

Host Adaptation of Codon Usage in SARS-CoV-2 From Mammals Indicate Natural Selection

Yanan Fu

Hubei University of Medicine

Yanping Huang

Hubei University of Medicine

Jingjing Rao

Hubei University of Medicine

Feng Zeng

Hubei University of Medicine

Ruiping Yang

Hubei University of Medicine

Zhixin Liu

Hubei University of Medicine

Weixing Du

Hubei University of Medicine

Long Liu (✉ liulong2015@outlook.com)

Hubei University of Medicine <https://orcid.org/0000-0002-1265-7603>

Research Article

Keywords: SARS-CoV-2, substitution, codon usage, host adaptation, receptor

Posted Date: December 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1125942/v1>

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

[Read Full License](#)

Abstract

The outbreak of COVID-19, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections, spread across hosts from humans to animals, transmitting particularly effectively in mink. How SARS-CoV-2 selects and evolves in the host, and the differences in the evolution of different animals are still unclear. To analysis the mutation and codon usage bias of SARS-CoV-2 in infected humans and animals. The SARS-CoV-2 sequence in mink (Mink-SARS2) and binding energy with receptor were calculated compared with human. The relative synonymous codon usage of viral encoded gene was analyzed to characterize the differences and the evolutionary characteristics. A synonymous codon usage analysis showed that SARS-CoV-2 is optimized to adapt in the animals in which it is currently reported, and all of the animals showed decreased adaptability relative to that of humans, except for mink. The neutrality plot showed that the effect of natural selection on different SARS-CoV-2 sequences is stronger than mutation pressure. A binding affinity analysis indicated that the spike protein of the SARS-CoV-2 variant in mink showed a greater preference for binding with the mink receptor ACE2 than with the human receptor, especially as the mutation Y453F and N501T in Mink-SARS2 lead to improvement of binding affinity for mink receptor. In summary, mutations Y453F and N501T in Mink-SARS2 lead to improvement of binding affinity with mink receptor, indicating possible natural selection and current host adaptation. Monitoring the variation and codon bias of SARS-CoV-2 provides a theoretical basis for tracing the epidemic, evolution and cross-species spread of SARS-CoV-2.

Introduction

SARS-CoV-2 is a β -coronavirus that emerged in 2019 and spread worldwide, leading to an ongoing global pandemic [1, 2]. As of November 29th 2021, the number of infected cases reached 261 million, and more than 5.2 million deaths have occurred (Johns Hopkins University statistics; <https://coronavirus.jhu.edu/map.html>). SARS-CoV-2 has a single-stranded positive-sense RNA genome containing 29,903 nucleotides and consisting of 11 open reading frames (ORFs) encoding 27 proteins [3]. The S glycoprotein is a fusion viral protein that functions in recognition of the host receptor ACE2 [4].

There is a broad host spectrum because SARS-CoV-2 binds a receptor common to humans and animals[5]. To date, the following animals have been reported to be susceptible to infection: cats, dogs, tigers, lions, ferrets, and mink [6-10]. SARS-CoV-2 infection of pets, including cats and dogs [8, 10], was the earliest reported animal infections in the epidemic. Later, in a report on SARS-CoV-2 infection in tigers, lions, and human keepers in a New York zoo, epidemiologic and genomic data indicated human-to-animal transmission [11]. Other animals, including snow leopards and gorillas, tested positive for SARS-CoV-2 after showing signs of illness [12, 13]. It is noteworthy that a study from The Netherlands reported the spread of SARS-CoV-2 from humans to mink and from mink back to humans in mink farms [14]. Eighty-eight mink and 18 staff members from sixteen mink farms were confirmed to be infected with SARS-CoV-2 as determined by high throughput sequence analysis. The adaptation of SARS-CoV-2 to bind the mink receptor and the viral evolution in the mink host are worthy of further study.

Codon usage bias refers to differences in the frequency of occurrence of synonymous codons during protein translation, which differs between hosts [15]. Codon usage bias is dominated by selection pressure in some bacteria, while it was driven by mutation in virus, such as Ebola virus [16, 17]. Viruses differ markedly in their specificity toward host organisms, and the analysis of the viral genome structure and composition contributes to the partial understanding of virus evolution and adaptation in the host [14, 18]. Further exploration of the codon usage pattern of SARS-CoV-2 in different hosts, especially the codon architecture of the *Spike* gene, indicate host adaptation related to cross-species transmission.

Surveillance of the substitution and selection of the SARS-CoV-2 genome is important for the study of viral evolution and for tracking viral transmission. In particular, study of the *Spike* gene helps to evaluate the immunization effect of vaccinations and to adjust the vaccine design in a timely manner. This study focuses on the divergence of the SARS-CoV-2 genome composition and codon usage in human and animal hosts to investigate the natural selection that might play a role in virus evolution, adaptability, and transmission.

