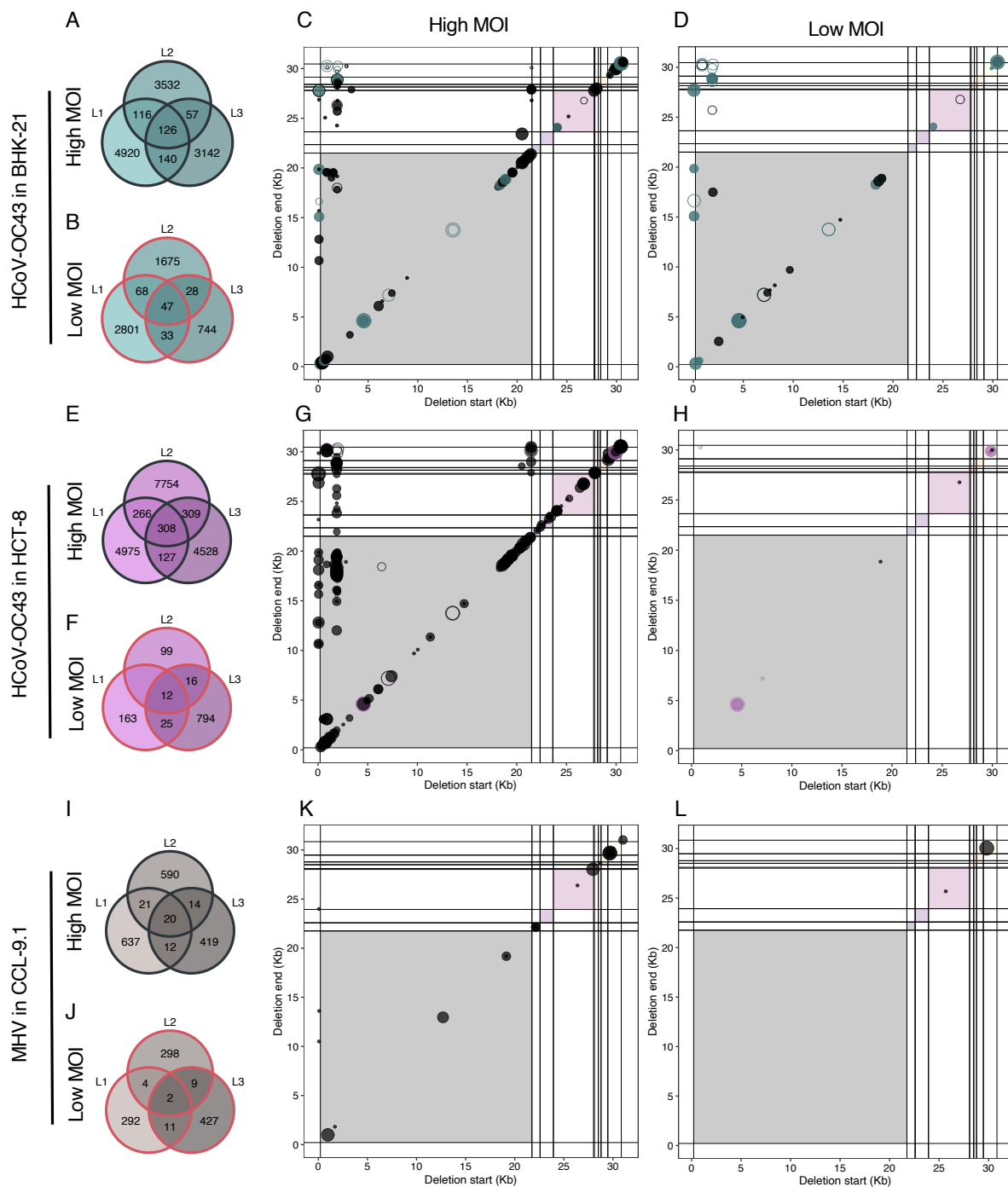
Figure 9. Convergencies between evolutionary lineages. Intersection size of deletions founded at lineages of HCoV-OC43 evolved in BHK-21 at (**A**) high and (**B**) low MOIs. BP and RP coordinates of common deletions shared by all the HCoV-OC43 lineages evolved in BHK-21cells at (**C**) high and (**D**) low MOIs. Intersection size of deletions founded at lineages of HCoV-OC43 evolved in HCT-8 at (**E**) high and (**F**) low MOIs. BP and RP coordinates of common deletions of parallel lineages of HCoV-OC43 in HCT-8 at (**G**) high and (**H**) low MOIs. Intersection size of deletions founded at lineages of MHV evolved in CCL-9.1 at (**I**) high and (**J**) low MOIs. BP and RP coordinates of common deletions of parallel lineages of HCoV-OC43 in HCT-8 at (**K**) high and (**L**) low MOIs. For all the scatter plots, the empty points indicate the deletion was already found in the ancestral virus; if the dot color is different from black, the deletion was convergent also between both MOIs. Background colors represent cistrons as in Figure 8A for HCoV-OC43 and Figure 8B for MHV. Symbols' size indicates the number of samples where the deletions were found.
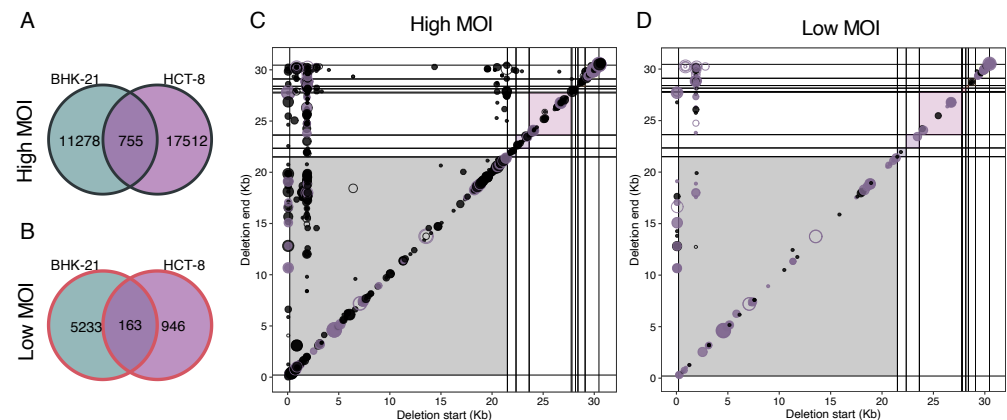
Figure 10. HCoV-OC43 deletion DVGs convergent among both cell types. (**A**,**B**) Intersection sizes at (**A**) high and (**B**) low MOIs. Hotspots of common deletions among cell types (intersection set in (**A**,**B**)) at (**C**) high and (**D**) low MOIs. Background colors indicate viral cistrons (as in Figure 8A), the size of the points indicate the number of samples where the deletion is present. Empty circles indicate that the deletions were already present in the ancestral stock and purple and black colors represent, respectively, common to both MOIs or unique to the MOI.

