



# Compelling Evidence Suggesting the Codon Usage of SARS-CoV-2 Adapts to Human After the Split From RaTG13

Evolutionary Bioinformatics  
Volume 17: 1–6  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11769343211052013



Yanping Zhang<sup>1,2</sup>, Xiaojie Jin<sup>1,2</sup>, Haiyan Wang<sup>1,2</sup>, Yaoyao Miao<sup>1,2</sup>, Xiaoping Yang<sup>1,2</sup>, Wenqing Jiang<sup>1,2</sup>  and Bin Yin<sup>1,2</sup> 

<sup>1</sup>Department of Respiratory Diseases, Qingdao Haici Hospital, Qingdao, Shandong, China. <sup>2</sup>The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China.

**ABSTRACT:** SARS-CoV-2 needs to efficiently make use of the resources from hosts in order to survive and propagate. Among the multiple layers of regulatory network, mRNA translation is the rate-limiting step in gene expression. Synonymous codon usage usually conforms with tRNA concentration to allow fast decoding during translation. It is acknowledged that SARS-CoV-2 has adapted to the codon usage of human lungs so that the virus could rapidly proliferate in the lung environment. While this notion seems to nicely explain the adaptation of SARS-CoV-2 to lungs, it is unable to tell why other viruses do not have this advantage. In this study, we retrieve the GTEx RNA-seq data for 30 tissues (belonging to over 17 000 individuals). We calculate the RSCU (relative synonymous codon usage) weighted by gene expression in each human sample, and investigate the correlation of RSCU between the human tissues and SARS-CoV-2 or RaTG13 (the closest coronavirus to SARS-CoV-2). Lung has the highest correlation of RSCU to SARS-CoV-2 among all tissues, suggesting that the lung environment is generally suitable for SARS-CoV-2. Interestingly, for most tissues, SARS-CoV-2 has higher correlations with the human samples compared with the RaTG13-human correlation. This difference is most significant for lungs. In conclusion, the codon usage of SARS-CoV-2 has adapted to human lungs to allow fast decoding and translation. This adaptation probably took place after SARS-CoV-2 split from RaTG13 because RaTG13 is less perfectly correlated with human. This finding depicts the trajectory of adaptive evolution from ancestral sequence to SARS-CoV-2, and also well explains why SARS-CoV-2 rather than other viruses could perfectly adapt to human lung environment.

**KEYWORDS:** SARS-CoV-2, RaTG13, mutation, synonymous, evolution

**RECEIVED:** August 20, 2021. **ACCEPTED:** September 17, 2021.

**TYPE:** Short Report

**FUNDING:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Bin Yin, Department of Respiratory Diseases, Qingdao Haici Hospital, Qingdao, Shandong 266033, China. Email: qd0532yb@163.com

## Introduction

Viruses need to adapt to hosts to survive and proliferate.<sup>1</sup> In order to rapidly replicate themselves, the key question for viruses is how to efficiently utilize the limited resources in host cells. The central dogma dictates that the virus proliferation includes (at least for RNA viruses like SARS-CoV-2) the replication and translation processes. While the elements required for replication is evolutionarily conserved across coronaviruses,<sup>2</sup> the mechanism of virus translation is quite mysterious. This prompts us to consider how the virus translate its own RNAs in the host cells. In fact, among the various biological processes, the most energy-consuming and rate-limiting step is mRNA translation.<sup>3</sup> Therefore, virus RNAs have to compete with host RNAs for translation machineries. It is believed that higher translation efficiency is favored by natural selection,<sup>4-6</sup> and this is especially crucial for viruses. Conceivably, if the high expression of viral genes is not achieved, then the optimization of viral protein sequence/structure is not meaningful because the viral genes must express to a sufficiently high amount before they exert their functions.

