

SARS-CoV-2 evasion from ADAR hyper-editing is both genome-encoded and sustained by the virus replication strategy

Enrico Bortoletto

University of Padua

Paola Venier

University of Padua <https://orcid.org/0000-0001-9634-3662>

Maxwell Shapiro

Stony Brook University

Andrea Leonardi

University of Padua <https://orcid.org/0000-0002-7246-8580>

Umberto Rosani (✉ umberto.rosani@unipd.it)

University of Padua

Brief Communication

Keywords: SARS-CoV-2, ADAR hyperediting

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-314516/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

SARS-CoV-2 is threatening the human society because of its capability to subvert antiviral defenses, causing cytokine hyper-activation and prolonged damage in multiple tissues with unpredictable outcomes in the mid-long term. Here, we evaluated the role of ADAR, an interferon-stimulated gene able to control the activation of the immune system and to directly modify exogenous dsRNAs, during *in-vivo* infection by SARS-CoV-2. After accurate analysis of 863 RNA-seq samples from different species, we identified ADAR-mediated hyper-editing of SARS-CoV-2 only in 49 human datasets at a low level (0.036‰ hyper edited reads on average) and preferentially on ORF6. Conversely, in mouse datasets we found abundant hyper-editing of viral reads (up to 1.16‰). The analysis of dinucleotide frequency along the ORFs of α , β , δ and γ coronaviruses highlighted the evolutionary pressure of ADAR enzymes, suggesting that the SARS-CoV-2 resistance to hyper-editing is both genome-encoded and supported by the viral transcription strategy.

Introduction

In the last two decades, coronaviruses have caused pandemics in multiple hosts worldwide. In 2003, Severe acute respiratory syndrome-related virus (SARS-CoV) killed approximately 774 people mostly in China (fatality rate of 9.6%)¹. In 2012, Middle East respiratory syndrome-related virus (MERS-CoV) killed approximately 800 people primarily in the middle east (fatality rate of 35%)² and, in 2018, Swine acute diarrhea syndrome coronavirus (SADS-CoV) killed more than 24,000 piglets in China³. Since the end of 2019, Severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) has rapidly spread worldwide, making COVID-19 the worst zoonotic disease of the modern history. Although considerably similar to coronaviruses of bats and pangolins⁴, SARS-CoV-2 has jumped from a still unrecognized host into humans, with the cross-species transmission mediated by an efficient interaction with human cell receptors⁵.

SARS-CoV-2 is a positive-sense, single stranded RNA virus with a genome length of about 30 kb. The replication of positive-strand RNA viruses involves the assembly of replication-transcription complexes (RTCs)⁶. Such RTCs are composed by a network of interconnected double membrane vesicles (DMV) derived from endoplasmic reticulum⁶⁻⁸. The role of these vesicles is not fully understood but, certainly, they provide a protected niche for preventing the exposure of viral replication intermediates, such as negative-stranded templates and dsRNAs, to cytosolic innate immune sensors, nucleases and, in general, antiviral host counteractions⁹⁻¹¹.

The impressive amount of whole-genome data reported for this virus helps tracing its micro-evolutionary dynamics^{12,13}, whereas dual RNA sequencing (RNA-seq) supports the identification of molecular host-virus interactions during infection¹⁴. Although the SARS-CoV-2 mutation rate is slower than other RNA viruses because of RNA proofreading activity¹⁵, genomic data produced in real time have shown a strong bias towards C-to-T substitutions, possibly resulting from APOBEC-mediated editing^{16,17}. SARS-CoV-2

mutations can crucially affect host-pathogen interactions by modifying the targets of RNA editing enzymes, the secondary structure of viral RNAs¹⁸ and stimulating innate immune responses via TLR7 and TLR8¹⁹. RNA sequencing (RNA-seq) and biochemical analyses revealed an extraordinary activation of the immune system during SARS-CoV-2 infection, known as “cytokine storm”, which has contributed to the SARS-CoV-2-induced mortality⁵.