It was noteworthy that 114 out of the 163 convergent deletions identified under low MOI conditions were also present in the high MOI set. More strictly, 60 deletions were shared across all lineages in both cell types at a high MOI and six at a low MOI. Additionally, the pattern of deletion coordinates shared between the two host cell types (Figure 10C,D) was similar to the pattern of common deletions among parallel lineages of the same cell type (Figure 9A) and also to the patterns of persistent deletions (Figure 8B). This consistency could be taken as further evidence that not all deletions were random occurrences, but were regulated by similar molecular mechanisms or selective pressures.

4. Discussion

During replication, viruses, especially RNA viruses, spontaneously generate varying versions of their genomes, including hypermutated genomes, insertions, deletions, and reorganizations (5′ and 3′ cb and sb). These unusual genomes cannot complete a whole viral cycle on their own but can play crucial roles in infection outcomes and viral evolution [11,12,14,46]. Understanding the dynamics of DVG formation and accumulation can provide insights into the underlying mechanisms of their appearance and function. To investigate this in betacoronaviruses, we conducted evolution experiments with HCoV-OC43 and MHV.

DVGs exhibited an overall increase in both richness and abundance in all evolving lineages. In HCoV-OC43, DVG richness initially spiked out but later stabilized, suggesting the existence of a diversity ceiling. However, DVG abundance at a high MOI remained consistently high in advanced passages. At a high MOI, HCoV-OC43 DVGs showed higher variability and abundance compared to a low MOI, likely due to potential *trans*-complementation and other sort of interactions between DVGs and helper viruses. Our observations support the hypothesis that at a high MOI, DVGs in HCoV-OC43 were not only generated de novo after each passage but some might also be selected, favoring the accumulation of the fittest DVGs while leading to a decrease in less fit variants. Indeed, at a high MOI, a much larger proportion of persistent DVGs was observed compared to a low MOI. Transmission at a low MOI likely occurred through secondary infections, as uninfected cells remained available after the first round of infection. Additionally, transmission could be facilitated by collective viral infection [47–49] or cell fusion [50–53]. Under these conditions, a DVG could co-infect with a helper virus and be *trans*-complemented for successful propagation, even at an initially low MOI; despite this, foundational effects could be more accused at a low MOI and, in some of the lineages, the viral load was undetectable.

