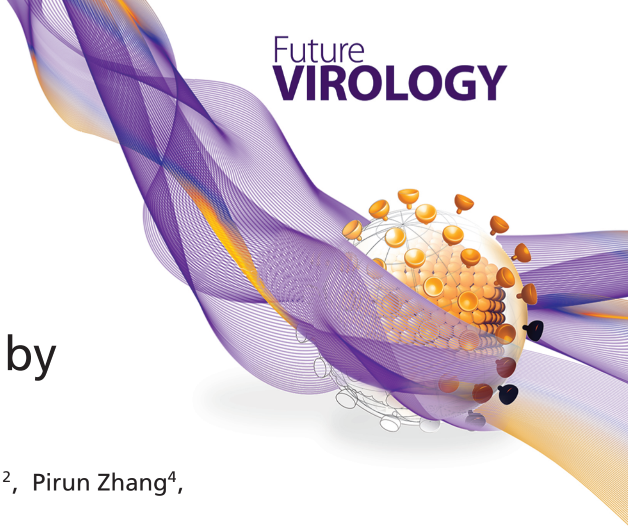
Future
VIROLOGY



Editorial

Fast evolution of SARS-CoV-2 driven by deamination systems in hosts

Yanping Zhang[1,2], Wen Jiang[1,2], Yan Li[3], Xiaojie Jin[1,2], Xiaoping Yang[1,2], Pirun Zhang[4], Wenqing Jiang*[,1,2] & Bin Yin**[,1,2]

[1]Department of Respiratory Diseases, Qingdao Haici Hospital, Shandong, China
[2]The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China
[3]Department of Cardiology, Qingdao Center Hospital, Shandong, China
[4]College of Basic Medicine, Shandong University of Traditional Chinese Medicine, Shandong, China
*Author for correspondence: Tel.: +86 0532 83777856; qdhospit87@163.com
**Author for correspondence: qd0532yb@163.com

> “These prevalent deamination events already caused ambiguity and uncertainty in evolutionary studies due to the fact that these deamination events and other natural mutations (replication errors) are technically indistinguishable”

Tweetable abstract: As an RNA virus, the fast evolution of SARS-CoV-2 is driven by the extensive RNA deamination by the host cells.

First draft submitted: 20 July 2021; Accepted for publication: 25 August 2021; Published online: 10 September 2021

Keywords: evolution • natural selection • RNA deamination • RNA structure • SARS-CoV-2

Cytidine-to-uridine deamination is overrepresented in SARS-CoV-2 compared with other organisms

Viruses are exposed to the cellular environment after infecting the cells. As many recent literatures revealed, SARS-CoV-2 is subjected to the hosts’ deamination systems, which include the cytidine-to-uridine (C-to-U) deamination by APOBECs [1,2] and the adenosine-to-inosine (A-to-I) deamination by *ADAR*s [3]. As I is equivalent to G, both C-to-U and A-to-I deamination would cause nucleotide substitution. These prevalent deamination events already caused ambiguity and uncertainty in evolutionary studies due to the fact that these deamination events and other natural mutations (replication errors) are technically indistinguishable [4,5]. Thus, understanding the occurrence and evolution of these host-driven mutations is helpful in predicting the co-evolution of the host-parasite arms-race as well as controlling the pandemic.

Among the polymorphic sites in the SARS-CoV-2 genome, the number of C-to-U substitution sites is about three-times higher than the number of A-to-I substitution sites [1]. This ‘C–U/A–I’ ratio = 3 is extraordinarily high compared with any other organisms with extensive RNA deamination studies. Here, we give evidence and data to show how unique SARS-CoV-2 is. Since most regions of the SARS-CoV-2 genome are coding sequences (CDS), the relative prevalence of C–U and A–I events in SARS-CoV-2 should be compared with deamination events in the CDS of other organisms. For instance, in the coding sequence of human (*Homo sapiens*), 128 C–U sites [6] and 1783 A–I sites [7] were found, resulting in a C–U/A–I ratio of 0.07. Similarly, in other species, the C–U/A–I ratio is less than 0.01 in mouse (*Mus musculus*) [8] and worm (*Caenorhabditis elegans*) [9], approximately 0.10 in pig (*Sus scrofa*) [10], approximately 0.04 in *Octopus vulgaris* (also named *Octopus sinensis*) [11] and approximately 0.06 in fly (*Drosophila melanogaster*) [12].