The footprint of APOBEC3 and ADAR deamination has been noted during SARS-CoV-2 infection *in-vivo*¹⁷. APOBEC and ADAR seem to deaminate in preferential contexts: APOBEC deaminates primarily TC or CC motifs in ssDNA, leading to TT or CT variations, whereas ADAR deaminates adenosine (A) to inosine (I) primarily in the context of WA (W=A/T) motifs in dsRNA, with the resulting inosine being read as guanine during translation. Depending on the host-virus combination, A-to-I hyper-editing (*he*) can either limit the virus replication by impairing viral dsRNAs and marking them for a rapid degradation²⁰ or favor the virus by acting as a negative-feedback in the host immune system, i.e. reducing the dsRNA detection mediated by the RIG-I-like receptor dsRNA helicase Melanoma Differentiation-Associated protein 5 (MDA5)²¹.

A previous analysis of some SARS-CoV-2 RNA-seq datasets of different origin outlined a low frequency of ADAR editing, and raises relevant questions about the ADAR role during infection²². In details: to what extent is SARS-CoV-2 susceptible to ADAR-mediated editing? Is the ADAR activity, in the form of *he*, traceable during SARS-CoV-2 infection? Here, we verified the *in-vivo* levels of ADAR *he* in 863 RNA-seq datasets from humans and other SARS-CoV-2 infection models. Analyzing the dinucleotide composition of the open-reading frames (ORFs) of SARS-CoV-2 and other coronaviruses, we have traced the evolutionary footprints of RNA editing enzymes and we estimated the current susceptibility of these coronaviruses to RNA editing.

Results And Discussion

We searched for SARS-CoV-2 transcription in 636 human-SARS-CoV-2 RNA-seq datasets, selecting the 328 of them with at least 1,000 SARS-CoV-2 reads (S. Table 1). We applied the *hyperediting* tool developed by E. Y. Levanon et al.²³ on these human datasets to trace genuine ADAR hyper-editing (*he*) and thus, we retrieved 14,273 SARS-CoV-2 *he* reads displaying 0.036 edited reads every thousand viral reads on average (S. Table 1). Along the SARS-CoV-2 genome sequence, *he* sites could be identified only by lowering the fraction of *minimum edited sites per read* to 0.03 instead of 0.05 and the resulting *he* levels poorly correlated with the coverage of SARS-CoV-2 in the same analyzed samples (Figure 1A, $r=0.42$, p -values 1.3×10^{-8}). Notably, only 49 of the analyzed datasets included more than 100 hyper-edited SARS-CoV-2 reads (total 9,568 *he* reads), with the *he* sites mostly clustering in two conserved hotspots along the SARS-CoV-2 genome: one localized on the RNA-dependent RNA polymerase (nsp12, position 14221:14331) and the second on nsp6 (position 11058:11162), both encoded within the polycistronic ORF1ab (Figure 1D). Normalizing *he* by gene expression levels, we demonstrated that *he* mostly impacted ORF6 (Figure 1C), whose encoded product is known to impair the induced transcription of ISGs by interacting with STAT1 and STAT2²⁴.

The analysis of additional 227 RNA-seq datasets referred to SARS-CoV-2 infection of non-human hosts, namely hamster, ferret, non-human primates, and mice, revealed relevant *he* levels exclusively in mouse datasets from one specific experiment (PRJNA646535, S. Table 1). SARS-CoV-2 infection in these mice was made possible by transfection of human Angiotensin I converting enzyme 2 (hACE2), while the role of Interferon signaling during infection was evaluated by knocking-out Interferon receptor (IFNR) or Interferon Regulatory Factor 3/7 (IRF3/7)²⁵. Notably, the IRF3/7 knock-out mice displayed the highest *he* levels (1.16‰), followed by control (0.26‰) and IFNR knock-out mice (0.11‰). In the mouse datasets we detected a higher hyper-editing efficiency (measured as the number of edited bases per *he* cluster) on SARS-CoV-2 RNAs than in human datasets; in addition, the normalized *he* levels impacted SARS-CoV-2 genes homogenously (Figure 1C and D).