An overall decrease in viral load in most cases could be explained by two alternative hypotheses. Firstly, it could be attributed to the progressive accumulation of DVGs that interfere with the replication of wild-type viruses [6,14,54,55]. This effect would be stronger at a high MOI, which seems not to be a general case in Figure 1. Secondly, it could be a consequence of strong transmission bottlenecks that turn on Muller's ratchet, leading to the fixation of deleterious mutations and, consequently, fitness declines, as is well established for other RNA viruses [56–60]. To further explore the role of DVGs in these dynamics, we performed RNA-seq from the supernatants of infected cells from each lineage at four equidistant evolutionary time points, as well as from the corresponding ancestral viruses, in order to reconstruct the DVG populations at each sample. Significant DVG accumulation was observed in both cell lines with HCoV-OC43; in BHK-21 cells, there was also a general increase with small fluctuations in infectious viral load. The positive trends for both full and defective viruses suggest that an interfering dynamic was not necessarily occurring. These dynamics reinforce the idea of DVGs as triggers and/or reservoirs of viral diversity, providing an adaptive advantage [12,44,61]. However, in HCT-8 cells, viral loads showed an overall decreasing trend. This could be attributed to technical limitations in testing viral loads, as it had to be assessed in BHK-21 cells due to the inability to form plaques in HCT-8 cells. Viral adaptation to HCT-8 cells during the passages might have resulted in fitness tradeoffs in the BHK-21 cells used for plaquing. Further studies are needed to investigate the potential role of DVG–complete virus cooperation at this stage.

In contrast to HCoV-OC43, the richness and abundance of DVGs in MHV were positively correlated and increased throughout evolution, while the overall trend for viral load indicated a decrease, leading to undetectable levels in some MOI lineages. We proposed two non-mutually exclusive hypotheses to explain the observed fluctuations in viral load, one calling for the increasing accumulation of DVGs and another one associating the effect with reductions in the inoculum size and the subsequent onset of Muller's ratchet [56–60]. Our finding of an accumulation of diverse populations of DVGs in evolving lineages, in both viruses but particularly in MHV, gives support to the first possibility, although does not rule out the second hypothesis, especially at a low MOI.

The MHV strain that we used in this experiment was interferon-sensitive. Although MHV has a short infection cycle and should not be notably affected by interferon-based cell defense [62], we cannot completely rule out the possibility that interplay between the virus and the cells' immune response may have impacted the generation and propagation of DVGs, contributing too to the divergent outcomes observed amongst the two betacoronaviruses. The interferon response has been identified as key in infections in which the presence of DVGs impacts the severity of symptoms [50], or in which infection does not resolve but becomes chronic [12,63,64].

The choice of cell lines for the HCoV-OC43 evolution experiments was primarily motivated by their high susceptibility and ability to produce high viral titers, which was essential for virus passaging at a high MOI. As a result, the outcome of viral DVG evolution in both cell lines showed similarities, indicating that they may not have imposed largely different constraints for viral adaptation. In retrospect, the use of more diverse cell lines with varying properties and susceptibilities might have provided additional insights into the impact of host factors on DVG evolution. Nevertheless, the results obtained from the selected cell lines still offered valuable information on DVG dynamics and accumulation in the context of HCoV-OC43 evolution. Future studies could consider employing a broader range of cell lines to gain a more comprehensive understanding of how host factors influence the evolution of DVGs. Indeed, Aguilar Rangel et al. [14] have shown that the Dengue virus (DENV) generates different spectra of DVGs in human and mosquito cells. Likewise, Hasiów-Jaroszewska et al. [65] also described variation in the composition of DVGs of tomato black ring virus in different host species. We could also compare our results for MHV in CCL-9.1 with those obtained in DBT-9 cells by Gribble et al. [44] and identify common hotspots of recombination, as well as differences.

Deletions were the most variable and abundant type of reconstructed DVGs, an observation that is consistent with previous studies with other positive-sense single-stranded RNA viruses: Aguilar Rangel et al. [14], in the cases of poliovirus and DENV; Zhou et al. [50], in the case of SARS-CoV-2; and Jaworski and Routh [66] for Flock House virus. For this reason, we conducted a more detailed analysis of this specific DVG type. The analysis, focusing on deletions longer than three nucleotides, distinguishes patterns in genome sizes that dominated in abundance. Over the course of HCoV-OC43 evolution at a high MOI, these patterns were refined within a specific size range, giving rise to dynamic peaks of high abundance. This dynamic behavior within a certain size range suggests adaptive evolution. In contrast, the accumulation pattern of certain sizes at a low MOI was not prominent, and, notably, MHV did not exhibit a clear pattern of size accumulation. This lack of a discernible pattern may indicate different evolutionary dynamics or constraints under a low MOI. Furthermore, short genomes with deletions (<2 kb) were particularly prevalent in the HCT-8 cell lineage. This observation underscores the potential influence of host-specific factors on the evolution dynamics of viral genomes with deletions.