A smart way to elevate mRNA translation efficiency is to optimize the synonymous codon usage.<sup>7-9</sup> Although synonymous codons encode identical amino acids, they do have different cognate tRNA concentrations.<sup>10</sup> Synonymous codons with

higher tRNA availability are advantageous as they are decoded and translated with faster rates. Since tRNA availability is highly correlated with relative synonymous codon usage (RSCU),<sup>11,12</sup> it is generally accepted that frequently used codons are optimized (optimal) codons that are suitable for fast translation as well as favored by natural selection.<sup>13</sup>

The RSCU of eukaryotes has been adapted to the tRNA concentration during long term evolution.<sup>12</sup> For viruses that invade the hosts, the viral codon usage is usually uncorrelated with host codon usage. It means that viral genes would be less capable of utilizing the tRNA pool and have lower translation efficiency. Indeed, manipulating codon usage has already been used as a strategy to control viral gene expression.<sup>14</sup> On the contrary, from the virus perspective, a wise way to efficiently invade the host is to make its codon usage similar to that of the host transcriptome.<sup>1</sup> Since RSCU is measured by the expressed genes in a particular tissue/sample, the different tissues of the same species could have distinct RSCU spectrum. A previous literature found an interesting co-evolution pattern between virus and host.<sup>15</sup> When the codon usage of virus is similar to that of the host, the translation efficiency of host genes would be suppressed so that the virus gains an advantage due to translation selection. However, the above analyses were majorly conducted at unicellular level (yeast and human cell), which does



not have a problem of tissue-specific gene expression. In reality, the infection of SARS-CoV-2 is also selective, reminding us that we should analyze different tissues respectively. Analogous conclusion was proposed based on proteomic evidence that although the GC content of SARS-CoV-2 is poorly correlated with human genome (possibly for avoiding human immune response<sup>16</sup>), the human genes with similar codon usage to SARS-CoV-2 were down-regulated upon infection.<sup>17</sup>

It is reported that although the codon usage of SARS-CoV-2 is not correlated with RSCU of human genome, it is correlated with codon usage of lung expressed genes.<sup>1</sup> In other words, SARS-CoV-2 is adapted to human lungs. This observation nicely explains why SARS-CoV-2 could be rampant in the lung environment. However, question comes that why other viruses do not have this advantage to infect human lungs? To answer this question, there are 2 steps to be accomplished. First is to look into multi-tissue transcriptome data to confirm that the codon usage of SARS-CoV-2 is really most similar to that of human lungs (compared with other human tissues). Next, one should analyze the codon usage of a close coronavirus to see whether the other viruses are less correlated with human lungs.

RaTG13 is a coronavirus found in bats (*Rhinolophus affinis*). Although its evolutionary divergence with SARS-CoV-2 is still under debate,<sup>18</sup> it is widely accepted that RaTG13 is the closest sequence relative to SARS-CoV-2.<sup>19-21</sup> In this study, by retrieving the GTEx<sup>22</sup> RNA-seq data of 30 human tissues (from over 17000 individuals) and analyzing the RSCU of SARS-CoV-2, RaTG13, and human tissues, we draw the conclusion that codon usage of SARS-CoV-2 is most adaptive to human lungs, and that this adaptation event took place after the split from RaTG13. Our finding nicely explains the uniqueness of SARS-CoV-2 and demonstrates the evolutionary trajectory of codon usage in coronaviruses.

## Materials and Methods

### *Availability of data and materials*

The RNA-seq data of human tissues were downloaded from GTEx portal.<sup>22</sup> The expression level of each gene (measured by TPM, transcripts per million) is given in the downloaded file. We downloaded the SARS-CoV-2 and RaTG13 genome from the NCBI website (<https://www.ncbi.nlm.nih.gov/genome/>). The coding sequence of human genome was downloaded from the Ensembl website of version hg19 ([ftp://ftp.ensembl.org/pub/release-75/fasta/homo\\_sapiens/cds/](ftp://ftp.ensembl.org/pub/release-75/fasta/homo_sapiens/cds/)).