To verify if SARS-CoV-2 could in some way interfere with the ADAR *he* of host dsRNAs, we systematically traced *he* events on both the mouse and human genomes. In mouse, ADAR *he* impacted preferentially non-coding genes, although the detected *he* levels were not influenced by SARS-CoV-2 infection for either protein coding genes or non-coding elements (S. Figure 1A). Although ADAR expression levels were low (<3 TPMs) and mildly downregulated in the knock-out mice, we observed a reduced *he*, in agreement to the reduced expression of interferon-related genes, which also include ADAR (S. Figure 1A). The efficiency of *he* was similar between SARS-CoV-2 and mouse RNAs in the genetic knock-out background (Figure 1B). Concerning the human datasets, the host reads have been unfortunately removed from most of the datasets displaying SARS-CoV-2 *he*. In 7 accessible human samples, we could show a higher *he* efficiency on host RNAs than on SARS-CoV-2, with an average of 5 edited bases per cluster compared to 3.5 (Figure 1B).

To further investigate the possible implications of ADAR *he* during SARS-CoV-2 infection in humans, we analyzed post-mortem RNA-seq data from patients differing in the SARS-CoV-2 infection level (N=57)²⁶, demonstrating that *he* of host dsRNAs was not influenced by SARS-CoV-2 infection (S. Figure 1B). In the same human samples, like in the mouse samples, the ADAR expression ranged from mid to low levels (<20 TPM) and was only mildly modulated by SARS-CoV-2 (fold change=1.8).

Given these results we hypothesized that the SARS-CoV-2 genome sequence could confer resistance to ADAR *he* and therefore we investigated the dinucleotide composition of the ORFs belonging to 70 coronaviruses, demonstrating a significant under-representation of the 'WA' (W=A/T), 'TC' and 'CG' motifs (S. Figure 2). While the 'CG' motif is known to be under-represented in coronaviruses, in agreement with the effect of the zinc finger antiviral protein (ZAP) and the low-CpG frequency characterizing vertebrate hosts²⁷, the 'WA' and 'TC' motifs are preferential targets for ADAR and APOBEC3 enzymes, respectively^{28,29}. No significant differences in the inter- and intra-genus under-representation were detectable for 'WA' and 'TC', except between alpha and delta-coronaviruses at 'WA' (p-value 0.0032, Figure 2, S. Table 2), with WA under-representation in ORFs 1ab and S being detected in 96% and 81% of the tested coronavirus genomes, respectively. The second metric that we considered was "the replacement transition fraction", or *repTrFrac*, which determines a significantly high mutation susceptibility leading to non-synonymous polymorphism (nsSNPs) in an ORF. Applying this metric, we showed that *repTrFrac* in

'WA' was significantly higher than in TC for most of the coronavirus genomes ('WA': $78.8 \pm 18\%$; 'TC': $31 \pm 34.6\%$, p -value 2.79×10^{-11} , Figure 2).

Among beta-coronaviruses, SARS-CoV-2 locates in the upper part of the distribution for 'WA' motifs, because of a significant WA under-representation in the ORFs covering 95% of the genome (ORF1ab, S, ORF3a, M, ORF6, ORF7a, N), whereas SARS-CoV-2 is in the lower part of the distribution for 'TC' motifs (ORF1ab and S, Figure 2). We also showed that most of the non-structural proteins encoded in the Orf1ab displayed 'WA' under-representation, except for nsp7, nsp9-12 and nsp16 (Figure 1D).

Conclusions

Following an accurate analysis of 863 well-selected RNA-seq datasets, we showed how ADAR and APOBEC have evolutionary contributed to minimize the RNA editing targets in the extant coronavirus genomes. We assume that ADAR has massively directed genome evolution towards less editable targets, with just few spaces left for additional synonymous variations, whereas APOBEC did leave room for additional synonymous variations. Our hypothesis is consistent with the strong bias towards C-to-T variations, traced during the SARS-CoV-2 pandemic and likely produced by APOBEC^{16,17}. An intrinsic resistance of the SARS-CoV-2 genome to ADAR *he* could explain the low frequency of *he* sites in the analyzed SARS-CoV-2 RNA reads, although the evidence of abundant *he* in mouse samples rather suggests that host-specific SARS-CoV-2 replication mechanisms contribute to reduce ADAR *he*. Actually, SARS-CoV-2 RNAs are protected in double membrane vesicles in humans³⁰, effectively masking the potential ADAR editing substrates (in mice the formation of such vesicles has not been investigated yet). Unmodified *he* levels towards human dsRNAs allow to discharge the hypothesis of a direct interaction between ADAR and SARS-CoV-2, according to the absence of ADAR in the viral interactome³¹. In the light of current knowledge, SARS-CoV-2 genotypes circulated during the pandemic and analyzed in this work most likely escape the pressure of ADAR *he* because of a combination of genome-encoded resistance and protected virus replication mechanisms in humans.