Furthermore, we observed that the cistrons *E* and *ns12.9* (or its syntenic *accessory protein 4* in MHV) were more impacted by deletions, while *ORF1ab* and *S* were less affected at both MOIs for both viruses. Coronaviruses *E* cistron encodes for a small structural protein involved in multiple stages of the replicative cycle; it participates in viral assembly, virion release, and pathogenesis [67]. In MHV, E is not essential for virus replication and a ΔE mutant showed a lower growth rate and a lower infectious titer than the wild-type virus, but it still was capable of producing viable particles [68]. *E* is highly expressed in infected cells, but only a small part is incorporated into the virion envelope. E also can form homotypic interactions, oligomerizing to generate viroporins, ion-channel proteins associated with pathogenicity. The HCoV-OC43 ns12.9 protein is associated with virus morphogenesis and pathogenesis by viroporin formation. The HCoV-OC43-Δns12.9 was defective in growth in vitro and reduced inflammation and virulence in the brains of infected mice, confirming its influence in virus pathogenesis [69]. Some studies had associated the inhibition of the E viroporins with a reduction in pathogenicity, highlighting its therapeutical value to face coronavirus infections [67], while others point towards different viroporins as potential targets for antiviral drugs [69]. The identification of *E*, *ns12.9*, and *accessory protein 4* as the most prone cistrons for deletions suggests the tantalizing possibility that DVGs in these cistrons may be selected for as a viral strategy to modulate virulence.

Furthermore, the E protein may interact with the host's immune system, implying that deletions in these regions could impact the virus' ability to replicate and evade immune responses. Our evolution experiment took place in cell culture and not in a whole organism with a complete immune system. The independent emergence of deletions of equivalent parts of the genome in different viruses and cell types suggests that the shift from the whole organism to cell culture was a more substantial landmark than the differences among various cell culture types. Additionally, these deletions were accumulated already in the first stages of the evolution, which suggests that the adaptive advantage of these variants was significant. However, the underlying mechanisms could be much more complex, potentially involving specific viral sequences and structures. The analysis of the most frequent sites for BP and RP, along with their surrounding sequences, suggests that the mechanism of DVG generation is not random and is influenced, in part, by the genome sequence but further by the resulting secondary RNA structure.

It is noteworthy that some hotspots of persistent deletions of BPs are in the first third of the viral genome, a region generally assumed to be crucial for viral replication. We identified a site close to the TRS-L as one of the most frequent BPs, suggesting that the RNA structures needed for replication would be intact. We should not rule out the effect that successive deletions have on the increase in frequency in some spots, or a special case of the latter, that sgRNAs are targets of deletions. Unfortunately, the sequencing technology used for this work limits our ability to investigate this further.

We have also paid attention to identifying deletions that consistently appeared over the course of evolution. In the case of HCoV-OC43, some of these persistent deletions likely represent instances where deletions appeared de novo multiple times in the same lineage, because our results also indicated that the process of DVG generation was not completely random. Notably, the pattern of coordinates for persistent deletions highlighted the existence of convergent deletions between parallel lineages and even between different cell types. This suggests that a subset of deletions arose independently of the host and lineage. Another portion of pervasive but not de novo generated deletions was likely composed of deletions that were transmitted along passages. When comparing samples at high vs. low MOIs, we observed more persistent deletions under high MOI conditions, suggesting that the virus, given the opportunity for trans-complementation, exhibits an increased tendency for supporting persistent deletions. In the case of MHV, most DVGs were not transmitted but de novo generated within each passage, not being under specific selection pressure, consistent with the less defined and more variable size distribution. This suggests, in general, that the dynamics of DVG accumulation differ among two closely related betacoronavirus species, with potentially distinct mechanisms governing their appearance, abundance, and fitness [70,71].

Furthermore, our findings indicated that the dynamics of persistent deletions in HCoV-OC43 at a high MOI maintained hotspots of deletion coordinates, becoming more abundant with passages. Certain genomic regions seemed more susceptible to deletions than others [24], with these hotspots showing an increased abundance during experimental evolution at a high MOI. In both viruses, regions susceptible to deletions were detected, and the abundance of deletions in these hotspots increased with the progress of experimental evolution at a high MOI. Despite a larger genome size being a disadvantage for replication competition, longer DVGs lacking structural proteins may have the advantage of being more readily complemented in trans and transmitted. Zhou et al. [50] also found hotspots of BPs and RPs in the genome of SARS-CoV-2 in samples from patients, which suggest that our in vitro results might reflect relevant processes in vivo.

The dynamics of deletions in parallel lineages showed a remarkable degree of convergence, despite the majority of shared deletions between the lineages were very short, potentially classified as sequencing errors [37]. After removing these short ones, we still found a number of DVGs of various sizes shared among lineages evolved in the same host. This finding prompted distinct patterns of deletion evolution within different cell types, combining stochastic generation of deletions within each passage and the potential persistence of positively selected deletions shared by different lineages. The observation that some deletions can be positively selected to modulate virus replication and pathogenesis, for instance, by altering coding reading frames and generating novel proteins [14], opens tantalizing possibilities for the development of DVGs with antiviral activity. In general, short DIPs, and not long DVG, have been considered for their interfering activity over wild-type viruses [72], although results by Levi et al. [73] suggest that long DVGs can be effective and broad-spectrum antiviral for alphaviruses.