### *Relative synonymous codon usage (RSCU)*

The RSCU is defined as the frequency of a codon divided by the mean frequency of all its synonymous codons.<sup>23</sup> Frequency is calculated with a particular set of genes, which are usually the highly expressed genes (HEG) in a given sample. Under this definition, for an amino acid with  $N$  synonymous codons, the sum of RSCU values of all its codons will exactly be  $N$ . For a codon,  $RSCU > 1$

represents frequently used codons, or termed optimal codons.  $RSCU < 1$  represents non-optimal codons. Considering the arbitrary definition of HEG in a sample, it is also feasible to calculate RSCU by weighting each gene by expression level instead of discarding the lowly expressed genes. In this study, the “all genes weighted” RSCU is calculated by weighting all gene expression, and the “tissue-specific HEG” has excluded the top 500 genes with the highest mean expression in all tissues. The excluded 500 genes represent house-keeping genes.

### *Statistics*

The statistics like Pearson correlation coefficient (PCC) is performed in R language and environment.  $\Delta PCC = PCC_{SARS-CoV-2} - PCC_{RaTG13}$ , where  $PCC_{SARS-CoV-2}$  is the PCC of RSCU between SARS-CoV-2 and human samples, and  $PCC_{RaTG13}$  is the PCC of RSCU between RaTG13 and human samples.

## Results

### *RSCU shows no positive correlation when all genes are used*

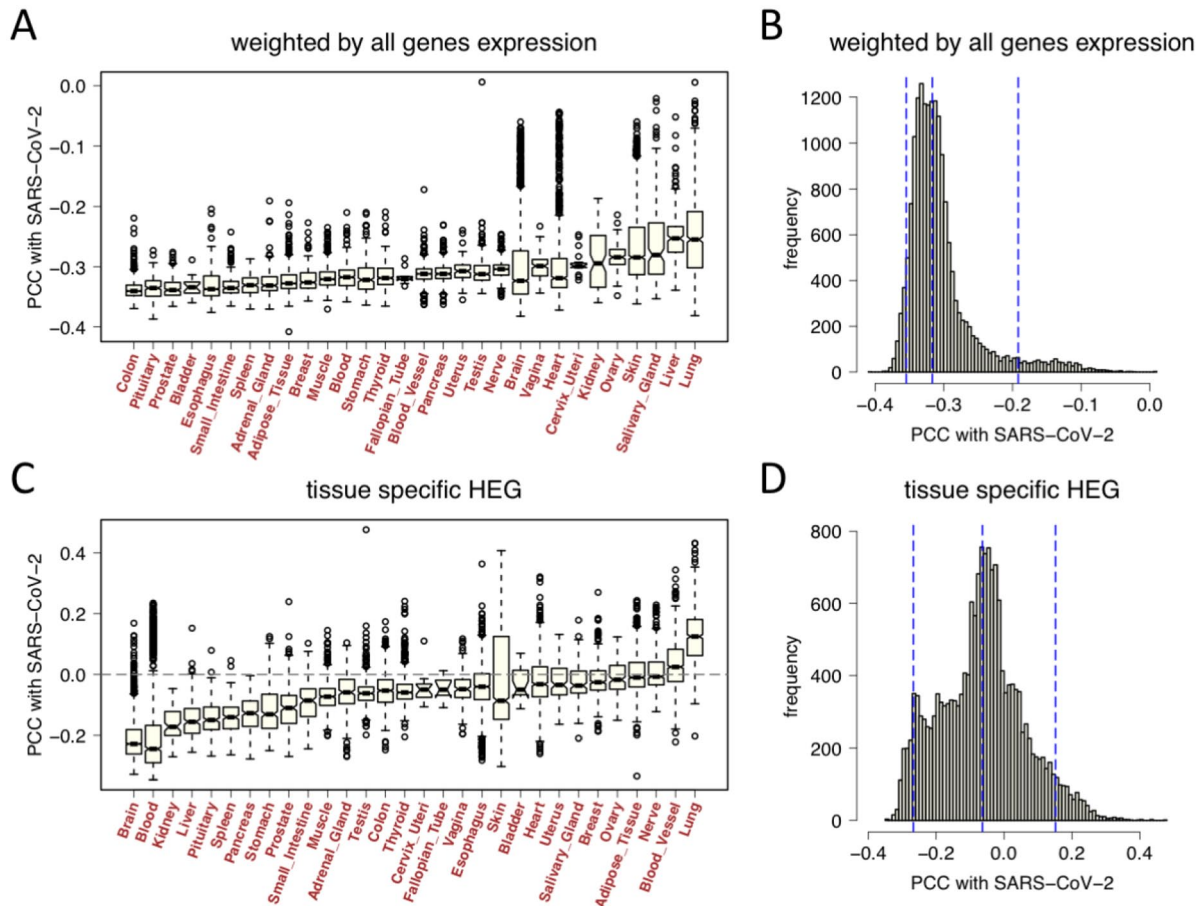
We retrieve GTEx transcriptome data of 30 human tissues (from over 17000 individuals) and intend to see the correlation of codon usage between human tissues and SARS-CoV-2. The RSCU values of SARS-CoV-2 are calculated with standard pipeline.<sup>1</sup> For human tissues, the RSCU value could be either calculated from the gene sequences of a set of arbitrarily selected HEG, or alternatively, we could weight each gene by its expression level so that the arbitrary cutoff is avoided. Theoretically, the 2 approaches should produce similar RSCU values because the HEG have higher weights.