Materials And Methods

Data retrieving. Genome sequences of 70 coronaviruses were downloaded from the NCBI genome database and parsed to extract 617 open reading frames (ORFs, S. Table 2). A total of 1,792 RNA-seq datasets of SARS-CoV-2 were retrieved from NCBI SRA archive (accessed 1st of December 2020) and screened as follows: RNA-seq metadata were used to extract samples of *in-vivo* infection in different host, while the number of SARS-CoV-2 reads per sample was assessed by mapping the reads to the MN908947 reference genome using *bwa* (github.com/lh3/bwa). Samples with at least 1,000 SARS-CoV-2 reads were considered positive.

ADAR hyper-editing analysis. The *hyperediting* tool²³ was applied after minimal modifications of the original version, which utilized *bwa*, SAMtools (github.com/samtools) and BEDTools (github.com/arq5x/bedtools2), implemented to overcome software incompatibilities. The tool parameters

were adapted to our model, applying: 3/5 for *Minimum of edited sites at Ultra-Edit read (%)*; 60 for *Minimum fraction of edit sites/mismatched sites (%)*; 25 for *Minimum sequence quality for counting editing event (PHRED)*; 60 for *Maximum fraction of same letter in cluster (%)*; 20 *Minimum of cluster length (%)*; and imposing that the *he* clusters should not be completely included in the first or last 20% of the read. Outputs in BED format were parsed using custom scripts, and further analyzed using CLC Genomic Workbench v.21 (Qiagen, US).

Gene expression analysis and *he* normalization. Quality-trimmed reads were mapped to the SARS-CoV-2 reference genome (MN908947) applying 0.8 and 0.8 for length and similarity parameters, respectively. Gene expression values were computed as Transcript Per Million (TPM) or as uniquely mapped reads, to normalize *he* levels.

Under-representation analysis. Under-representation and replacement transition fraction analysis were performed using the n3 module of the *Cytidine Deaminase Representation Reporter* (CDUR)³². Briefly, this reporter received as input a coding sequence, which was shuffled 1,000 times by switching nucleotides in the third positions of the codons while maintaining the integrity of the amino-acid sequence as well as the genome GC content. We measured the relevant statistics (e.g., *belowTA* and *repTrFracTA*) as follows. The “*below*” metrics counted the number of hotspots (e.g., TA) in the input and compared this number to the distribution of hotspots observed in the shuffled sequences to obtain an empirical P-value. The replacement transition fraction, or *repTrFrac*, compared the ratio of possible non-synonymous mutations that can occur at the hotspot (e.g., TA) to the observed number of hotspots, obtaining a P-value in a similar way. This fraction was compared to the distribution resulting from the shuffled sequences, to obtain a second empirical P-value.

Statistical analysis. Since the data were not normally distributed (Shapiro-Wilk test), we applied the Kruskal-Wallis test and the Mann–Whitney U test for analyzing the specific sample pairs for stochastic dominance. To examine the correlation between the number of viral and edited reads we calculated the Spearman’s rank correlation coefficient. All the statistical analyses were performed using R (version 4.0.3)³³.

References

1. Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y. & Yuen, K. Y. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin. Microbiol. Rev.***20**, 660–694 (2007).
2. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). in *StatPearls* (StatPearls Publishing, 2021).
3. Zhou, P. *et al.* Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature***556**, 255–258 (2018).

