Studies on Human-Coronavirus protein-protein interaction network from the perspective of viral adaptation in a novel host

Running title: Evolution of coronaviruses in a novel host

Debarun Acharya* (debarunacharya@gmail.com)

Tapan K Dutta* (tapan@jcbose.ac.in)

Affiliation

Department of Biological Sciences,

Bose Institute

EN-80, Sector V, Bidhannagar, Kolkata 700091, West Bengal, India.

*Corresponding authors.

Abstract

Host-pathogen interaction is the best example of an evolutionary arms race where pathogen and host continuously coevolve to survive and exert negative effects on each other. The adaptability of both host and pathogen is critical for this association. In this study, we explored the adaptation of the severe acute respiratory syndrome (SARS) coronavirus (CoV) in humans from the genomic and evolutionary perspectives based on a comparative analysis of SARS-CoV2-human and Other-CoV-human interactions. We observed that human proteins that are part of the SARS-CoV2-human association are less enriched in hubs and bottlenecks. Again, they also take part in fewer protein complexes and show faster evolutionary rates compared to the Other-CoV-associated human proteins. The human proteins involved in the interaction with SARS-CoV2 are mostly longer proteins harboring long disordered stretches and a higher level of disordered protein binding sites. Codon usage analysis also revealed that the novel coronavirus is least adapted to codons, used in housekeeping genes and genes that get expressed in lung tissues, compared to other two deadly coronaviruses, SARS CoV1 and MERS CoV. We conclude that the signatures showed by SARS-CoV2-human protein interaction network represent a model for understanding the evolutionary feature of an early stage of host-virus association in comparison to that obtained from relatively long-term-associated host-virus interactions, achieving higher levels of adaptation.

Importance

The current study focuses on the evolution of viruses, using novel coronavirus as a model. The host-pathogen interaction is better viewed from molecular perspective, where host and pathogen proteins interact and co-evolve with each other. This is even more important for viruses, which use the host's cellular machinery for protein synthesis. As for a deadly virus, killing the host is inadvertent as it is also suicidal for the virus itself. A more stable association

that is beneficial for both host and virus will require host adaptation, achievable through evolutionary time. In this study, the deadly novel human coronavirus has been viewed as the early stage of host viral interaction, which has been achieved through a more random way, with lower adaptation to host codons compared to the other human coronaviruses, having more stable host-association for being evolutionarily older.

Keywords

Host-pathogen interaction; SARS-CoV2; viral evolution; codon usage; protein intrinsic disorder.

1. Introduction

The recent COVID-19 pandemic caused by the novel coronavirus SARS-CoV2 has been a serious threat to global public health, with its effects persisting throughout most countries and continents. Since the last two and a half years, a plethora of research articles were published to explore the origin and the nature of the virus (40, 50, 70, 77), the molecular basis underlying the infection (59, 64, 74), the epidemiology of the disease (6, 24, 32), the impact on human population (2, 3, 42), the protection strategies from the virus (1, 66, 95) among the many aspects of SARS-CoV2 infection. The high infectivity and mortality rates among the SARS-CoV2 patients had influenced these studies and accelerated the development of therapeutic and prophylactic measurements against SARS-CoV2 throughout the globe (7, 38). This pandemic situation further initiated the discovery of efficient treatments like convalescent plasma therapy (17, 33), the use of monoclonal antibodies (56, 58, 92), the development of vaccines (45, 62, 83) and the repurposing of drugs (20, 41, 51, 82, 91, 98). A huge number of repurposed drugs (293 drugs as of 10th June 2022) (60) and more than 30 vaccines (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines) were approved globally. However, the challenges had been very tough due to the accumulation of mutations leading to the generation of different SARS-CoV2 variants including the emergence of alpha and beta variants in December 2020 to the omicron in November 2021 (53, 78).

The horseshoe bats (*Rhinolophus* spp.) are the reservoir hosts of coronaviruses, including the SARS CoV1 and SARS-CoV2 (72, 90, 97), with a few reports indicating pangolins as the intermediate host (47, 55, 94). The high genomic sequence similarity (~99.99%) of SARS-CoV2 isolates from different patients suggests a very recent host-shift to humans. Thus, from the perspective of molecular evolution, it is showing a real-time view of viral evolution to adapt to a novel host from a naïve population without any prior exposure to the virus.