We first weight each gene by expression level and calculate the RSCU for each sample. Then we perform correlation analysis and calculate the Pearson correlation coefficient (PCC) between SARS-CoV-2 and every human sample. Totally 30 tissues are ranked by median PCC. While lung samples generally have the highest RSCU correlation with SARS-CoV-2 (Figure 1A), it is striking that the PCC is globally negative, with a median PCC = -0.32 and 95% quantile [-0.36, -0.19] for all human samples (Figure 1B).

Since we have calculated the human RSCU by weighting each gene by expression level, we wonder whether another approach could improve the correlation. We select the top 50% genes with higher expression levels and calculate RSCU from the raw sequences of these genes (not weighted by expression level). It turns out that the RSCU under this approach does not differ much from the weighted one, with median PCC = -0.31 and 95% quantile [-0.35, -0.19] when compared to SARS-CoV-2.

### *Tissue-specific HEG is suitable for calculating RSCU*

We have observed unexpected results where the RSCU values are negatively correlated between human and SARS-CoV-2. Given the fact that SARS-CoV-2 is very successful in invading



**Figure 1.** Pearson correlation coefficient (PCC) of RSCU between SARS-CoV-2 and human tissues. Each dot represents a human individual. Thirty human tissues are displayed separately and ranked by the median value. (A) The RSCU calculated by weighting each gene by expression level. Boxplots are shown for each tissue. (B) The RSCU calculated by weighting each gene by expression level. Histogram is shown for all observations with median =  $-0.32$ . The vertical lines also show the 95% quantile [ $-0.36, -0.19$ ]. (C) The RSCU calculated by tissue-specific HEG. Boxplots are shown for each tissue. (D) The RSCU calculated by tissue-specific HEG. Histogram is shown for all observations with median =  $-0.06$ . The vertical lines also show the 95% quantile [ $-0.26, 0.15$ ].

human lungs, there should be some features that are unique to lungs to allow SARS-CoV-2 to take advantage of them.