4. Lam, T. T.-Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature***583**, 282–285 (2020).
5. Song, P., Li, W., Xie, J., Hou, Y. & You, C. Cytokine storm induced by SARS-CoV-2. *Clin. Chim. Acta Int. J. Clin. Chem.***509**, 280–287 (2020).
6. Knoops, K. *et al.* SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLOS Biol.***6**, e226 (2008).
7. Snijder, E. J. *et al.* A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLOS Biol.***18**, e3000715 (2020).
8. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.***19**, 155–170 (2021).
9. Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun.***11**, 5885 (2020).
10. Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science***369**, 1395–1398 (2020).
11. de Breyne, S. *et al.* Translational control of coronaviruses. *Nucleic Acids Res.***48**, 12502–12522 (2020).
12. Wang, R., Hozumi, Y., Zheng, Y.-H., Yin, C. & Wei, G.-W. Host Immune Response Driving SARS-CoV-2 Evolution. *Viruses***12**, (2020).
13. Zhu, Z. *et al.* Rapid Spread of Mutant Alleles in Worldwide SARS-CoV-2 Strains Revealed by Genome-Wide Single Nucleotide Polymorphism and Variation Analysis. *Genome Biol. Evol.***13**, (2021).
14. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell***181**, 1036-1045.e9 (2020).
15. Sevajol, M., Subissi, L., Decroly, E., Canard, B. & Imbert, I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res.***194**, 90–99 (2014).
16. Simmonds, P. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere***5**, (2020).
17. Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G. & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.***6**, eabb5813 (2020).
18. Simmonds, P. Pervasive RNA Secondary Structure in the Genomes of SARS-CoV-2 and Other Coronaviruses. *mBio***11**, (2020).
19. Kosuge, M., Furusawa-Nishii, E., Ito, K., Saito, Y. & Ogasawara, K. Point mutation bias in SARS-CoV-2 variants results in increased ability to stimulate inflammatory responses. *Sci. Rep.***10**, 17766 (2020).
20. Shevchenko, G. & Morris, K. V. All I's on the RADAR: role of ADAR in gene regulation. *FEBS Lett.* (2018) doi:10.1002/1873-3468.13093.
21. Pestal, K. *et al.* Isoforms of the RNA editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity***43**, 933–944 (2015).

22. Picardi, E., Mansi, L. & Pesole, G. A-to-I RNA editing in SARS-CoV-2: real or artifact? *bioRxiv* 2020.07.27.223172 (2020) doi:10.1101/2020.07.27.223172.
23. Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.***5**, 4726 (2014).
24. Miorin, L. *et al.* SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc. Natl. Acad. Sci. U. S. A.***117**, 28344–28354 (2020).
25. Israelow, B. *et al.* Mouse model of SARS-CoV-2 reveals inflammatory role of type I interferon signaling. *J. Exp. Med.***217**, (2020).
26. Desai, N. *et al.* Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection. *Nat. Commun.***11**, 6319 (2020).
27. Nchioua, R. *et al.* SARS-CoV-2 Is Restricted by Zinc Finger Antiviral Protein despite Preadaptation to the Low-CpG Environment in Humans. *mBio***11**, (2020).
28. Lamers, M. M., van den Hoogen, B. G. & Haagmans, B. L. ADAR1: ‘Editor-in-Chief’ of Cytoplasmic Innate Immunity. *Front. Immunol.***10**, 1763 (2019).
29. Chen, J. & MacCarthy, T. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genes. *PLoS Comput. Biol.***13**, (2017).
30. Wolff, G., Melia, C. E., Snijder, E. J. & Bárcena, M. Double-Membrane Vesicles as Platforms for Viral Replication. *Trends Microbiol.***28**, 1022–1033 (2020).
31. Mourier, T. *et al.* Host-directed editing of the SARS-CoV-2 genome. *Biochem. Biophys. Res. Commun.***538**, 35–39 (2021).
32. Shapiro, M., Meier, S. & MacCarthy, T. The cytidine deaminase under-representation reporter (CDUR) as a tool to study evolution of sequences under deaminase mutational pressure. *BMC Bioinformatics***19**, 163 (2018).
33. R: The R Project for Statistical Computing. <https://www.r-project.org/>.

Declarations

Acknowledgments

University of Padova Strategic Research Infrastructure Grant 2017: “CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l’Innovazione”. DOR 2020 granted to PV.

Author contributions

EB and UR designed the analysis; EB and UR analyzed hyper-editing data; EB and MS analyzed coronavirus genome data; UR analyzed RNA-seq data; EB, MS, AL and PV contributed to discussion and to the homogenization of the different analyses; UR wrote the manuscript; PV and EB revised and improved the manuscript.

Competing Interests statement

The authors declare no competitive interests.