Materials And Methods

SARS-COV-2 sequences and data collection

A total 209 SARS-COV-2 genome sequences from humans, cats, dogs, tigers, lions, hamster and minks were used for the genetic analysis (The strains information was recorded in the Supplementary Table S1). All the genomic sequences selected by the hosts were obtained from the GISAID database (<https://www.gisaid.org/>). Isolate Wuhan/WIV04 was used as the reference strain.

Evolutionary analysis

209 SARS-COV-2 genomes were used for phylogenetic analysis. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [19]. The tree with the highest log likelihood (-42442.50) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.0500)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 29903 positions in the final dataset. Evolutionary analyses were conducted in MEGA-X [20].

Identification of mutations

The sequences were aligned using MEGA-X, and the single nucleotide polymorphisms were analyzed using the SNIPlay pipeline by uploading aligned Fasta format file (https://sniplay.southgreen.fr/cgi-bin/analysis_v3.cgi)[21]. All the sequences including coding regions, 5'UTR and 3'UTR were used for the analysis.

Estimation of nonsynonymous and synonymous substitution rates

The number of nonsynonymous substitutions per synonymous site (dN) and the number of synonymous substitutions per nonsynonymous site (dS) for each coding site were calculated using the Nei–Gojobori method (Jukes–Cantor) in MEGA-X. The Datamonkey adaptive evolution server (<http://www.datamonkey.org>) was used to identify sites where only some of the branches have undergone selective pressure. The mixed-effects model of evolution (MEME) and fixed effects likelihood (FEL) approaches were used to infer the nonsynonymous and synonymous substitution rates.

Codon usage analysis

The codon adaptation index (CAI) of a given coding sequence was calculated using R script [22]. A CAI analysis of those coding sequences from different hosts was performed using DAMBE 5.0 and the CAI [23, 24]. The codon usage data of different hosts were retrieved from the codon usage database (<http://www.kazusa.or.jp/codon/>), and the relative synonymous codon usages (RSCUs) were analyzed using MEGA-X software. Animal-SARS2 means SARS-CoV-2 isolated from the indicted animals, the access IDs from mink, cat, dog, tiger and lion are MT457401.1, MT747438.1, MT215193.1, MT704316.1 and MT704312.1. Bat-CoV refers to RaTG13 (GenBank: MN996532.2) and the corresponding host (bat), Pangolin-COV refers to pangolin coronavirus (GenBank: QLR06867.1).

Neutral evolution analysis

Neutrality plot analysis was constructed to determine the codon bias influenced by natural selection. It reflected the neutrality effect of directional mutation pressure equilibrium in shaping the codon usage bias [25]. Mostly the GC3 position has an equal probability of appearing A/T and G/C nucleotides. There will be variation between GC12 against GC3 regression values due to the directional mutational pressure. Here GC12 against GC3 were plotted with a regression line.

Spike protein sequence and structure reconstruction

The crystal structure of the SARS-CoV-2 receptor-binding domain (RBD) in complex with human ACE2 (PDB ID: 6M0J) was used for structural analysis. Structures of ACE2 and the viral spike from mink were constructed by the SWISS model server (<https://swissmodel.expasy.org/>). Comparisons of the predicted protein structures and pairwise comparisons were analyzed using PyMOL software.

Molecular dynamics

For the binding free energy (E), we simulated the minimized annealing energy through molecular dynamics (MD) simulation in YASARA [26]. We performed three iterations of energy minimization for the set of wild-type residues in the viral spike protein bound with human ACE2 and mutant residues in the Mink-SARS2 spike with mink ACE2. The relative binding energy (ΔE) are reported as the mean and standard deviation values across three replicates.

Selective coefficient index

The selection coefficient index (S) of all SARS-CoV-2 codons was estimated by the FMutSel0 model in the program CODEML (PAML package) [27]. The fitness parameter of the most common residues at each location is fixed to 0, while the other fitness parameters are limited to $-20 < F < 20$.

Plasmids construction and cell culture

The gene fragment of RBD of SARS-COV-2 (NCBI ID: MN996528.1) inserted into pCAGGS plasmid was donated by Prof. Jianguo Wu, RBD fragments of SARS-COV-2 and Mink-ACE2 (NCBI ID: MW269526.1) were synthesized by Genscript Inc. The mutants of RBD genes were constructed using the Mut Express II Fast Mutagenesis Kit (Vazyme, C214). The wild type and mutant genes were cloned into pCAGGS vector with a 6´ Histidine tags using the *EcoRI* and *XhoI* restriction sites. Mink-ACE2 fragment was inserted into vector pcDNA3.1-eGFP (named pcDNA3.1-mACE2-eGFP). The products were subsequently transformed into competent *E.coli* Top10 strain. Recombinant plasmids were verified by Next-generation sequencing. Plasmid pcDNA3.1-hACE2-eGFP expressing hACE2 fused with eGFP was ordered from Fubio Ltd (MC_0101086). HEK293T and BHK-21 cells were cultured by high-sugar DMEM medium in 5% CO₂ atmosphere of 37°C incubator.

