



Retrospect of the Two-Year Debate: What Fuels the Evolution of SARS-CoV-2: RNA Editing or Replication Error?

Lai Wei¹

Received: 22 October 2022 / Accepted: 19 March 2023 / Published online: 28 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Mutation is one of the mechanisms of the evolutionary divergence of an organism. Under this global COVID-19 pandemic, the fast evolution of SARS-CoV-2 became one of the most worrying issues. Some researchers believed that the hosts' RNA deamination systems (APOBECs and ADARs) are the major source of mutations and have driven the evolution of SARS-CoV-2. However, apart from RNA editing, the RDRP (RNA-dependent RNA polymerase)-mediated replication errors may also contribute to the mutation of SARS-CoV-2 (just like the single-nucleotide polymorphisms/variations in eukaryotes caused by DNA replication errors). Unfortunately, it is technically unable to distinguish RNA editing and replication errors (SNPs) in this RNA virus. Here comes a fundamental question: we indeed observed the fast evolution of SARS-CoV-2, but what exactly fuels its evolution: RNA editing or replication errors? This debate lasts for 2 years. In this piece, we will retrospect the 2-year debate on RNA editing versus SNPs.

Introduction

The continuous mutation and evolution of SARS-CoV-2 should be one of the major threats to humans under this global pandemic. Newly emerged virus strains might acquire the ability to escape the current vaccines. Understanding the molecular mechanisms underlying the rampant mutation of SARS-CoV-2 is urgently needed.

As SARS-CoV-2 is a typical RNA virus, it is believed that the hosts' (humans') RNA deamination systems (APOBECs and ADARs) have driven the mutation and evolution of SARS-CoV-2 [1–3]. Both enzyme families would cause nucleotide changes in RNA sequences. APOBECs drive the C-to-U(T) alteration in RNAs, whilst ADARs drive the A-to-I(G) alteration in RNAs. However, apart from RNA editing mediated by APOBECs and ADARs, the RDRP (RNA-dependent RNA polymerase)-mediated replication errors may also contribute to the mutation profile of SARS-CoV-2. To better understand what replication errors are, one could consider the single-nucleotide polymorphisms (SNPs) in eukaryotes. The SNPs in eukaryotes are essentially DNA mutations introduced during replication, leading to

single-nucleotide variants (SNVs) between a given sequence and the reference genome sequence. When studying DNA organisms, many well-established pipelines could help researchers distinguish DNA SNPs and RNA editing events (when genome sequencing and RNA-sequencing are available) [4–6]. However, for RNA viruses, like SARS-CoV-2, it is technically unable to distinguish RNA editing events and replication errors caused by RDRP [7] because both processes take place on RNAs.

The difficulty in verifying the mutation source of SARS-CoV-2 has led to debates. What exactly fuels the evolution of SARS-CoV-2: RNA editing or replication errors (SNPs)? This debate lasts for 2 years. In this short article, we will retrospect the 2-year debate on the driving force of SARS-CoV-2 evolution: RNA editing versus replication errors (SNPs).

Stage 1: the “Trigger” of this Debate

The debate began with a paper published in *Science Advances* written by Di Giorgio et al. [8]. This paper identified SNVs (between RNA-sequencing data and the reference sequence of SARS-CoV-2) and showed a typical symmetric profile in the transcriptome of SARS-CoV-2. Normally, only the replication errors could lead to a symmetric SNV profile because the polymerase machinery make mistakes equally

✉ Lai Wei
wl@bnu.edu.cn

¹ College of Life Sciences, Beijing Normal University, Beijing, China

on both strands (positive or negative strand) during replication [9, 10]. In contrast, when replication errors (or SNPs) in SARS-CoV-2 are excluded, the remaining SNVs (if any) should belong to RNA editing events. A typical SNV profile of RNA editing should significantly skew to a particular type of mutation, such as A-to-G or C-to-T [11–13]. Amazingly, even with a symmetric SNV profile in hand, the Di Giorgio et al. paper concluded that what they found were RNA editing events. It seems that the data presented by Di Giorgio et al. were actually supporting the SNP-driving view of the fast evolution of SARS-CoV-2 (although they claimed RNA editing). Soon after its publication, this paper incurred a tsunami of criticisms.