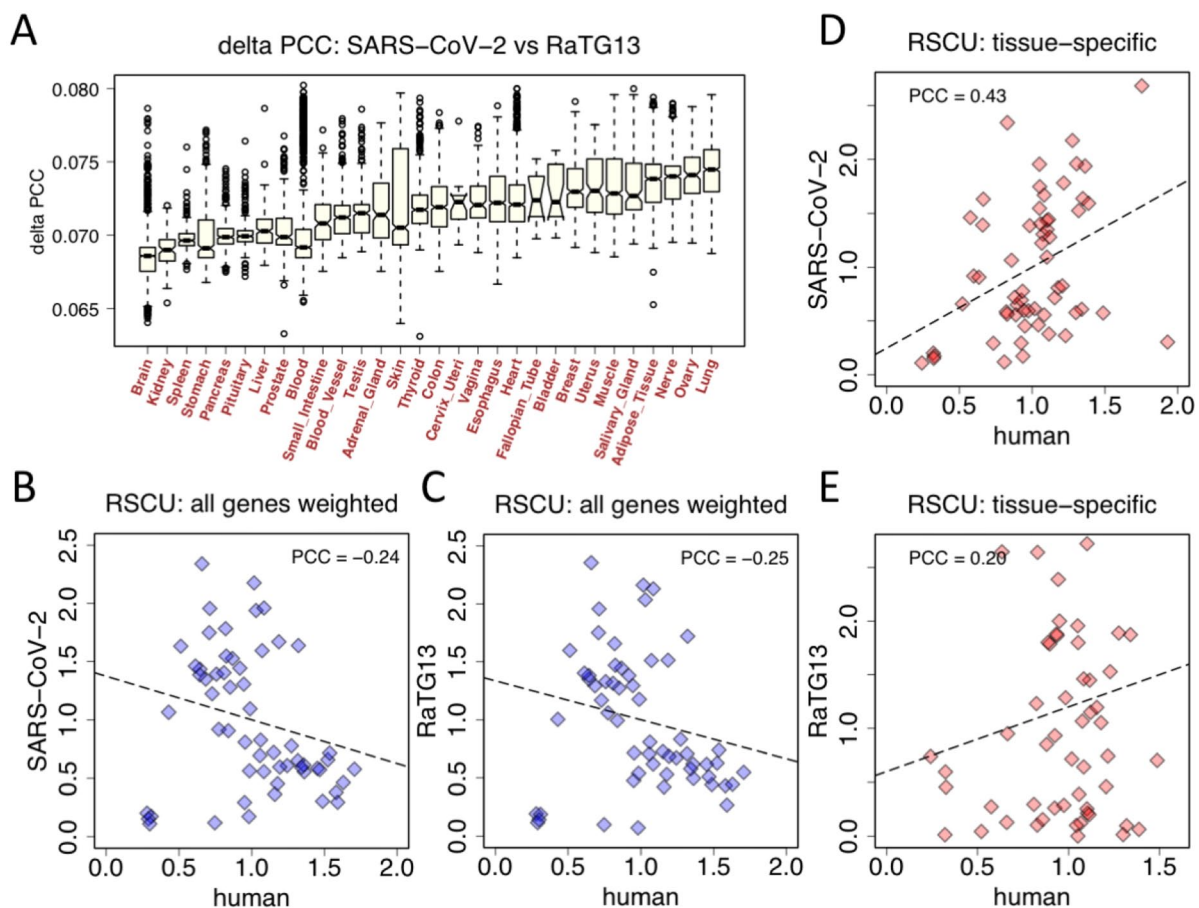
It is known that some house-keeping genes are constantly highly expressed in all tissues (such as constitutively expressed RPL family genes) while some other genes are selectively highly expressed in particular tissues (termed tissue-specific HEG). Our purpose is to try to reflect the tissue-specific codon usage, but the weighting method would obviously overrate the contribution of house-keeping genes and underestimate the importance of tissue-specific HEG. Among the tens of thousands of genes in the human genome, we rank the genes with mean expression across all GTEx samples, and empirically find that the top 500 genes are constantly highly expressed in most samples. Therefore, we regard the top 500 genes as house-keeping genes. Notably, the robustness of this cutoff would be discussed later.

In order to reflect the contribution of tissue-specific HEG, we remove the top 500 house-keeping genes and perform the weighting method to calculate RSCU. The PCC values between SARS-CoV-2 RSCU and human RSCU are calculated for each tissue respectively. Strikingly, the correlation is

no longer “all negative” but shows considerable fractions of positive correlation in each tissue (Figure 1C). Lung is still the most highly correlated tissue, and the median PCC across all tissues is  $-0.06$  with 95% quantile [ $-0.26, 0.15$ ] (Figure 1D). For all the lung samples, 93% of them show positive correlation with SARS-CoV-2 and 3.6% show significantly positive PCC. These results agree with our intuitive expectation.

To test the robustness of the tissue-specific HEG approach, we alternatively selected the top 300, top 800, and top 1000 genes as the house-keeping genes, and calculated the RSCU and PCC for the remaining genes weighted by expression level. The aforementioned patterns are robust. Lung is always the tissue showing the highest correlation with SARS-CoV-2, and 87% to 94% of the lung samples show significantly positive PCC.