Protein expression and purification

The recombinant RBDs of SARS-COV-2 mutants were expressed in CHO cells. Recombinant pCAGGS plasmids were transfected into CHO cells in 245 mm dishes according to the manufacturer's recommendations; the supernatant was collected and centrifuged after 5 days. The soluble proteins were purified by using HisPur™ Ni-NTA Resin (Thermo Scientific, 88221). Purified proteins were eluted in buffer consisting of 20 mM Tris-HCl (pH 8.0) and 150 mM NaCl.

Flow cytometry

The plasmids pcDNA3.1-hACE2-GFP and pcDNA3.1-mACE2-eGFP were transfected into BHK-21 cells using Liposome 6000 (Beyotime, C0526) according to the manufacturer's instructions. The cells expressing hACE2-GFP and mACE2-GFP were collected after transfection for 24 h and suspended in the PBS buffer. Then, the cells were incubated with the purified His-tagged RBDs at a final concentration of 30 µg/mL at 37°C for 30 min. After washed by PBS twice and incubation with anti-His/APC antibodies (1:5000), the cells were determined using a BeckMan CytoFLEX and the data were analyzed using FlowJo V10 software.

Statistical analysis and mapping

Statistical analyses were performed using ANOVA followed by Turkey's post hoc test (Fig 2C&2F) or Student's t-test (Fig 3E), and the data were considered significantly different if the p -value was less than 0.05. The SPSS 20.0 software was used for regression curve fit. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns. means no significant. The figures were mapped by the software PRISM GraphPad 5.0.

Results

Sequence and analysis of SARS-CoV-2 isolated from animals

As of Jun 20th 2021, more than 2.53 million SARS-CoV-2 genome sequences had been uploaded to the GISAID database. It is important to study the mutation rates and selective pressures on the SARS-CoV-2 genome during the spread of the epidemic. The evolutionary entropy increased at specific sites in the whole genome of SARS-CoV-2. In addition to humans, SARS-CoV-2 infects other animals (Fig. 1A) and evolves in these animals. A phylogenetic tree was reconstructed based on animal-derived whole genome consensus sequences compared with the SARS-CoV-2 human isolate WIV04 (Fig. 1B). Most SARS-CoV-2 clade isolates from the same animal clustered together, and the same clade contained sequences from all the mink regardless of their geographic region.

The cluster of SARS-CoV-2 from mink (Mink-SARS2) has more substitutions compared to the reference sequence WIV04 (Supplementary Table S2), and the substitutions of cytidine in Mink-SARS2 account for nearly 50% of the substitutions, while in other animals, cytidine accounts for only 30% of the substitutions (Fig. 1C). The substitution of adenine in SARS-CoV-2 in other animals is threefold higher than that in Mink-SARS2. To track how the substitutions occurred in the Mink-SARS2 genome, we recorded all the mutations in the Mink-SARS2 genome in reference to the WIV04 genome. The results in Fig. 1D & 1F show that the cytidine-to-uracil transition occurred more than 40% of the time and was eightfold higher than the uracil-to-cytidine substitution. Notably, the substitutions of guanine and adenine were more than threefold higher in nonsynonymous mutations than in synonymous mutations (Fig. 1E).

Mutational spectra of Spike protein in human and animal samples

The variation in the spike gene was evident when all the included sequences isolated from humans and animals were recorded in our study, which led to the identification of a number of highly variable residues, including the relatively high-frequency amino acid variation sites. C-to-U substitutions were scattered throughout the SARS-CoV-2 genome and accounted for 24.06% of the substitutions in the spike gene in all epidemic strains analyzed as of February 2, 2021 (Fig. 2A). Because of the widespread transmission of D614G (GAT>GGT), A-to-G substitutions accounted for 56.12% of all the monitored strains. The result of $dN-dS$ indicates the natural selection for mutations in these specific sites in the *spike* gene ($dN-dS > 0$ indicates positive selection, and $dN-dS < 0$ indicates purification selection). Fig. 2B shows that sites 222, 262, 439 and 614 were exposed to strong positive selection pressure, while positions 294, 413, 1018 and 1100 were subjected to purifying selection during evolution.