In the animal species listed above, the A–I sites are much more frequent than C–U sites (usually over 20-fold), while SARS-CoV-2 possesses a larger number of C–U sites over A–I sites (three fold). Totally, compared with A-to-I events, the C-to-U deamination is over-represented in SARS-CoV-2 for roughly 60-fold . As we know, the occurrence of deamination events is determined by *cis* elements and *trans* factors. In fact, when SARS-CoV-2 infects human cells, the host mRNAs and the viral RNAs are subjected to the same *trans* environment. Therefore,

Future Medicine

the answers to why the C–U and A–I deamination events differ so much between host and virus should be found in the *cis* sequence features. We propose that the genome component and RNA structure may serve as the basis of such different deamination spectrums between SARS-CoV-2 and animals.

RNA structure may determine the prevalence of C-to-U deamination in SARS-CoV-2

Genome architecture varies widely from species to species. In the majority of animals, the GC content in the genome is around 60%, suggesting that GC is slightly favored over AT by animals [13]. In contrast, SARS-CoV-2 has an AT (AU) content above 50%, which means that the virus prefers AT. Since the GC base pairs are biochemically more stable than AT base pairs, an RNA sequence with higher GC content is more likely to fold into a stable secondary structure and form dsRNAs [14]. Therefore, the mRNA pools in animal cells with high GC content are likely to form dsRNA structures while the RNA virus like SARS-CoV-2 with low GC content will tend to stay in unfolded linear state or less stable secondary structures.

It is known that the RNA deaminases have their own preference on RNA structures. The adenosine deaminase *ADAR* has high affinity to stable dsRNA structures. *ADAR* binds to dsRNAs and converts adenosines to inosines within the secondary structure [15]. The cytidine deaminase with similar function to APOBEC tends to bind to ssRNAs [16] and converts cytidines to uridines. In a cell that is invaded by SARS-CoV-2, the endogenous mRNAs with more dsRNA structures are more likely to be bound by *ADAR* while the unstructured linear viral RNAs are preferentially targeted by APOBEC. Although both host RNAs and viral RNAs are staying in the same environment, they do undergo different fates due to their distinct sequence features. Consequently, C-to-U deamination becomes predominant in SARS-CoV-2 genome.

An extraordinarily high fraction of the SARS-CoV-2 genome is deaminated, indicating fast evolution

For animal species like humans, if one wonders what percentages of the human genome (coding sequence in mRNAs) are subjected to A-to-I or C-to-U, then the answer would be less than 0.01% for both deamination types [6,7]. That means, less than 0.01% of the adenosines in human coding regions are potentially 'deaminatable'. Similarly, less than 0.01% of the cytidines in CDS belong to the potential C-to-U candidates. In sharp contrast, among the 30,000 nucleotides in the SARS-CoV-2 genome, at least 10% of them was found to be candidates for deamination across world-wide isolations [1]. This 0.01 versus 10% difference seems striking, but it becomes plausible under the central dogma and the nature of different species.

No matter for human or SARS-CoV-2, the mutation calling pipeline is identical. The deamination sites in the RNA-seq data are searched against the reference genome sequence. The reference sequence is unchanging while the short reads in the RNA-seq data could be largely different according to the specimen and samples collected. Intriguingly, the central dogma dictates that the human mRNAs should be transcribed from DNA before they are subjected to RNA deamination. Therefore, human RNAs are not inheritable. The RNA deamination sites in one generation might not necessarily be deaminated in the next generation. The human RNA deamination sites could not accumulate generation by generation. When it comes to SARS-CoV-2, which is an RNA virus, the RNAs are the inheritable genetic materials. SARS-CoV-2 proliferates by RNA replication, so that the viral RNAs subjected to A-to-I or C-to-U conversions would be directly used as the template for the next generation. Mutations in the template RNA are faithfully transmitted to the daughter-strand RNA. Under this mechanism, the A–I and C–U substitution sites are accumulated rapidly in the SARS-CoV-2 sequence. This also gives an explanation for the fact that the virus strains isolated from different patients often carry completely different mutation sites. That is caused by the random nature of deamination exerted by the host.

Natural selection constraining the host-driven evolution of SARS-CoV-2

Based on the nature of RNA replication strategy used by SARS-CoV-2, the A-to-I and C-to-U substitutions are rapidly accumulating in the virus genome. Intriguingly, the AT-rich regions, which are usually single stranded and likely to be targeted by APOBEC, are the hotspots for C-to-U deamination. Then, the deamination events will lead to a higher AT (AU) content as C is converted to U. In the next generation, this local RNA region with elevated AT content will be more likely to be targeted by APOBEC and then subjected to C-to-U deamination again. This endless loop seems unstoppable until all cytidines are converted to uridines.