The protein-protein interactions are crucial components in the molecular crosstalk between host and pathogen during infection. Such interactions are necessary for the establishment of the pathogen and its multiplication inside the host. For a novel pathogen, interactions are often random and may lead to severe consequences in the host. However, from the perspective of pathogens, killing the host is suicidal and extreme virulence is, therefore, not a sustainable evolutionary solution. Instead, the pathogen undergoes evolutionary selection within the host's body, gradually adapting to the host environment over time. This process shapes the coevolutionary dynamics of host-pathogen interactions. Thus, the increasing multiplication and exposure to the human subpopulations provides the novel virus a large evolutionary space to form mutant variants, evolutionarily equipped to survive and reproduce (53, 78).

The novel coronavirus SARS-CoV2 is the newest member of the coronavirus family, belonging to the Sarbecovirus subgenus (genus *Betacoronavirus*) (49). It is a close relative to the animal coronaviruses discovered early in the twentieth century and the third coronavirus known to be pathogenic for humans (79). Although, until recent times the effect of human-infecting coronaviruses (namely Human coronaviruses HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1) was only limited to mild symptoms comparable to the common cold (54). On the other hand, the outbreak of severe acute respiratory syndrome (SARS) coronavirus in 2003 (SARS CoV) caused the death of more than 700 people globally. The Middle East Respiratory Syndrome Coronavirus (MERS-CoV) was initially identified in Saudi Arabia in 2012. This virus follows a zoonotic mode of transmission, being spread to humans from dromedary camels that were their major reservoir host. The disease has a high death rate of about 35% of the infected persons, affecting about 1400 individuals in South Korea and Middle East countries, like Saudi Arabia, United Arab Emirates, Jordan, Qatar, Oman, and Iran. The severity of infections caused by these viruses is much lower compared to SARS-CoV2, where the host-virus interaction is expressing a naïve state compared to the evolutionarily older coronavirus strains. The recent pandemic caused by SARS-CoV2 lead to nearly 7 million deaths and over 700 million infected

persons globally as of July 2023.TheSARS-CoV2 has 79.7% sequence similarity with the SARS-CoV of 2003 (98), the highest compared to other pathogenic coronaviruses. The devastations were as extreme as more than 100,000 global deaths per day. Apart from the huge number of casualties, the socioeconomic distress caused by the pandemic is likely to last longer than anticipated.

Additionally, it does not change the fact that the interactions with human proteins for both the SARS-CoV2 and other coronaviruses may control the severity of these infections, influencing the biological function of the human proteins. However, with time, human-coronavirus interaction is supposed to undergo coevolution leading to their coexistence. Such a scenario may be feasible by purifying and stabilizing selections and/or changing the pattern of the host-pathogen interaction from “by chance” to “by choice”. Moreover, other preventive measures of the disease such as passive immunization or vaccination will certainly help to stop the disease transmission or may imply further selection pressure to generate vaccine-resistant variants (57). Irrespective of the outcome, it is obvious that the first report of human-SARS-CoV2 interaction data published in 2020 represents an early stage of host-pathogen interaction compared to the available interactome data of human and other coronavirus proteins accumulated over time.

Studies with pathogen-human protein-protein interaction networks have suggested a high degree centrality of human proteins interacting with pathogen proteins, which is consistent for viral, bacterial, protozoan and fungal pathogens (4, 22, 30, 63). The novel coronavirus SARS-CoV2, however, is only recently infected its human host and hence, has earned very little time to evolve with the host, which should be reflected to the pathogen-human protein-protein interaction (PHPPI) network. The human-SARS-CoV2 PPI network, therefore, represents an early stage of viral adaptation to the host environment. On the contrary, the other deadly coronaviruses with relatively long-term association with human hosts have undergone coevolutionary processes, stabilizing the host-virus interaction. In this study, we explored the human coronavirus interactome data for SARS-CoV2 and other coronaviruses and compared

the attributes of the human proteins involved in these two groups. The study aims to understand how the interactions between human and SARS-CoV proteins become feasible by accounting the roles of human proteins and their associated protein-interaction network features that facilitate their interaction with the nCoV proteins, despite humans being a novel host for the viruses.