Therefore, we conclude that the tissue-specific HEG, rather than the constitutively highly expressed house-keeping genes, are suitable for distinguishing the codon usage of different tissues. Lung is still the most highly correlated tissue with SARS-CoV-2. A large fraction of lung samples shows positive correlation with the codon usage of SARS-CoV-2.



**Figure 2.** Comparison between SARS-CoV-2 and RaTG13. Their correlation with human tissues is compared. (A) We define  $\text{delta PCC} = \text{PCC}_{\text{SARS-CoV-2}} - \text{PCC}_{\text{RaTG13}}$ , where  $\text{PCC}_{\text{SARS-CoV-2}}$  is the PCC of RSCU between SARS-CoV-2 and human samples, and  $\text{PCC}_{\text{RaTG13}}$  is the PCC of RSCU between RaTG13 and human samples. RSCU is calculated with tissue-specific HEG. (B) RSCU between SARS-CoV-2 and human sample Lung\_GTEX-QDT8-0926-SM-32PL2. The RSCU of all genes is used. (C) RSCU between RaTG13 and human sample Lung\_GTEX-QDT8-0926-SM-32PL2. The RSCU of all genes is used. (D) RSCU between SARS-CoV-2 and human sample Lung\_GTEX-QDT8-0926-SM-32PL2. The RSCU of tissue-specific HEG is used. (E) RSCU between RaTG13 and human sample Lung\_GTEX-QDT8-0926-SM-32PL2. The RSCU of tissue-specific HEG is used.

### *SARS-CoV-2 adapts to human after the split from RaTG13*

Given the observation that the codon usage of SARS-CoV-2 conforms to that of the human lung expressed genes, one would intuitively ask why only SARS-CoV-2 rather than other closely related coronaviruses has succeeded in the adaptation. To answer this question, we retrieve the sequence of RaTG13, the closest coronavirus to SARS-CoV-2. We calculate the correlation of RSCU between RaTG13 and human tissues where the tissue-specific HEG (excluding the top 500 house-keeping genes) are used. For a given human sample, we define  $\text{delta PCC} = \text{PCC}_{\text{SARS-CoV-2}} - \text{PCC}_{\text{RaTG13}}$ . A positive delta PCC means SARS-CoV-2 is more adaptive to human. We rank the human tissues with increasing delta PCC values. Amazingly, all samples have a positive delta PCC value, among which lungs generally have the highest delta PCC (Figure 2A). This result demonstrates that SARS-CoV-2 always has better codon adaptation to human tissues compared with RaTG13, and that the extent of difference peaks in lungs.

Note that the RSCU above is calculated with the tissue-specific HEG. We have proved that using this set of genes is

better than using all genes. To reemphasize this notion, we use a lung sample (ID: Lung\_GTEX-QDT8-0926-SM-32PL2) to display how the choice of gene sets would affect the RSCU and correlation. First, we calculate RSCU with all genes weighted by expression level in this lung sample. The PCC with SARS-CoV-2 (Figure 2B) and the PCC with RaTG13 (Figure 2C) are both negative and do not differ much. However, if we calculate RSCU with the tissue-specific HEG by excluding the top 500 house-keeping genes, then the PCC with both SARS-CoV-2 and RaTG13 has greatly improved but the PCC with SARS-CoV-2 (Figure 2D) is obviously higher than the PCC with RaTG13 (Figure 2E). This pattern, where SARS-CoV-2 is better correlated with human, suggests that the adaptation has taken place after SARS-CoV-2 split from RaTG13.

### Discussion

From the perspectives of structural biology, biochemistry, and molecular biology, there are already plenty of explanations for how SARS-CoV-2 could perfectly adapt to the human environment. However, several essential links are still missing. (1)

Why SARS-CoV-2 rather than other closely related coronaviruses could have this perfectness? (2) If the high expression of viral genes is not achieved, then the optimization of protein sequence/structure is almost futile (because quantity and quality are both important for the functioning of viral genes). The answer to these questions could be addressed in the light of evolution.<sup>24</sup> Conceivably, mutation is the source of natural selection and evolution.<sup>25–27</sup> Different mutated sequences were subjected to natural selection to eliminate the less adaptive ones.<sup>28</sup> The transition from ancient sequence to the present SARS-CoV-2 sequence reliably reflects this long-lasting evolution process.