5. Conclusions

In conclusion, the buildup of DVGs seems to be governed by dynamic processes that set a cap on the diversity of DVGs capable of accumulating in each sample. This restriction is affected by both the passage number and, in specific instances, the MOI. Throughout evolution at a high MOI, a selective process appears to take place, giving preference to certain DVGs for accumulation while others diminish over successive passages. This observation implies that DVGs possess the ability to transmit under high MOI conditions in a virus- and cell-type-dependent manner, resulting in their enduring presence and potential influence on viral evolution.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v16040644/s1, Figure S1: Percentage of sgRNAs over the total

number of deletions observed per sample. Thick lines are only shown for illustrating the average trend over time; Figure S2: Distribution of deletion type DVGs at a low MOI. To facilitate visualization only deletions bigger than 50 nucleotides are shown. The number represent the exact length of the DVG and in parenthesis the lineages where they were founded; Table S1: Statistics describing the variation in MOI for each experimental evolution lineage. Median values $\pm$ IQR are shown. Last column shows the fold difference between high and low MOIs (averaging lineages); Table S2: Fitting of viral load data to the best-fitting ARIMA(p, d, q) model based on the minimum BIC criterium. Errors represent ± 1 SE.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Excel files with viral titers and dilution schemes are available at https://git.csic.es/sfelenalab/dvg-accumulation-in-betacoronaviruses-experimental-evolution (accessed on 17 April 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Duffy, S.; Shackelton, L.A.; Holmes, E.C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* **2008**, *9*, 267–276. [CrossRef] [PubMed]
2. Sanjuán, R.; Nebot, M.R.; Chirico, N.; Mansky, L.M.; Belshaw, R. Viral mutation rates. *J. Virol.* **2010**, *84*, 9733–9748. [CrossRef] [PubMed]
3. Belshaw, R.; Sanjuán, R.; Pybus, O.G. Viral mutation and substitution: Units and levels. *Curr. Opin. Virol.* **2011**, *1*, 430–435. [CrossRef] [PubMed]
4. Stampfer, M.; Baltimore, D.; Huang, A.S. Absence of interference during high-multiplicity infection by clonally purified vesicular stomatitis virus. *J. Virol.* **1971**, *7*, 409–411. [CrossRef]
5. Huang, A.S. Defective interfering viruses. *Annu. Rev. Microbiol.* **1973**, *27*, 101–118. [CrossRef] [PubMed]
6. Felt, S.A.; Achouri, E.; Faber, S.R.; López, C.B. Accumulation of copy-back viral genomes during respiratory syncytial virus infection is preceded by diversification of the copy-back viral genome population followed by selection. *Virus Evol.* **2022**, *8*, veac091. [CrossRef]
7. Barrett, A.D.T.; Dimmock, N.J. Modulation of Semliki Forest virus-induced infection of mice by defective-interfering virus. *J. Infect. Dis.* **1984**, *150*, 98–104. [CrossRef] [PubMed]
8. Roux, L.; Simon, A.E.; Holland, J.J. Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. *Adv. Virus Res.* **1991**, *40*, 181–211. [PubMed]
9. Smith, C.M.; Scott, P.D.; O'Callaghan, C.; Easton, A.J.; Dimmock, N.J. Defective interfering influenza RNA inhibits infectious influenza virus replication in human respiratory tract cells: A potential new human antiviral. *Viruses* **2016**, *8*, 237. [CrossRef]
10. Huang, A.S.; Baltimore, D. Defective viral particles and viral disease processes. *Nature* **1970**, *226*, 325–327. [CrossRef]
11. Rezelj, V.V.; Levi, L.I.; Vignuzzi, M. The defective component of viral populations. *Curr. Opin. Virol.* **2018**, *33*, 74–80. [CrossRef] [PubMed]