Stage2: Debate: What is a Reliable SNV Profile of RNA Editing?

Three independent papers were simultaneously submitted to three different journals with similar indication that the findings in [8] paper were unreliable.

In detail, (1) [14] pointed out that as [8] has just provided a symmetric SNV profile, the strong evidence for RNA editing has not yet been provided. The view held by Picardi et al. echoes our previous introduction about the relationship between SNV profile and the confidence of RNA editing.

(2) [15] disproved the so-called “RNA editing motifs” shown by [8] and meanwhile, Song et al. displayed an SNV profile with slight enrichment on A-to-G mutation. Genuine RNA editing sites tend to reside in a particular sequence context due to the binding preference of the editing enzymes [16–18]. This feature usually serves as supporting evidence to show the reliability of RNA editing sites. Song et al. claimed that Di Giorgio et al. failed to support the RNA-editing sites with the sequence context.

(3) [10] directly concluded that [8] has proved nothing but mechanically running a series of bioinformatic pipelines. Zong et al.’s views might be opaque to common readers. Their key point of Zong et al. is, the same bioinformatic pipelines (i.e. the variant calling pipeline) could be applied to any datasets regardless of the “biological meaning” of the output results. Di Giorgio et al. just ran the pipeline and obtained a non-informative result (the symmetric SNV profile) but they interpreted the “null result” as RNA editing events. Zong et al. claimed that the entire Di Giorgio paper was based on the mis-interpretation of the SNV profile.

From their debate, we could understand that their core argument lies in the mutation (SNV) profile. An SNP profile caused by replication errors is symmetric [8], whilst an RNA editing profile is skewed to a particular type of variation, like A-to-G (representing A-to-I RNA editing) [15] or C-to-T (representing C-to-U RNA editing) [1, 19]. As many bioinformatic methods have claimed, the accurate

identification of RNA editing events requires multiple steps of hard filters to exclude the replication errors (SNPs) or even sequencing errors [13, 20, 21]. However, even with stringent pipelines, one could not always obtain an SNV profile enriched with a particular mutation type [22]. The non-optimal SNV profile could not be regarded as evidence for RNA editing [23]. Under this common sense shared by the RNA editing community, Di Giorgio et al. definitely failed to provide evidence for RNA editing in SARS-CoV-2.

Given that Di Giorgio et al. failed to show a reliable SNV profile to support the existence of RNA editing, we would expect that those critical papers [10, 14, 15] could improve the pipeline and find some genuine RNA-editing sites in SARS-CoV-2. However, the A-to-G enrichment shown by [15] is still very weak and it is hard to say that [15] has made much improvement compared to [8]. It is still unclear whether replication errors (SNPs) or RNA editing events dominate the mutations (SNVs) found in SARS-CoV-2 RNA.

Stage3: Argument on False Positive and True Positive

After the harsh criticism by [10], the group of [8] has responded. Di Giorgio et al. argued that although there might be false-positive RNA-editing sites in the SNVs they found [8], there must be true RNA-editing sites in the SNV profile [24]. The definition of true/false-positive rates was based on the enrichment of the desired type of mutation. For instance, Li et al. found that 96% of the SNVs in the ant transcriptome was A-to-G sites so that the true-positive rate of A-to-I RNA editing was 96%, whilst false-positive rate was 4% [12]. According to this definition, Di Giorgio et al.’s response paper [24] was still weak and pale although they smartly circumvented the key criticism raised by [10]. Martignano et al. still failed to give a confidence interval of the accuracy (true-positive rate) of the so-called RNA-editing sites they identified. Based on the SNV profile shown by [8], the true-positive rate of RNA-editing sites was actually lower than 50% by definition. This almost represented a “random result”.

From another aspect, even one acknowledges the statement of [24], it is irrefutable that the existence of true-positive sites does not “forgive” the large number of false-positive sites in the SNV profile [25]. The reason is every clear: readers would presume that the A-to-G variations shown in the paper are all A-to-I editing sites instead of thinking that “Oh, the A-to-G variations may have 50% false-positive rates by default...”.