2. Materials and Methods

2.1. Sequence data

The coding sequences (CDS) for MERS and SARS coronaviruses were obtained using the NCBI GenBank (10) (NCBI Reference Sequences NC_019843.3, NC_004718.3 and NC_045512 for MERS-CoV, SARS-CoV1 and SARS-CoV2, respectively). The Human CDS sequences were obtained using the Biomart interface of Ensembl (93). Human protein sequences were obtained from the UniProt (85) [proteome ID: UP000005640]. We have selected the 'reviewed' proteome only, which leads to a list of 20360 proteins and their corresponding amino acid sequences.

2.2. Protein-protein interaction data

The human-SARS-CoV2 protein-protein interaction data was obtained from recent large-scale studies (8, 9, 12-14, 18, 23, 35-37, 43, 44, 46, 48, 50, 73, 76, 80, 81) dealing with at least 10 interactions. After removing the redundant interactions, we obtained a total of 12271 human-SARS-CoV2 interacting pairs involving 4669 human proteins. The interaction of human proteins with other coronavirus proteins was obtained from the H2V database (96) (http://www.datjar.com:40090/h2v/downloaddata/) that includes MERS-CoV and SARS-CoV1 and their interacting protein partners in human. A total of 296 and 1162 human-virus interactions were obtained for MERS-CoV and SARS-CoV1 respectively, involving a total of 1306 human

proteins interacting with coronaviruses other than SARS-CoV2 (mentioned as Other-CoVs throughout the manuscript). Finally, we classified human proteins based on their interaction statuses with coronavirus proteins to obtain 3704 SARS-CoV2 interacting proteins, 341 Other-CoV interacting proteins, 965 proteins interacting with both these groups (SARS-CoV2 and Other-CoV) and 15371 proteins (noninteracting) without any reported interaction with any coronaviruses.

The human protein-protein interaction network was obtained from the supplementary data of Cheng et al. (19), which provides a comprehensive list of human protein-protein interaction networks collected from established PPI databases, namely IntAct (65), InnateDB (11), PINA (21), HPRD (68), BioGRID (16), MINT (52) etc. The Entrez gene IDs were converted to corresponding UniProt IDs using UniProt's ID conversion tool. The final data contains 214412 interactions involving 15580 human proteins. The human PPI network was analyzed using the NetworkAnalyzer plugin of Cytoscape 3.8.1 (75) to determine the degree and betweenness centrality of human proteins. The top 20% proteins having the higher degree and betweenness centrality were considered as hubs and bottlenecks, respectively (4).

2.3. Protein intrinsic disorder and disordered protein binding sites

The intrinsic disorder of human proteins was predicted using the IUPred 2A (26), where each amino acid residue in a protein is provided with a disorder score ranging from 0 to 1. In IUPred, the residues with a score of ≥ 0.5 are considered disordered residues, and the consecutive stretches of ≥ 30 amino acid-long disordered residues are considered as long-disordered regions. An in-house PERL script was used to determine long disordered regions of human proteins (4). The disordered binding regions were calculated using the ANCHOR2 webserver (27). In addition, the Molecular Recognition Features (MoRFs) of human disordered proteins interacting with coronavirus proteins were obtained from thefMoRFpred webserver (25), where

each amino acid in a disordered protein is classified as MoRF or non-MoRF residue. Consecutive stretches of ≥ 5MoRF residues were considered as the regions responsible for disordered protein binding (MoRF regions) and the number of such regions in each human protein were calculated using the in-house PERL script.

2.4. Evolutionary Rate

The evolutionary rates of human proteins were obtained from the Ensembl Biomart (93) as the ratio of the nonsynonymous nucleotide substitutions per nonsynonymous sites and synonymous nucleotide substitutions per synonymous sites (i.e., dN/dS ratio) for each human protein, with their 1:1 orthologous mouse proteins. The mutation saturation was controlled by removing all proteins having dS values >3 (5).

2.5. Functional enrichment

Functional enrichment analysis was accomplished using the Gene Ontology (GO) terms (34) for all the GO domains using the GOrilla webserver (31). We compared human proteins interacting with SARS-CoV2 with those interacting with Other-CoVs using the former as target and the latter as background in the GOrilla to obtain enriched functions along with the enrichment scores.

2.6. Housekeeping and lung-specific genes

Human housekeeping genes were obtained from the HRTAtlas 1.0 database (39), which presents the human housekeeping genes that are stably expressed in 52 tissues and cell types. We obtained a final list of 2071 non-redundant protein-coding housekeeping genes. The genes expressed in the lungs were obtained from the Human Protein Atlas using the Tissue Atlas for lung transcriptome (84). We have selected a final list of 190 protein-coding genes that are selectively elevated in the lungs.