12. Vignuzzi, M.; López, C. Defective viral genomes are key drivers of the virus-host interaction. *Nat. Microbiol.* **2019**, *4*, 1075–1087. [CrossRef] [PubMed]
13. Di Gioacchino, A.; Legendre, R.; Rahou, Y.; Najburg, V.; Charneu, P.; Greenbaum, B.D.; Tangy, F.; van der Werf, S.; Cocco, S.; Komarova, A.V. sgDI-tector: Defective interfering viral genome bioinformatics for detection of coronavirus subgenomic RNAs. *RNA* **2022**, *28*, 277–289. [CrossRef]
14. Aguilar Rangel, M.; Dolan, P.T.; Taguwa, S.; Xiao, Y.; Andino, R.; Frydman, J. High-resolution mapping reveals the mechanism and contribution of genome insertions and deletions to RNA virus evolution. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2304667120. [CrossRef]
15. Jennings, P.A.; Finch, J.T.; Robertson, J.S. Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* **1983**, *34*, 619–627. [CrossRef] [PubMed]
16. Saira, K.; Lin, X.; DePasse, J.V.; Halpin, R.; Twaddle, A.; Stockwell, T.; Angus, B.; Cozzi-Lepri, A.; Delfino, M.; Dugan, V.; et al. Sequence analysis of in vivo defective interfering-like RNA of influenza a H1N1 pandemic virus. *J. Virol.* **2013**, *87*, 8064–8074. [CrossRef]
17. Poirier, E.Z.; Mounce, B.C.; Rozen-Gagnon, K.; Hooikaas, P.J.; Stapleford, K.A.; Moratorio, G.; Vignuzzi, M. Low-fidelity polymerases of alphaviruses recombine at higher rates to overproduce defective interfering particles. *J. Virol.* **2016**, *90*, 2446–2454. [CrossRef]
18. Li, T.; Pattnaik, A.K. Replication signals in the genome of vesicular stomatitis virus and its defective interfering particles: Identification of a sequence element that enhances DI RNA replication. *Virology* **1997**, *232*, 248–259. [CrossRef] [PubMed]
19. Sun, Y.; Kim, E.J.; Felt, S.A.; Taylor, L.J.; Agarwal, D.; Grant, G.R.; López, C.B. A specific sequence in the genome of respiratory syncytial virus regulates the generation of copy-back defective viral genomes. *PLoS Pathog.* **2019**, *15*, e1007707.
20. Brian, D.A.; Baric, R.S. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **2005**, *287*, 1–30.
21. Yang, D.; Leibowitz, J.L. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus Res.* **2015**, *206*, 120–133. [CrossRef] [PubMed]
22. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The architecture of SARS-CoV-2 transcriptome. *Cell* **2020**, *181*, 914–921. [CrossRef]
23. Repass, J.F.; Makino, S. Importance of the positive-strand RNA secondary structure of a murine coronavirus defective interfering RNA internal replication signal in positive-strand RNA synthesis. *J. Virol.* **1988**, *72*, 7926–7933. [CrossRef] [PubMed]
24. Lin, Y.J.; Lai, M.M. Deletion mapping of a mouse hepatitis virus defective interfering RNA reveals the requirement of an internal and discontiguous sequence for replication. *J. Virol.* **1993**, *67*, 6110–6118. [CrossRef] [PubMed]
25. Liu, Q.; Johnson, R.F.; Leibowitz, J.L. Secondary structural elements within the 3′ untranslated region of mouse hepatitis virus strain JHM genomic RNA. *J. Virol.* **2001**, *75*, 12105–12113. [CrossRef] [PubMed]
26. De Haan, C.A.M.; Volders, H.; Koetzner, C.A.; Masters, P.S.; Rottier, P.J.M. Coronaviruses maintain viability despite dramatic rearrangements of the strictly conserved genome organization. *J. Virol.* **2002**, *76*, 12491–12502. [CrossRef] [PubMed]
27. Goebel, S.J.; Taylor, J.; Masters, P.S. The 3′ cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J. Virol.* **2004**, *78*, 7846–7851. [CrossRef] [PubMed]
28. Liu, P.; Li, L.; Millership, J.J.; Kang, H.; Leibowitz, J.L.; Giedroc, D.P. A U-turn motif-containing stem-loop in the coronavirus 5′ untranslated region plays a functional role in replication. *RNA* **2007**, *13*, 763–780. [CrossRef] [PubMed]
29. Sethna, P.B.; Hung, S.L.; Brian, D.A. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons'. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5626–5630. [CrossRef]
30. Hofmann, M.A.; Sethna, P.B.; Brian, D.A. Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. *J. Virol.* **1990**, *64*, 4108–4114. [CrossRef]
31. Van der Most, R.G.; Bredenbeek, P.J.; Spaan, W.J.M. Domain at the 3′ end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. *J. Virol.* **1991**, *65*, 3219–3226. [CrossRef] [PubMed]
32. Fosmire, J.A.; Hwang, K.; Makino, S. Identification and characterization of a coronavirus packaging signal. *J. Virol.* **1992**, *66*, 3522–3530. [CrossRef] [PubMed]
33. Zhao, X.; Shaw, K.; Cavanagh, D. Presence of subgenomic mRNAs in virions of coronavirus IBV. *Virology* **1993**, *196*, 172–178. [CrossRef] [PubMed]
34. Reed, L.J.; Muench, H. A simple method of estimating fifty per cent endpoints. *Am. J. Epidemiol.* **1938**, *27*, 493–497. [CrossRef]
35. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**. [CrossRef]
36. Olmo-Uceda, M.J.; Muñoz-Sánchez, J.C.; Lasso-Giraldo, W.; Arnau, V.; Díaz-Villanueva, W.; Elena, S.F. DVGfinder: A metasearch tool for identifying defective viral genomes in RNA-Seq data. *Viruses* **2022**, *14*, 1114. [CrossRef] [PubMed]
37. Schirmer, M.; D'Amore, R.; Ijaz, U.Z.; Hall, N.; Quince, C. Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinform.* **2016**, *17*, 215. [CrossRef] [PubMed]
38. VanderPlas, J.T. Understanding the Lomb-Scargle periodogram. *Astrophys. J. Suppl. Ser.* **2018**, *236*, 16. [CrossRef]
39. Pagès, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. Biostrings: Efficient Manipulation of Biological Strings. R Package Version 2.70.1. 2024. Available online: https://bioconductor.org/packages/Biostrings (accessed on 20 March 2024).
40. Lorenz, R.; Hofacker, I.L.; Stadler, P.F. RNA folding with hard and soft constraints. *Algorithms Mol. Biol.* **2016**, *11*, 8. [CrossRef]
41. Johnson, P.Z.; Kasprzak, W.K.; Shapiro, B.A.; Simon, A.E. RNA2Drawer: Geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. *RNA Biol.* **2019**, *16*, 1667–1671. [CrossRef]