Stage4: The Logic Problem and the Golden Standard

Since the response by Di Giorgio et al. group [24] was unsatisfactory, it incurred new criticisms. [9] asked an ultimate question: if the symmetric SNV profile shown by [8] could be regarded as evidence for RNA editing, then (1) what will be the golden standard for RNA editing detection? (2) Why should others try so hard to filter out false-positive sites in order to enrich a particular type of mutation like A to G?

The logic is, the symmetric SNV profile could be obtained “by default” (without any additional efforts in bioinformatics), because the polymerase errors (including sequencing errors plus replication errors) would intrinsically produce a symmetric SNV profile. Only RNA editing is able to increase the number of a particular type of SNV (A to G or C to T), leading to an asymmetric SNV profile where the true-positive rates equals the enrichment of the target mutation type. Providing a symmetric SNV profile almost proves nothing.

Despite the intensive criticism by [9], they did not propose an alternative methodology to improve the RNA editing detection pipeline. Given the nearly random results shown by [8], it remains unclear are there any other highly reliable data suggesting the RNA-editing origin of the mutations in SARS-CoV-2.

Stage5: It Turns Out To Be C-to-U RNA Editing that Fuels SARS-CoV-2 Evolution

Although the [8] paper was imperfect in many aspects, none of the critical papers have raised any convincing enough evidence to show that RNA editing fuels the evolution of SARS-CoV-2. Researchers were willing to believe that RNA editing events exist in the SARS-CoV-2 transcriptome but the non-optimal SNV profile [8] is always an obstacle that prevents the scientists from reaching a solid conclusion.

Several recent papers finally ended this debate [26, 27] by providing compelling evidence and explanations. The common point of these recent papers is, they used the millions of world-wide SARS-CoV-2 sequences from GISAID [28] instead of using the intra-host transcriptome data. The polymorphic sites in global SARS-CoV-2 population exhibited a striking peak towards C to T, representing C-to-U RNA editing [26]. As previously introduced/discussed about the commonly accepted criteria of RNA editing detection, this strongly asymmetric mutation profile could only be explained by the rampant C-to-U RNA editing. No alternative theories could explain such abundant/excessive C-to-T mutation sites.

Therefore, the debate is ended. The intra-host transcriptome data [8] might contain unknown confounding factors that obscured signals of RNA editing. The world-wide SARS-CoV-2 sequences successfully showed enrichment for C-to-U editing sites. Thus, the driving force of SARS-CoV-2 evolution turns out to be C-to-U RNA editing. From the beginning where a symmetric SNP profile was provided [8] to this end where a clear enrichment of C-to-T sites was shown [26], two years have passed. Researchers have discussed true/false-positive rates, golden standard of RNA editing detection, the methodology, and many logic issues. Retrospect of this debate is helpful for future studies on RNA editing and SARS-CoV-2 evolution.

Acknowledgements We thank all medical workers who fight against SARS-CoV-2 during this pandemic.

Author Contributions LW designed and supervised this study. LW wrote the manuscript. LW approved the submission of the manuscript.

Funding This study was not supported by funding.

Declarations

Conflict of interest The author declares no conflict of interest.

Ethical Approval Not applicable.

Informed Consent Not applicable.