2.7. Codon usage

The frequency of each codon were calculated using the 'cusp' function of the EMBOSS, an open-source software package (71). The function cusp calculates the codon usage for one or more nucleotide coding sequences and provides a tabulated output, using desired CDS sequences. The codon composition of human coronavirus CDS and that of the human housekeeping and lung-specific genes were calculated using the 'cusp' function. Similarly, the CDS sequences of viral proteins for the three coronaviruses were used to determine their codon usages. The expected number of each codon, given the input sequence(s), per 1000 bases were correlated for each coronavirus CDSs with that of human housekeeping (HK) and lung-specific (Lung) genes, thereby obtaining the Pearson correlation coefficient (PCC) for each combination (HK-MERS, HK-SARS1, HK-SARS2 and Lung-MERS, Lung-SARS1, Lung-SARS2).

3. Results

3.1. Network centrality of human proteins interacting with coronaviruses

Human proteins were differentiated based on their interaction statuses with the proteins of human coronaviruses using the H2V database (96) and the available high-throughput protein interaction datasets (see Methods). The analyses resulted in 3704 SARS-CoV2 interacting proteins, 341 other CoV interacting proteins, 965 proteins interacting with both these groups (SARS-CoV2 and Other-CoV) and 15371 non-interacting proteins which imply no reported interaction with none of the coronaviruses. We identified the central proteins in the human protein-protein interaction (PPI) network using the human high-throughput PPI network data reported by Cheng et al. (19). We obtained the hubs and the bottlenecks in human PPI network using the degree and betweenness centrality values, respectively (see methods). These central proteins are thought to be the prime targets of pathogens and are involved in pathogen-host

protein-protein interaction network (4).Our comparative analysis with high-throughput protein interaction data suggested that the noninteracting group contains the lowest proportion of hubs and bottlenecks (Figure 1). However, human proteins interacting with SARS-CoV2 contain a lower proportion of hubs and bottlenecks than those interacting with other coronaviruses (Figure1). The result indicates two observations: firstly, the SARS coronaviruses mainly target host proteins with higher connectivity, as revealed in earlier studies with other pathogens (4, 29). Secondly, the SARS-CoV2 interacts with less-central human proteins than other-CoVs. This is expected as the novel SARS-CoV2 has no prior exposure to host cellular conditions, whereas the virus-human PPI network in evolutionarily 'older' coronaviruses (Other-CoV) represents a relatively more stable coevolutionary association.

3.2. Intrinsic disorder of human proteins interacting with coronaviruses

Protein intrinsic disorder plays an important role in the host-pathogen protein-protein interaction network (69) as it provides flexibility to protein structure to promote low-affinity interactions between proteins (4, 67). The increased flexibility of the host's disordered proteins allows host-pathogen protein interactions, which otherwise remain unfeasible (4). For a detailed understanding of this scenario from the coronavirus perspective, the human proteins interacting with SARS-CoV2 proteins and those interacting with other coronaviruses were examined, using the IUPred protein intrinsic disorder prediction algorithm. Interestingly, we observed that the human proteins interacting with SARS-CoV2 have significantly more long disordered stretches and more disordered residues than those interacting with other coronaviruses, followed by proteins interacting with both CoVs and non-interacting proteins (Table 1). However, when we compared the length of human proteins, we observed that the SARS-CoV2-interacting human proteins are significantly longer (Table 1, $P = 7.89 \times 10^{-152}$, Mann-Whitney test between two groups) than those interacting with Other-CoVs.

3.3. Molecular Recognition Features

Intrinsically disordered regions within a protein often contain specialized regions that undergo disorder-to-order transition during the physical interaction with their binding partners. These regions, known as Molecular Recognition Features (MoRFs), contain the amino acid residues which are important for protein-protein interactions (25, 86). The amino acid residues in an intrinsically disordered protein can be classified into MoRF and non-MoRF residues, with the former playing vital roles in protein-protein interaction (67), including the pathogen-host protein-protein interactions during infections (4). The abundance of MoRF residues in human-disordered proteins interacting with coronaviruses was calculated using the fMoRFpred webserver (25). Interestingly, it was observed that the SARS-CoV2-interacting human disordered proteins contain a higher number of MoRF residues which may contribute to their interaction with the viral proteins (Table 1). Also, it suggests that the novel host-virus interaction utilizes the disorder-binding residues of host proteins to facilitate their association.