CAI was used to quantify the codon usage similarities between different coding sequences based on a reference set of highly expressed genes [28]. To clarify the optimization of SARS-CoV-2 in different hosts, we calculated the average CAI of the SARS-CoV-2 whole genome (Fig. 2C). Interestingly, SARS-CoV-2 in bat hosts has a higher value of CAI relative to humans, while dogs had an obviously decreased CAI value

compared to humans (Fig. 2C). The ENC-plot analysis helps us to further understand the influence factors of SARS-CoV-2 codon usage bias. The result (Fig. 2D) showed most sources of SARS-CoV-2 located slightly below the standard curve ($R^2=0.7563$, $P=0.005$), indicating that the codon usage bias is not only affected by mutation factors, but also affected by evolutionary pressure. This is consistent with Rupam's report [25]. The neutrality plot showed the linear regression coefficient of SARS-CoV-2 sequences was -0.1161 (Fig. 2E), indicating that the pressure for codons at first and second position are different from those at third position in the evolutionary process. The first and second position of codons are mainly affected by mutation and third position is mainly affected by selection. Considering codon usage in the spike gene in different hosts, Fig. 2F shows that pangolins, cats, dogs, tigers, and lions all had a lower CAI value than humans. The bias of codon usage in the spike mutants are shown in Supplementary Table S3. These results indicated that SARS-CoV-2 optimized codon usage to adapt to the animals in which infection has been reported, but all of them showed a downward trend in adaptability relative to humans except for mink.

SARS-CoV-2 spike mutant shows greater preference for binding with mink receptor

The spike interacts with human and mink ACE2 through the amino acids H34Y, Y41, L79H, M82T and G354R (Fig. 3A&3B), which form electric charge attraction and hydrophobic interactions with residues Asn439, Tyr453, Phe486 and Asn501 on spike. To distinguish the differentiation of receptor sequences between different animals and humans, the ACE2 amino acid sequences in humans, mink, ferrets, tigers, cats, and dogs were aligned (Fig. 3C). The results showed that the critical mutations H34Y, L79H and G354R appear in mink and ferret ACE2 (Fig. 3B). On the other hand, viral variation is another important factor that should also be considered when analyzing infection differences between animals and humans. Corresponding to the contact residues on the receptors, alignment of the viral sequence contacts of ACE2 on spike indicated that residues binding receptor are conserved (Fig. 3C), while residues at site 453, which interact with those at position 34 in ACE2 (Fig. 3D), showed a higher binding affinity for F453-Y34 in mink than for Y453-H34 in humans (Fig. 3E). The interaction between T501 and R354 showed increased binding energy with mink receptor than humans (Fig. 3E). The flow cytometry results showed that Y453F, F486L and N501T mutants lead to the stronger overt fluorescent shift of cells with mink ACE2 than wild type RBD, while only F486L and double mutant Y453F&F486L have a similar function with binding to human receptor (Fig. 3F). These variations indicate that SARS-CoV-2 spike shows a greater preference for binding with the mink receptor than humans ACE2 after this mutation occurs.

Codon usage and fitness analysis of each amino acid in SARS-COV-2 encoded proteins

Amino acid substitutions within the SARS-CoV-2 Spike RBM may have contributed to host adaption and cross-species transmission. N439K, S477N and N501Y were the most abundant variations throughout the RBM regions (Fig. 4A). N439 does not bind directly with ACE2 but functions in the stabilization of the 498–505 loop [21], but the N439K substitution is absent in animal CoVs (Fig. 3C). Previous computational analysis combined with entropy analysis of the spike showed that S477N may have

decreased stability compared with the wild type [29]. Since human SARS-CoV-2 and Mink-SARS2 do not show very different codon usage bias (Fig. 2C) and because viral codon bias depends on the host, we compared the codon usage frequency of SARS-CoV-2 and SARS-CoV (Fig. 4B), for which ferrets are common hosts. Because substitutions N501T (AAU>ACU) in mink and N501Y (AAU>UAU) in humans occurred nonsynonymously in the first and second positions and since these substitutions had a lower frequency than other noted substitutions (Fig. 1F & 2A), further study on the relationship of these substitutions is needed. The results of selective coefficient index in Fig. 4C show the differences of relative fitness in the SARS-CoV-2 codons, CGA and CGG have the high fitness score in all codon-specific estimates, and the fitness of T (ACU) was close to Y (UAU). In addition, it was observed that among the 12 ORFs encoded by SARS-CoV-2, the codons encoding different ORFs of SARS-CoV-2 possess diversity RSCU values (Supplementary Table S4). It is worth noting that UCA encoding Ser have excessive bias in ORF7b (Fig. 4D), and AGG encoding Arg overbiased in ORF6 (Supplementary Table S4).