However, in real observation, the evolution of SARS-CoV-2 sequence is not as fast as this model predicts. It is true that the viral sequence evolves much faster than eukaryotic genomes, but the C-to-U deamination machineries have not yet ‘erased’ all the cytidines in SARS-CoV-2. Evolution is a balance between mutation and natural selection. While mutations take place randomly, natural selection eliminates the majority of individuals carrying deleterious mutations. Under particular circumstances, a SARS-CoV-2 sequence with a C-to-U substitution might be less adaptive compared with the viral sequence before mutation. For instance, a viral sequence with surplus AT(AU)-ending codons is unsuitable for viral RNA translation [17,18], and this sequence (if resulted from extensive C-to-U deamination) would be eliminated by the host cell and does not have a chance to transmit to the next generation.

Future perspective

Natural selection slows down the observed nucleotide substitution rate in viruses. Nevertheless, the host-driven deamination on SARS-CoV-2 sequences provides numerous chances for the virus to test which sequence is more suitable for invading human cells. If one viral sequence (out of a million different sequences) shows higher virulence and adaptiveness, then this sequence would rapidly spread throughout the infected cells and dominate the viral population. In other words, one success out of a million trials might be enough for the virus to evolve.

Once the virus invades the hosts, we are unable to shut down the cellular RNA deamination system. Therefore, the fast mutation of viral RNA is inevitable at that late stage. However, a feasible way to alleviate the fast mutation of a virus is to block virus transmission. Without accession to host cells, viral RNAs could not change their own sequences autonomously. Then, the virus mutation rate is largely reduced. Altogether, in the light of evolutionary and molecular biology, the fast evolution of SARS-CoV-2 sequence is driven by the deamination systems in hosts. Even with the recent success in many sorts of vaccines, preventing the virus transmission is still one of the powerful approaches to control the pandemic.

Acknowledgments

The authors thank the members in the group who have provided suggestions for this article.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1. Li Y, Yang X, Wang N *et al.* Mutation profile of over 4500 SARS-CoV-2 isolations reveals prevalent cytosine-to-uridine deamination on viral RNAs. *Future Microbiol.* 15(14), 1343–1352 (2020).
2. Harris RS, Dudley JP. APOBECs and virus restriction. *Virology* 479–480, 131–145 (2015).
3. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.* 6(25), eabb5813 (2020).
4. Li Y, Yang XN, Wang N *et al.* The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. *Future Virol.* 15(6), 341–347 (2020).
5. Li Y, Yang XN, Wang N *et al.* Pros and cons of the application of evolutionary theories to the evolution of SARS-CoV-2. *Future Virol.* 15(6), 369–372 (2020).
6. Liu Z, Zhang J. Human C-to-U coding RNA editing is largely nonadaptive. *Mol. Biol. Evol.* 35(4), 963–969 (2018).
7. Xu G, Zhang J. Human coding RNA editing is generally nonadaptive. *Proc. Natl Acad. Sci. USA* 111(10), 3769–3774 (2014).
8. Licht K, Kapoor U, Amman F *et al.* A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing. *Genome Res.* 29(9), 1453–1463 (2019).
9. Zhao HQ, Zhang P, Gao H *et al.* Profiling the RNA editomes of wild-type *C. elegans* and *ADAR* mutants. *Genome Res.* 25(1), 66–75 (2015).
10. Jiang Y, Cao X, Wang H. Mutation profiling of a limbless pig reveals genome-wide regulation of RNA processing related to bone development. *J. Appl. Genet.* doi:10.1007/s13353-021-00653-0 (2021) (Epub ahead of print).
11. Liscovitch-Brauer N, Alon S, Porath HT *et al.* Trade-off between transcriptome plasticity and genome evolution in cephalopods. *Cell* 169(2), 191–202 e111 (2017).
12. Yu Y, Zhou H, Kong Y *et al.* The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. *PLoS Genet.* 12(7), e1006191 (2016).

13. Li Y, Yang X, Wang N *et al.* GC usage of SARS-CoV-2 genes might adapt to the environment of human lung expressed genes. *Mol. Genet. Genomics* 295(6), 1537–1546 (2020).
14. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res.* 31(13), 3429–3431 (2003).
15. Bass BL. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71(1), 817–846 (2002).
16. Yin P, Li QX, Yan CY *et al.* Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504(7478), 168–171 (2013).
17. Yu YY, Li Y, Dong Y, Wang XK, Li CX, Jiang WQ. Natural selection on synonymous mutations in SARS-CoV-2 and the impact on estimating divergence time. *Future Virol.* 16(7), 447–450 (2021).
18. Wang Y, Gai Y, Li Y, Li C, Li Z, Wang X. SARS-CoV-2 has the advantage of competing the iMet-tRNAs with human hosts to allow efficient translation. *Mol. Genet. Genomics* doi:10.1007/s00438-020-01731-4 (2020) (Epub ahead of print).