42. Lawrence, M.; Huber, W.; Pagès, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.T.; Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **2013**, *9*, e1003118. [CrossRef] [PubMed]

43. Elshina, E.; te Velthuis, A.J.W. The influenza virus RNA polymerase as an innate immune agonist and antagonist. *Cell. Mol. Life Sci.* **2021**, *78*, 7237–7256. [CrossRef] [PubMed]

44. Gribble, J.; Stevens, L.J.; Agostini, M.L.; Anderson-Daniels, J.; Chappell, J.D.; Lu, X.; Pruijssers, A.J.; Routh, A.L.; Denison, M.R. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **2021**, *17*, e1009226. [CrossRef] [PubMed]

45. Zuker, M. On finding all suboptimal foldings of an RNA molecule. *Science* **1989**, *244*, 48–52. [CrossRef] [PubMed]

46. Elena, S.F. The role of indels in evolution and pathogenicity of RNA viruses. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2310785120. [CrossRef] [PubMed]

47. Maeda, J.; Maeda, A.; Makino, S. Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells. *Virology* **1999**, *263*, 265–272. [CrossRef]

48. Cuevas, J.M.; Durán-Moreno, M.; Sanjuán, R. Multi-virion infectious units arise from free viral particles in an enveloped virus. *Nat. Microbiol.* **2017**, *2*, 17078. [CrossRef]

49. Andreu-Moreno, I.; Sanjuán, R. Collective infection of cells by viral aggregates promotes early viral proliferation and reveals a cellular-level Allee effect. *Curr. Biol.* **2018**, *28*, 3212–3219. [CrossRef]

50. Zhou, T.; Gilliam, N.J.; Li, S.; Spandau, S.; Osborn, R.M.; Connor, S.; Anderson, C.S.; Mariani, T.J.; Thakar, J.; Dewhurst, S.; et al. Generation and functional analysis of defective viral genomes during SARS-CoV-2 infection. *mBio* **2023**, *14*, e00250-23. [CrossRef]

51. Rosell, A.; Havervall, S.; von Meijenfeldt, F.; Hisada, Y.; Aguilera, K.; Grover, S.P.; Lisman, T.; Mackman, N.; Thålin, C. Patients with COVID-19 have elevated levels of circulating extracellular vesicle tissue factor activity that is associated with severity and mortality. *Arterioscler. Thromb. Vasc. Biol.* **2021**, *41*, 878–882. [CrossRef]

52. Zavan, B.; Crupi, M.J.F.; Ilkow, C.S.; Bell, J.C.; Boulton, S. Extracellular vesicles and viruses: Two intertwined entities. *Int. J. Mol. Sci.* **2023**, *24*, 1036. [CrossRef] [PubMed]

53. Buchrieser, J.; Dufloo, J.; Hubert, M.; Monel, B.; Planas, D.; Rajah, M.M.; Planchais, C.; Porrot, F.; Guivel-Benhassine, F.; van der Werf, S.; et al. Syncytia formation by SARS-CoV-2-infected cells. *EMBO J.* **2020**, *39*, e106267. [CrossRef] [PubMed]

54. Leroy, H.; Han, M.; Woottum, M.; Bracq, L.; Bouchet, J.; Xie, M.; Benichou, S. Virus-mediated cell-cell fusion. *Int. J. Mol. Sci.* **2020**, *21*, 9644. [CrossRef] [PubMed]

55. Rajah, M.M.; Bernier, A.; Buchrieser, J.; Schwartz, O. The mechanism and consequences of SARS-CoV-2 spike-mediated fusion and syncytia formation. *J. Mol. Biol.* **2022**, *434*, 167280. [CrossRef] [PubMed]

56. Clarke, D.K.; Duarte, E.A.; Moya, A.; Elena, S.F.; Domingo, E.; Holland, J.J. Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* **1993**, *67*, 222–228. [CrossRef] [PubMed]