References

- Li Y, Yang XN, Wang N, Wang HY, Yin B, Yang XP, Jiang WQ (2020) The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. *Future Virol* 15:341–347
- Zhang YP, Jiang W, Li Y, Jin XJ, Yang XP, Zhang PR, Jiang WQ, Yin B (2021) Fast evolution of SARS-CoV-2 driven by deamination systems in hosts. *Future Virol* 16:587–590
- Zhao M, Li C, Dong Y, Wang X, Jiang W, Chen Y (2022) Nothing in SARS-CoV-2 makes sense except in the light of RNA modification? *Future Virol* 17:769
- Eisenberg E (2012) Bioinformatic approaches for identification of A-to-I editing sites. *Curr Top Microbiol Immunol* 353:145–162
- Picardi E, Pesole G (2013) REDIttools: high-throughput RNA editing detection made easy. *Bioinformatics* 29:1813–1814
- Porath HT, Carmi S, Levanon EY (2014) A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat Commun* 5:4726
- Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, Jiang W (2020) SNPs or RNA modifications? Concerns on mutation-based evolutionary studies of SARS-CoV-2. *PLoS ONE* 15:e0238490
- Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG (2020) Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv* 6:eabb5813
- Cai H, Liu X, Zheng X (2022) RNA editing detection in SARS-CoV-2 transcriptome should be different from traditional SNV identification. *J Appl Genet* 63:587–594
- Zong J, Zhang Y, Guo F, Wang C, Li H, Lin G, Jiang W, Song X, Zhang X, Huang F et al (2022) Poor evidence for host-dependent

- regular RNA editing in the transcriptome of SARS-CoV-2. *J Appl Genet* 63:413–421
11. Duan Y, Cai W, Li H (2023) Chloroplast C-to-U RNA editing in vascular plants is adaptive due to its restorative effect: testing the restorative hypothesis. *RNA* 29:141–152
 12. Li Q, Wang Z, Lian J, Schiott M, Jin L, Zhang P, Zhang Y, Nygaard S, Peng Z, Zhou Y et al (2014) Caste-specific RNA editomes in the leaf-cutting ant *Acromyrmex echinatior*. *Nat Commun* 5:4943
 13. Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB (2013) Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods* 10:128–132
 14. Picardi E, Mansi L, Pesole G (2021) Detection of A-to-I RNA Editing in SARS-COV-2. *Genes (Basel)* 13:41
 15. Song Y, He X, Yang W, Wu Y, Cui J, Tang T, Zhang R (2022) Virus-specific editing identification approach reveals the landscape of A-to-I editing and its impacts on SARS-CoV-2 characteristics and evolution. *Nucleic Acids Res* 50:2509–2521
 16. Chen CX, Cho DS, Wang Q, Lai F, Carter KC, Nishikura K (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* 6:755–767
 17. Palladino MJ, Keegan LP, O'Connell MA, Reenan RA (2000) dADAR, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* 6:1004–1018
 18. Porath HT, Knisbacher BA, Eisenberg E, Levanon EY (2017) Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome Biol* 18:185
 19. Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, Jiang W (2020) Mutation profile of over 4500 SARS-CoV-2 isolations reveals prevalent cytosine-to-uridine deamination on viral RNAs. *Future Microbiol* 15:1343–1352
 20. Ramaswami G, Lin W, Piskol R, Tan MH, Davis C, Li JB (2012) Accurate identification of human Alu and non-Alu RNA editing sites. *Nat Methods* 9:579–581
 21. Zhang Q, Xiao X (2015) Genome sequence-independent identification of RNA editing sites. *Nat Methods* 12:347–350
 22. Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG (2011) Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333:53–58
 23. Lin W, Piskol R, Tan MH, Li JB (2012) Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science* 335:1302 (**author reply 1302**)
 24. Martignano F, Di Giorgio S, Mattiuz G, Conticello SG (2022) Commentary on “Poor evidence for host-dependent regular RNA editing in the transcriptome of SARS-CoV-2.” *J Appl Genet* 63:423–428
 25. Wei L (2022) Reconciling the debate on deamination on viral RNA. *J Appl Genet* 63:583–585
 26. Liu X, Liu X, Zhou J, Dong Y, Jiang W, Jiang W (2022) Rampant C-to-U deamination accounts for the intrinsically high mutation rate in SARS-CoV-2 spike gene. *RNA* 28:917–926
 27. Zhu L, Wang Q, Zhang W, Hu H, Xu K (2022) Evidence for selection on SARS-CoV-2 RNA translation revealed by the evolutionary dynamics of mutations in UTRs and CDSs. *RNA Biol* 19:866–876
 28. Shu Y, McCauley J (2017) GISAID: Global initiative on sharing all influenza data—from vision to reality. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.