By comparing the codon usage profiles of SARS-CoV-2, RaTG13, and numerous human tissues, we successfully find that SARS-CoV-2 always has better correlation with human tissues compared with its closest relative RaTG13. This observation nicely reconstructs the evolution history of SARS-CoV-2. After split from RaTG13, SARS-CoV-2 has optimized its codon usage to become more similar to that of human tissues, especially for lungs. This allows SARS-CoV-2 to utilize the resources (like tRNA pools) of human cells more efficiently, and explains why SARS-CoV-2 is so unique compared with other viruses. The adaptive codon usage of SARS-CoV-2 facilitates the viral RNA translation. A high translation rate of viral genes guarantees the high amount of viral proteins, which serves as the prerequisite of successful viral invasion and proliferation. To the host, the optimization of viral sequence is undesired, but for the virus itself, this codon optimization might be a milestone in evolution. Interestingly, the previous study on co-evolution of host-virus also found that the codon usage correlation is higher between virus and symptomatic hosts (compared to other natural hosts),<sup>15</sup> this simply support our conclusion that SARS-CoV-2 has optimized the codon usage after splitting from RaTG13.

We also emphasize that the correlation of codon usage between host and parasite should be investigated tissue by tissue, that is, taking the tissue specific gene expression into account, rather than using the whole reference genome. As we have shown, different human tissues exhibit completely different correlation with the virus. SARS-CoV-2 is highly correlated with human lungs, indicating its perfect adaptation to lung environment. The whole genome sequence (of a host species) is used only when there are no high-quality tissue-specific transcriptome data available.

Note that codon optimization is only a necessary but not sufficient requirement for successful viral invasion. The codon optimization ensures a relatively high level of viral gene expression. However, the expressed viral protein should also be structurally optimal to invade the host cells. The natural selection on codon optimization is exerted on synonymous mutations, while the selection on protein sequence/structure is exerted on missense mutations. One could conclude that any mutations in the viral coding sequence are potentially subjected to natural

selection. Therefore, under this global pandemic, any single mutation in the SARS-CoV-2 sequence should not be ignored since it might change the viral adaptiveness and even virulence.

## Conclusions

The codon usage of SARS-CoV-2 has adapted to human lungs to allow fast decoding and translation. This adaptation probably took place after SARS-CoV-2 split from RaTG13 because RaTG13 is less perfectly correlated with human. This finding depicts the trajectory of adaptive evolution from ancestral sequence to SARS-CoV-2, and also well explains why SARS-CoV-2 rather than other viruses could perfectly adapt to human lung environment.

## Acknowledgements

We thank the members in our group that have given suggestions to our project. At this SARS-CoV-2 time we should especially thank all the medical workers fighting against SARS-CoV-2.

## Author Contributions

Conceptualization: W.J.; Methodology: Y.Z., X.J., H.W., Y.M., and X.Y.; Software: Y.Z.; Validation: W.J.; Forma analysis: Y.Z.; Investigation: Y.Z.; Resources: Y.Z.; Data curation: Y.Z.; Writing—original draft preparation: Y.Z.; Writing—review and editing: W.J.; Visualization: Y.Z.; Supervision: W.J.; and Project administration: B.Y. All authors have read and agreed to the published version of the manuscript.

## Data Accessibility

The RNA-seq data of human tissues were downloaded from GTEx portal.<sup>22</sup> The expression level of each gene (measured by TPM, transcripts per million) is given in the downloaded file. We downloaded the SARS-CoV-2 and RaTG13 genome from the NCBI website (<https://www.ncbi.nlm.nih.gov/genome/>). The coding sequence of human genome was downloaded from the Ensembl website of version hg19 ([ftp://ftp.ensembl.org/pub/release-75/fasta/homo\\_sapiens/cds/](ftp://ftp.ensembl.org/pub/release-75/fasta/homo_sapiens/cds/)).