Figures

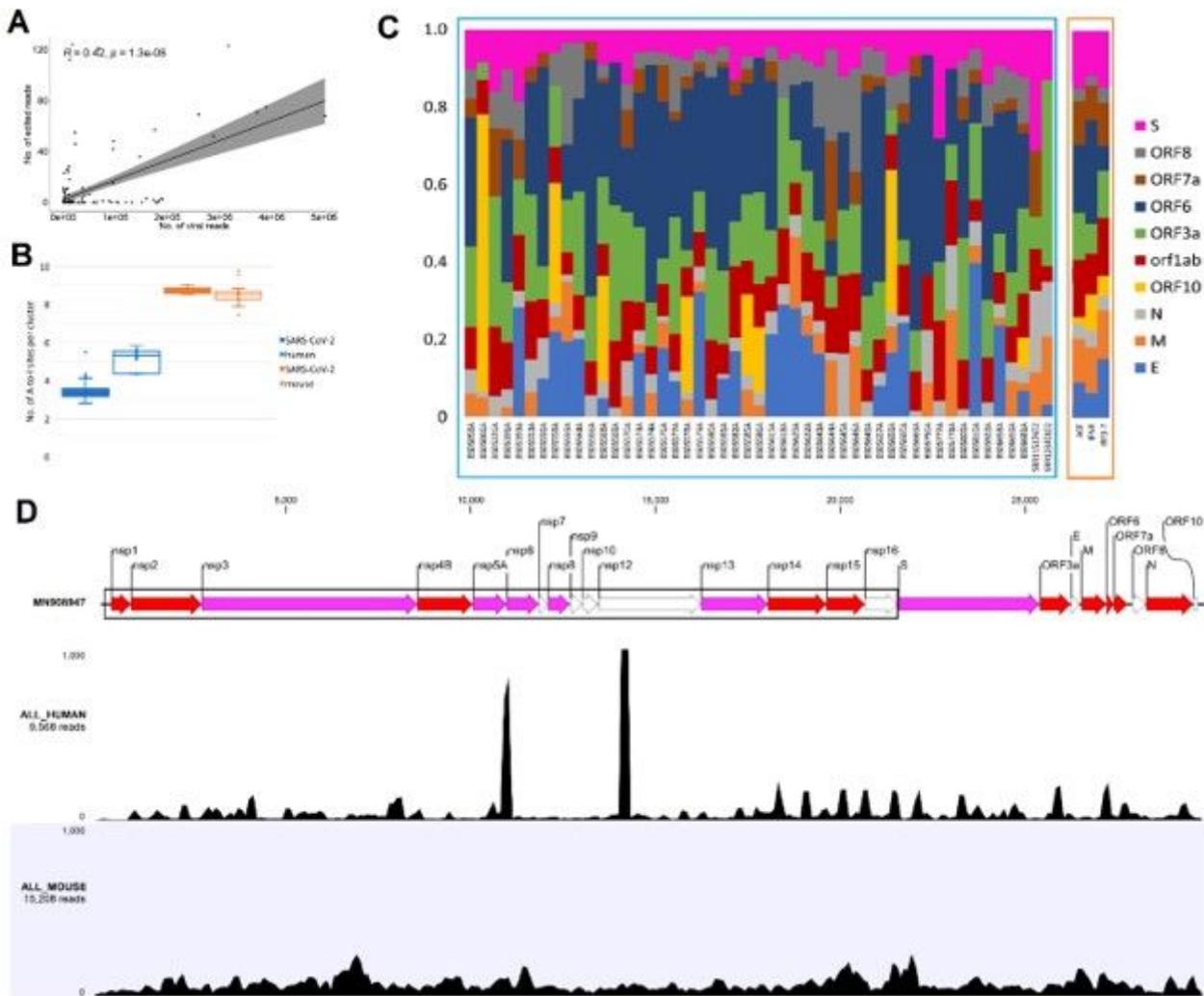


Figure 1

ADAR hyper-editing on SARS-CoV-2. A. The number of he reads per sample was plotted against the number of SARS-CoV-2 reads for samples having at least one he read. B. Hyper-editing efficiency was measured as number of he sites per cluster in samples with at least 100 he reads. The plot illustrates data for SARS-CoV-2 in human (N=49) and mouse (N=3) hosts, for the available human host (paired to the analyzed SARS-CoV-2 datasets, N=7) and for paired mouse host (N=8). C. The relative normalized he level per SARS-CoV-2 ORF is reported for samples with at least 100 he reads. He levels were normalized by the number of unique mapped reads per ORF. Human and mouse samples are framed in blue and orange, respectively. D. SARS-CoV-2 genome (MN908947) and its coding regions with two linear profiles