Discussion

Tracking animal variants arising from human contact or produced from animal bodies is an interesting topic and allows for better understanding of the evolutionary mechanism and selection fitness of SARS-CoV-2 in the host. Regardless of the probability of contact between different animals and SARS-CoV-2, the transmission of the virus between animals is inseparable from susceptibility and host adaptability. Some viruses with codon bias observed in the process of adaptation are related to their hosts, including Rotavirus, HPV, and Marburg virus et al. [30-32]. Mink were the first extensively farmed species to be affected by the COVID-19 epidemic, indicating that mustelids, including mink and ferrets, are more sensitive to SARS-CoV-2 than other animals [33]. Several mink farms in The Netherlands, Denmark, USA, and Spain all reported infection cases [34-38], indicating mink-to-mink and mink-to-human (Netherlands, Denmark) transmission. Other animals, including tigers and lions, are also susceptible to SARS-CoV-2 infection [39]. Hence, comparison of the susceptibility and the natural evolutionary pressure in different hosts for SARS-CoV-2 is meaningful and helpful for clarifying the host adaptation mechanisms and monitoring the epidemic.

Viral genes and genomes exhibit varying numbers of synonymous codons depending on the host [40]; hence, the codon usage bias of the virus has a strong relationship with its host. Studying the preferred synonymous codon usage and base substitutions helps to provide an understanding of the codon patterns of the virus in relation to their hosts and in relation to viral genome evolution. The convergence effect of virus codon preference on the host is widely recognized and is also one of the main natural selection forces for the coevolution of viruses and hosts [41]. In this study, we compared the codon bias of SARS-CoV-2 in mink with that of SARS-CoV in ferrets. Residues threonine (T) and tyrosine (Y) had similar codon biases in SARS-CoV-2 and SARS-CoV (Fig. 4B), which both have the capability to infect mink and ferrets. The N501T variation mostly appeared in mink, while the N501Y mutation present only in humans cannot be explained from the perspective of codon bias and indicates that these two variations belong to two separate lineages.

The WebLogo diagram in Fig. 4B shows that SARS coronaviruses preferentially have U- or A-ending codons. This is consistent with a previous report [25], and the G or C nucleotides in the third position of the preferred SARS-CoV-2 codons are not well represented. This feature may lead to an imbalance in the tRNA pool in infected cells, resulting in reduced host protein synthesis. The substitution rate of C-to-U was the highest in most of the reported sequences in animal species (Fig. 1C). This may be because the surrounding context of cytidine in the sequence strongly influences the possibility of its mutation to U [42]. In the mink sequences, we observed an 8-fold increase in C-to-U substitution compared with the U-to-C substitution, which was higher than the reported 3.5-fold increase in mink [36][36][37][38], suggesting host adaptation of SARS-CoV-2 in mink over time and the ongoing outbreaks in multiple mink farms. In mink, the variations in G and A with nonsynonymous substitutions were higher than those with synonymous substitutions, which needs to be further analyzed. In addition, the sequences of other animal-CoVs are limited, such as those in the dogs and lions in the GISAID database, which is a limiting factor for comparison of base substitutions.

CAI was used to measure the synonymous codon similarities between the virus and host coding sequences. For each animal source of the SARS-CoV-2 sequence, we calculated the average genome and *spike* gene values in the CAI (Fig. 2C & 2D). Bat-CoV (RaTG13) and SARS-CoV-2 (from humans) had higher CAI values, which indicates that the viruses adapt to their hosts (bat and human) with optimized or preferred chosen codons, while the dog source of SARS-CoV-2 had lower CAI values, suggesting that SARS-CoV-2 adapts to dogs with random codons. This finding was consistent with the conclusion that, compared to dogs, humans are favored hosts for adaptation [43]. The whole genome or spike sequence in Mink-SARS2 had a similar substitution level to human SARS-CoV-2, pointing to the ongoing adaptation of SARS-CoV-2 to the new host and using the preferred chosen codons.

The spike protein is critical for virus infection and host adaptation. We observed that three nonsynonymous mutations in the RBM domain, Y453F, F486L and N501T, independently emerged but were rarely observed in human lineages; these residues are directly involved in contact with the surface of the S-ACE2 complex and therefore are relevant to new-host adaptation. Other mutations within the RBM domain should also be monitored to prevent viral transmission and to prevent the virus escaping from host antibody immunity. Such as the B.1.351 SARS-CoV-2 variant carrying the E484K and N501Y mutations, which can bind to K31 and Y41 on ACE2 to promote virus entry into the cell [44], had the ability to re-infect the recovered patients or people who have been vaccinated[45].

The current data show that the preventive effect of vaccines on Delta variation was reduced [46], but it can still prevent the occurrence of severe cases. The codon bias and frequency of mutants should be further evaluated to optimize the vaccine and block transmission.

Declarations

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

YNF, FZ, YPH, JJR and LL contributed to the design of experiments. YNF, YPH, JJR, FZ, RPY, YNF and ZXL contributed to the conduction of experiments. JJR, RPY and FZ contributed to the reagents. WXD and ZXL contributed to the analyses of the data. LL, WXD and ZXL contributed to the writing the paper. WXD and LL contributed to the editing the paper.