## ORCID iDs

Wenqing Jiang  <https://orcid.org/0000-0002-9605-5799>

Bin Yin  <https://orcid.org/0000-0002-1762-7585>

## REFERENCES

- 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*. 2020;295:1537–1546.
- Gilbert C, Tengs T. No species-level losses of s2m suggests critical role in replication of SARS-related coronaviruses. *Sci Rep*. 2021;11:16145.
- 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*. 2021;296:113–118.
- Seward EA, Kelly S. Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome Biol*. 2016;17:226.
- Seward EA, Kelly S. Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria. *Genome Biol*. 2018;19:102.

6. Zhao S, Song S, Qi Q, Lei W. Cost-efficiency tradeoff is optimized in various cancer types revealed by genome-wide analysis. *Mol Genet Genomics*. 2021;296:369-378.
7. Taghinezhad-S S, Razavilar V, Keyvani H, Razavi MR, Nejadstarrari T. Codon optimization of Iranian human papillomavirus type 16 E6 oncogene for *Lactococcus lactis* subsp *cremoris* MG1363. *Future Virol*. 2017;12:499-511.
8. Sorensen MA, Kurland CG, Pedersen S. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol*. 1989;207:365-377.
9. Yu CH, Dang Y, Zhou Z, et al. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell*. 2015;59:744-754.
10. dos Reis M, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*. 2004;32:5036-5044.
11. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985;2:13-34.
12. Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol*. 2001;53:290-298.
13. Yu Y, Li Y, Dong Y, Wang X, Li C, Jiang W. Natural selection on synonymous mutations in SARS-CoV-2 and the impact on estimating divergence time. *Future Virol*. Published online May 19, 2021. doi:10.2217/fvl-2021-0078
14. Baker SF, Nogales A, Martínez-Sobrido L. Downregulating viral gene expression: codon usage bias manipulation for the generation of novel influenza A virus vaccines. *Future Virol*. 2015;10:715-730.
15. Chen F, Wu P, Deng S, et al. Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection. *Nat Ecol Evol*. 2020;4:589-600.
16. Roy A, Guo F, Singh B, et al. Base composition and host adaptation of the SARS-CoV-2: insight from the codon usage perspective. *Front Microbiol*. 2021;12:548275.
17. Alonso AM, Diambra L. SARS-CoV-2 codon usage bias downregulates host expressed genes with similar codon usage. *Front Cell Dev Biol*. 2020;8:831.
18. Li Y, Yang X, Wang N, et al. Pros and cons of the application of evolutionary theories to the evolution of SARS-CoV-2. *Future Virol*. 2020;15:10.2217/fvl-2020.
19. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect Genet Evol*. 2020;79:104212.
20. Zhou P, Yang XL, Wang XG, et al. Addendum: a pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;588:E6-273.
21. Li Y, Yang X, Wang N, et al. The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. *Future Virol*. 2020;15:10.2217/fvl-2020.
22. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45:580-585.
23. Sharp PM, Li WH. The codon adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987;15:1281-1295.
24. Dobzhansky T. Nothing in biology makes sense except in the light of evolution. *Am Biol Teach*. 1973;35:125-129.
25. Chang S, Li J, Li Q, Yu CP, Xie LL, Wang S. Retrieving the deleterious mutations before extinction: genome-wide comparison of shared derived mutations in liver cancer and normal population. *Postgrad Med J*. Published online April 9, 2021. doi:10.1136/postgradmedj-2021-139993
26. 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*. Published online July 19, 2021. doi:10.1007/s13353-021-00653-0
27. Jiang Y, Cao X, Wang H. Comparative genomic analysis of a naturally born serpentine pig reveals putative mutations related to limb and bone development. *BMC Genomics*. 2021;22:629.
28. Zhang Y, Jiang W, Li Y, et al. Fast evolution of SARS-CoV-2 driven by deamination systems in hosts. *Future Virol*. Published online September 10, 2021. doi:10.2217/fvl-2021-0181