depicting the distribution of ADAR he on SARS-CoV-2 genome in human and mouse samples. The ORFs were colored by the susceptibility of ADAR editing, violet for ORFs with “maximized” ‘WA’ under-representations, red for ‘WA’ “under-represented” ORFs and white for the others.

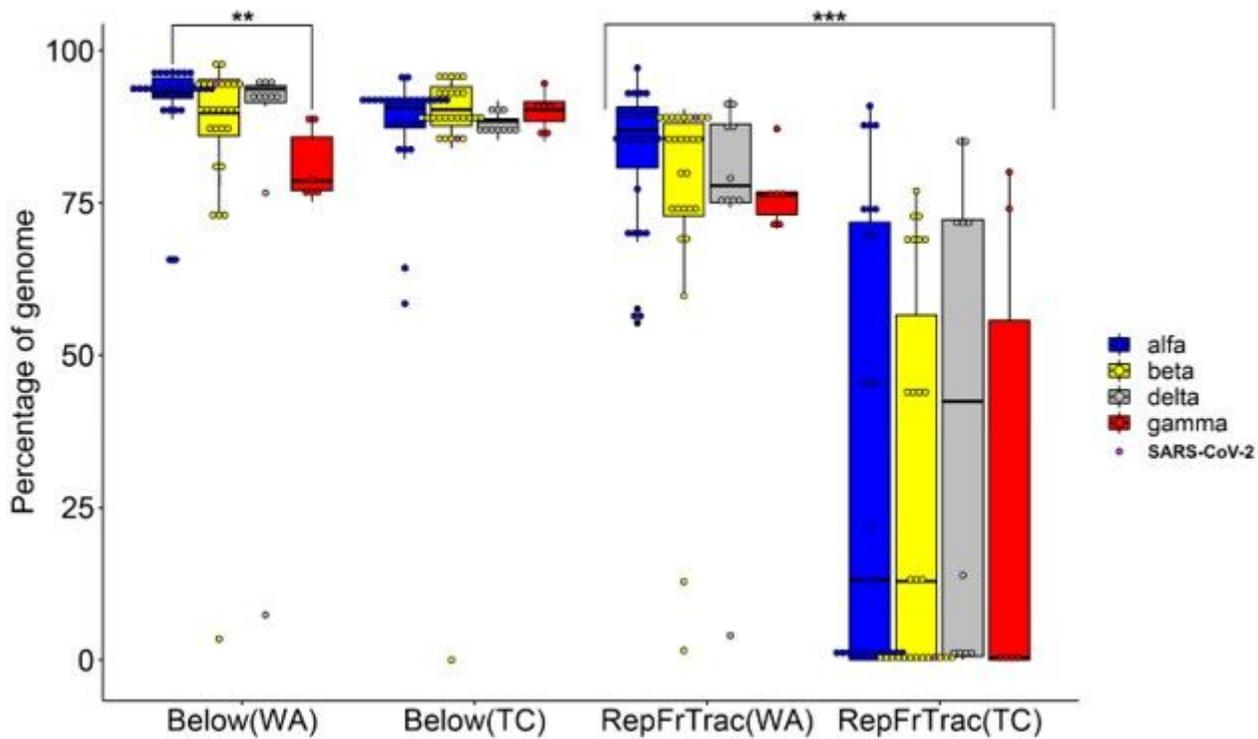


Figure 2

Percentage of viral genomes that display under-representation of the ‘WA’ or ‘TC’ motifs (below metric) or that have maximized the under-representations of these dinucleotides (repFrTrac metric). The different coronavirus genera are considered as separate groups. Significant p-values are reported, <0.01 ** or $<0,001$ ***.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S.Table1.xlsx](#)
- [S.Table2.xlsx](#)
- [SupplementaryFiguresv6.pdf](#)