FUNDING

This work was supported by Natural Science Foundation of China (82002149 and 81902066), Foundation of Health Commission of Hubei Province (WJ2021M059), the Principle Investigator Program at Hubei University of Medicine (HBMUPI202102) and the Cultivating Project for Young Scholar at Hubei University of Medicine (2018QDJZR32). The funder has no function in the study design, data collection and analysis, manuscript writing and submission.

References

1. Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R (2020) Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 85:104445
2. Rabaan AA, Al-Ahmed SH, Haque S, Sah R, Tiwari R, Malik YS, Dhama K, Yattoo MI, Bonilla-Aldana DK, Rodriguez-Morales AJ (2020) SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. *Le infezioni in medicina* 28:174-184
3. Lopic I, Rogic D, Plebani M (2020) Erythrocyte sedimentation rate is associated with severe coronavirus disease 2019 (COVID-19): a pooled analysis. *Clinical chemistry and laboratory medicine* 58:1146-1148
4. Nikolettou V, Markaki M, Palikaras K, Tavernarakis N (2013) Crosstalk between apoptosis, necrosis and autophagy. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1833:3448-3459
5. Arai Y, Kawashita N, Hotta K, Hoang PVM, Nguyen HLK, Nguyen TC, Vuong CD, Le TT, Le MTQ, Soda K, Ibrahim MS, Daidoji T, Takagi T, Shioda T, Nakaya T, Ito T, Hasebe F, Watanabe Y (2018) Multiple polymerase gene mutations for human adaptation occurring in Asian H5N1 influenza virus clinical isolates. *Sci Rep* 8:13066
6. Venkatesh D, Bianco C, Nunez A, Collins R, Thorpe D, Reid SM, Brookes SM, Essen S, McGinn N, Seekings J, Cooper J, Brown IH, Lewis NS (2020) Detection of H3N8 influenza A virus with multiple mammalian-adaptive mutations in a rescued Grey seal (*Halichoerus grypus*) pup. *Virus evolution* 6:veaa016

7. Eaaswarkhanth M, Al Madhoun A, Al-Mulla F (2020) Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 96:459-460
8. Sit THC, Brackman CJ, Ip SM, Tam KWS, Law PYT, To EMW, Yu VYT, Sims LD, Tsang DNC, Chu DKW, Perera R, Poon LLM, Peiris M (2020) Infection of dogs with SARS-CoV-2. *Nature* 586:776-778
9. Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G, tenOever BR, Sanjana NE (2020) The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. [bioRxiv:2020.2006.2014.151357](https://doi.org/10.1101/2020.2006.2014.151357)
10. Takahiko Koyama DPaLP (2020) Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ* 98:495-504
11. Wang L, Mitchell PK, Calle PP, Bartlett SL, McAloose D, Killian ML, Yuan F, Fang Y, Goodman LB, Fredrickson R, Elvinger F, Terio K, Franzen K, Stuber T, Diel DG, Torchetti MK (2020) Complete Genome Sequence of SARS-CoV-2 in a Tiger from a U.S. Zoological Collection. *Microbiol Resour Announc* 9
12. DALY N (2021) Several gorillas test positive for COVID-19 at California zoo—first in the world.
13. Schaecher SR, Diamond MS, Pekosz A (2008) The Transmembrane Domain of the Severe Acute Respiratory Syndrome Coronavirus ORF7b Protein Is Necessary and Sufficient for Its Retention in the Golgi Complex. *Journal of virology* 82:9477-9491
14. Munnink BBO, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolsma P, Rietveld A, Brouwer M, Bouwmeester-Vincken N, Harders F, Hakze-van der Honing R, Wegdam-Blans MCA, Bouwstra RJ, GeurtsvanKessel C, van der Eijk AA, Velkers FC, Smit LAM, Stegeman A, van der Poel WHM, Koopmans MPG (2021) Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 371:172-177
15. Alwosaibai K, Alshaer W (2017) PAX2 maintains the differentiation of mouse oviductal epithelium and inhibits the transition to a stem cell-like state. *Oncotarget* 8:76881-76897
16. Chen Y, Shi Y, Deng H, Gu T, Xu J, Ou J, Jiang Z, Jiao Y, Zou T, Wang C (2014) Characterization of the porcine epidemic diarrhea virus codon usage bias. *Infection, Genetics and Evolution* 28:95-100
17. Cristina J, Moreno P, Moratorio G, Musto H (2015) Genome-wide analysis of codon usage bias in Ebolavirus. *Virus research* 196:87-93
18. Gharaibeh L, Elmadany N, Alwosaibai K, Alshaer W (2020) Notch1 in Cancer Therapy: Possible Clinical Implications and Challenges. *Molecular pharmacology* 98:559-576
19. Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, Teschendorff AE (2016) dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic acids research* 45:D812-D818
20. Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25:402-408
21. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J,

- Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395:497-506
22. Sharp PM, Li WH (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research* 15:1281-1295
 23. Taghiloo S, Aliyali M, Abedi S, Mehravaran H, Sharifpour A, Zaboli E, Eslami-Jouybari M, Ghasemian R, Vahedi-Larijani L, Hossein-Nattaj H, Amjadi O, Rezazadeh H, Ajami A, Asgarian-Omran H (2020) Apoptosis and immunophenotyping of peripheral blood lymphocytes in Iranian COVID-19 patients: Clinical and laboratory characteristics.
 24. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England journal of medicine* 382:727-733
 25. Dutta R, Buragohain L, Borah P (2020) Analysis of codon usage of severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) and its adaptability in dog. *Virus research* 288:198113
 26. Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, Chen L, Li M, Liu Y, Wang G, Yuan Z, Feng Z, Zhang Y, Wu Y, Chen Y (2020) Reduction and Functional Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). *Frontiers in immunology* 11:827
 27. Yang Z, Nielsen R (2008) Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol Biol Evol* 25:568-579
 28. Bian X-W, Team TC-P (2020) Autopsy of COVID-19 patients in China. *National Science Review* 7:1414-1418
 29. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z (2020) Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama* 323:1061-1069
 30. Kattoor JJ, Malik YS, Sasidharan A, Rajan VM, Dhama K, Ghosh S, Bányai K, Kobayashi N, Singh RK (2015) Analysis of codon usage pattern evolution in avian rotaviruses and their preferred host. *Infection, Genetics and Evolution* 34:17-25
 31. Ma YP, Liu ZX, Hao L, Ma JY, Liang ZL, Li YG, Ke H (2015) Analysing codon usage bias of cyprinid herpesvirus 3 and adaptation of this virus to the hosts. *Journal of Fish Diseases* 38:665-673
 32. Nasrullah I, Butt AM, Tahir S, Idrees M, Tong Y (2015) Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC evolutionary biology* 15:174
 33. Liu DX, Fung TS, Chong KK, Shukla A, Hilgenfeld R (2014) Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral research* 109:97-109
 34. Zhou Z, Ren L, Zhang L, Zhong J, Xiao Y, Jia Z, Guo L, Yang J, Wang C, Jiang S, Yang D, Zhang G, Li H, Chen F, Xu Y, Chen M, Gao Z, Yang J, Dong J, Liu B, Zhang X, Wang W, He K, Jin Q, Li M, Wang J (2020) Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. *Cell host & microbe* 27:883-890.e882

35. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Møller R, Panis M, Sachs D, Albrecht RA, tenOever BR (2020) SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. *bioRxiv:2020.2003.2024.004655*
36. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolsma P, Rietveld A, Brouwer M, Bouwmeester-Vincken N, Harders F, Hakze-van der Honing R, Wegdam-Blans MCA, Bouwstra RJ, GeurtsvanKessel C, van der Eijk AA, Velkers FC, Smit LAM, Stegeman A, van der Poel WHM, Koopmans MPG (2021) Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 371:172-177
37. Luu K, Greenhill CJ, Majoros A, Decker T, Jenkins BJ, Mansell A (2014) STAT1 plays a role in TLR signal transduction and inflammatory responses. *Immunology and cell biology* 92:761-769
38. Lei X, Dong X, Ma R, Wang W, Xiao X, Tian Z, Wang C, Wang Y, Li L, Ren L, Guo F, Zhao Z, Zhou Z, Xiang Z, Wang J (2020) Activation and evasion of type I interferon responses by SARS-CoV-2. *Nature communications* 11:3810
39. Narayanan K, Huang C, Makino S (2008) SARS coronavirus accessory proteins. *Virus research* 133:113-121
40. Lloyd AT, Sharp PM (1992) Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic acids research* 20:5289-5295
41. Li JY, Liao CH, Wang Q, Tan YJ, Luo R, Qiu Y, Ge XY (2020) The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus research* 286:198074
42. Zhang J, Cruz-cosme R, Zhuang M-W, Liu D, Liu Y, Teng S, Wang P-H, Tang Q (2020) A systemic and molecular study of subcellular localization of SARS-CoV-2 proteins. *Signal Transduction and Targeted Therapy* 5:269
43. Stukalov A, Girault V, Grass V, Bergant V, Karayel O, Urban C, Haas DA, Huang Y, Oubraham L, Wang A, Hamad SM, Piras A, Tanzer M, Hansen FM, Enghleitner T, Reinecke M, Lavacca TM, Ehmann R, Wölfel R, Jores J, Kuster B, Protzer U, Rad R, Ziebuhr J, Thiel V, Scaturro P, Mann M, Pichlmair A (2020) Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *bioRxiv:2020.2006.2017.156455*
44. Ren W, Zhu Y, Wang Y, Shi H, Yu Y, Hu G, Feng F, Zhao X, Lan J, Wu J, Kenney DJ, Douam F, Tong Y, Zhong J, Xie Y, Wang X, Yuan Z, Zhou D, Zhang R, Ding Q (2021) Comparative analysis reveals the species-specific genetic determinants of ACE2 required for SARS-CoV-2 entry. *PLoS pathogens* 17:e1009392
45. Gomez CE, Perdiguero B, Esteban M (2021) Emerging SARS-CoV-2 Variants and Impact in Global Vaccination Programs against SARS-CoV-2/COVID-19. *Vaccines* 9
46. Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, Stowe J, Tessier E, Groves N, Dabrera G, Myers R, Campbell CNJ, Amirthalingam G, Edmunds M, Zambon M, Brown KE, Hopkins S, Chand M, Ramsay M (2021) Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med*