57. Duarte, E.A.; Clarke, D.K.; Moya, A.; Elena, S.F.; Domingo, E.; Holland, J.J. Many-trillionfold amplification of single RNA virus particles fails to overcome the Muller's ratchet effect. *J. Virol.* **1993**, *67*, 3620–3623. [CrossRef] [PubMed]

58. Novella, I.S.; Elena, S.F.; Moya, A.; Domingo, E.; Holland, J.J. Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *J. Virol.* **1995**, *69*, 2869–2872. [CrossRef]

59. Elena, S.F.; González-Candelas, F.; Novella, I.S.; Duarte, E.A.; Clarke, D.K.; Domingo, E.; Holland, J.J.; Moya, A. Evolution of fitness in experimental populations of vesicular stomatitis virus. *Genetics* **1996**, *142*, 673–679.

60. De la Iglesia, F.; Elena, S.F. Fitness declines in tobacco etch virus upon serial bottleneck transfers. *J. Virol.* **2008**, *81*, 4941–4947. [CrossRef]

61. Lin, C.H.; Tam, H.M.H.; Yang, C.Y.; Hsieh, F.C.; Wang, J.L.; Yang, C.C.; Hsu, H.W.; Liu, H.P.; Wu, H.Y. Evolution of the coronavirus spike protein in the full-length genome and defective viral genome under diverse selection pressures. *J. Gen. Virol.* **2023**, *104*, 001920. [CrossRef]

62. Kyuwa, S.; Sugiura, Y. Role of cytotoxic T lymphocytes and interferon-γ in coronavirus infection: Lessons from murine coronavirus infections in mice. *J. Vet. Med. Sci.* **2020**, *82*, 1410–1414. [CrossRef] [PubMed]

63. Mura, M.; Combredet, C.; Najburg, V.; Sánchez David, R.Y.; Tangy, F.; Komarova, A.V. Nonencapsidated 5′ copy-back defective interfering genomes produced by recombinant measles viruses are recognized by RIG-I and LGP2 but not MDA5. *J. Virol.* **2017**, *91*, e00643-17. [CrossRef]

64. Wignall-Fleming, E.B.; Vasou, A.; Young, D.; Short, J.A.L.; Hughes, D.J.; Goodbourn, S.; Randall, R.E. Innate intracellular antiviral responses restrict the amplification of defective virus genomes of parainfluenza virus 5′. *J. Virol.* **2020**, *94*, e00246-20. [CrossRef] [PubMed]

65. Hasiów-Jaroszewska, B.; Minicka, J.; Zarzyńska-Nowak, A.; Budzyńska, D.; Elena, S.F. Defective RNA particles derived from tomato black ring virus genome interfere with the replication of parental virus. *Virus Res.* **2018**, *250*, 87–94. [CrossRef] [PubMed]

66. Jaworski, E.; Routh, A. Parallel ClickSeq and nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS Pathog.* **2017**, *13*, e1006365. [CrossRef] [PubMed]

67. Schoeman, D.; Fielding, B.C. Coronavirus envelope protein: Current knowledge. *Virol. J.* **2019**, *16*, 69. [CrossRef] [PubMed]

68. Kuo, L.; Masters, P.S. The small envelope protein E is not essential for murine coronavirus replication. *J. Virol.* **2003**, *77*, 4597–4608. [CrossRef] [PubMed]

69. Zhang, R.; Wang, K.; Ping, X.; Yu, W.; Qian, Z.; Xiong, S.; Sun, B. The ns12.9 accessory protein of human coronavirus OC43 is a viroporin involved in virion morphogenesis and pathogenesis. *J. Virol.* **2015**, *89*, 11383–11395. [CrossRef] [PubMed]

70. De Groot, R.J.; van der Most, R.G.; Spaan, W.J.M. The fitness of defective interfering murine coronavirus DI-a and its derivatives is decreased by nonsense and frameshift mutations. *J. Virol.* **1992**, *66*, 5898–5905. [CrossRef]
71. Brian, D.A.; Spaan, W.J.M. Recombination and coronavirus defective interfering RNAs. *Semmin. Virol.* **1997**, *8*, 101–111. [CrossRef]
72. Notton, T.; Sardanyés, J.; Weinberger, A.D.; Weinberger, L.S. The case of transmissible antivirals to control population-wide infectious disease. *Trends Biotechnol.* **2014**, *32*, 400–405. [CrossRef] [PubMed]
73. Levi, L.I.; Rezelj, V.V.; Henrion-Lacritick, A.; Erazo, D.; Boussier, J.; Vallet, T.; Bernhauerová, V.; Suzuki, Y.; Carrau, L.; Weger-Lucarelli, J.; et al. Defective viral genomes from chikungunya virus are broad-spectrum antivirals and prevent virus dissemination in mosquitoes. *PLoS Pathog.* **2021**, *17*, e1009110. [CrossRef] [PubMed]