Figures

Figure 1

Composition and substitution analysis of SARS-CoV-2 isolated from animals. (A) The reported animals infected with SARS-CoV-2 and the defined transmission route from human to animal. (B) Phylogenetic tree using the maximum likelihood method and Tamura-Nei model performed by MEGA-X. The tree was provided with 500 bootstraps. (C) The proportions of uracil, guanine, thymine, and cytidine substitutions (nonsynonymous) in mink SARS-CoV-2 and other animals were separately counted. (D) Base pair changes observed in the mink SARS-CoV-2 genomes. All transitions and transversions were recorded and analyzed (see Supplementary Table S2). (E) The synonymous and nonsynonymous substitutions of Mink-SARS2 were counted and analyzed. (F) The relative proportions of all transitions and transversions were separately analyzed.

Figure 2

The mutation spectra of the spike protein and the selection pressure. (A) The substitutions in the animal viral genome in this study were analyzed, including uracil, guanine, thymine, and cytidine as substituted with other bases. (B) The dN-dS value was calculated using the Datamonkey tool. (C) The genomic CAI value was calculated using SARS-CoV-2 sequences in humans and animals. Bat-CoV refers to RaTG13 and the corresponding host (bat), Pangolin-COV refers to pangolin coronavirus (GenBank: QLR06867.1), other animal-SARS2 means SARS-CoV-2 isolated from the indicted animals, the first SARS-CoV-2 refers the human host. (D) ENC-Plot analysis of each coronaviruses sequences from animals and human. (E) Neutral plot analysis of each coronaviruses sequences from animals and human. (F) The CAI value of spike sequences in SARS-CoV-2 from human and animals.

Figure 3

Receptors and binding analysis of spike with human and mink. (A) Receptors ACE2 interacted with the SARS-CoV-2 Spike. (B) Alignments of receptors ACE2 in humans and animals. The single letter amino acid (aa) that functions in the spatial interaction is indicated. (C) Alignment of the SARS-CoV-2 RBD sequences in humans and in animals reported to have been infected. The residues in contact with ACE2 are indicated. (D) Results of the comparison of the spike structure from Mink-SARS2 with the reference strain WIV04. Visualization of the changed residues within Mink-SARS2 are shown as colored balls. (E) The binding free energy of the wild-type RBD and three mutants with the human receptor and mink receptor, respectively. (F) Characterization of the binding between human ACE2 (hACE2, upper) and mink ACE2 (mACE2, lower) with spike RBD mutants by FACS. His-tagged wild type RBD, RBD mutants and NTD

proteins were incubated with CHO cells expressing with eGFP-fused ACE2s. Cells stained with the wild type, mutant RBD and NTD proteins are shown in bright green, red and orange, respectively. The spike NTD was set as the negative control.

Figure 4

Codon usage and fitness analysis of the viral encoded proteins. (A) Total mutations of the RBM variants were recorded and counted after analysis by MEGA-X software. The frequency was calculated using the Datamonkey server, and the figure was produced by WebLogo (<https://weblogo.berkeley.edu/logo.cgi>). (B) The synonymous codon usage bias of SARS-CoV-2 was produced by WebLogo, comparing the mink SARS-CoV-2 sequence (GenBank ID MT396266) with SARS-CoV strain Toronto-2. (C) All codon-specific estimates of selective coefficient index were calculated, the indicated mutant codons of N501 were marked in purple, and the highest fitness codons were colored in red. (D) Analysis of relative synonymous codon usage in the coding residues. The figure shows the RSCU values of every codon for each amino acid.

Supplementary Files

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

- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS3.docx](#)
- [SupplementaryTableS4.xlsx](#)