3.4. Protein complex association

The protein complex association measures the association of a protein to macromolecular complexes involving other proteins to perform the destined function of the complexes. The involvement of a protein in different complexes, often known as protein complex number (PCN) (15), can be a suitable measurement of the number of functions (multifunctionality) in which the protein is involved. Here, the protein complex association of human proteins interacting with SARS-CoV2 and other coronaviruses revealed that the latter group is associated with more protein complexes followed by SARS-CoV2 ($PCN_{OtherCoV}$ = 3.729; $PCN_{SARS\text{-}CoV2}$ = 3.469; $PCN_{Noninteracting}$ = 3.028; PCN_{Both} = 2.976; $P = 2.57 \times 10^{-4}$, Kruskal-Wallis test). Nonetheless, this is fairly expected for the novel virus and on the other hand, relatively more evolved other coronaviruses are capable to hijack more protein complexes in the human PPI network.

3.5. Evolutionary rate

The evolutionary rate of a protein depicts the change in its amino acid sequence over the evolutionary timescale. Highly expressed proteins evolve slower, due to the functional constraints imposed on them, particularly, because of their high cellular demand (28). Such proteins are more often targeted by pathogenic proteins, as their slower rate of evolution is beneficial to the pathogens for sustained interaction over evolutionary timescale. For a novel pathogen, the interactions with host proteins may represent an early stage of this coevolution, where the target proteins are mostly random, and the evolutionary rate of such targets may be higher than that emerged from coevolved interactions. To further explore this, we have studied the evolutionary rate of human proteins interacting with different coronaviruses, using 1:1 mouse orthologs (see materials and methods). We observed that the novel coronavirus interacts with slightly faster evolving human proteins than other coronaviruses, although the difference of interaction between these two groups is small ($P_{dN} = 4.76 \times 10^{-2}$, $P_{dN/dS} = 1.09 \times 10^{-1}$, Mann Whitney test). Nevertheless, the noninteracting human proteins evolve at quite a fast rate (Figure 2, $P_{dN} = 2.68 \times 10^{-102}$, $P_{dN/dS} = 5.97 \times 10^{-73}$, Kruskal-Wallis test). This trend supports our hypothesis that the novel coronavirus, being only recently adapted to the human host is yet to undergo extended co-evolutionary processes and represent the initial stage of host-pathogen coevolution.

3.6. Functional enrichment analysis

The functional enrichment of SARS-CoV2 and Other CoV-interacting human proteins were compared using the GOrilla webserver using default parameters for all the Gene Ontology (GO) terms. The GO biological process and molecular function domains did not show any difference between the groups. The difference was observed only for GO cellular component, where the SARS-CoV2 interacting human genes were localized in cell membrane, as well as in the

endoplasmic reticulum and Golgi apparatus (Figure 3), indicating that these human proteins may be involved in cellular transport system, and are hijacked by the SARS-CoV2 upon infection.

3.7. Codon usage

The function of genes relies on their encoded proteins, which form the desired three-dimensional structure attributed due to the interactions of its constituent amino acids. However, for functional attributes, the desired expression of these proteins, often represented as the 'translational efficiency', is crucial. After the mRNA recognition by the ribosomes, an efficient translation rate depends on the abundance of aminoacylated tRNAs, which interact with codons on mRNAs via their anticodons. Therefore, all possible codon combinations that encode the same amino acid chain do not necessarily have the same translational efficiency, though they share identical sequences (87). Viral pathogens that depend on host cellular machinery including the host aminoacylated tRNAs for the translation of their proteins should prefer a higher translation rate to facilitate their multiplication and propagation in new host cells/tissues (61), and this can be facilitated by using similar codon usage with host proteins. We used the EMBOSS (71) tool to determine the frequency of each codon from MERS-CoV, SARS-CoV1 and SARS-CoV2. The codon usage pattern was correlated with human housekeeping (N = 2070) and lung-specific (N = 186) genes using the Pearson correlation coefficient (PCC) to understand the codon adaptability of these coronavirus genes. The SARS-CoV2 showed the lowest codon adaptation ($PCC_{housekeeping}$ = 0.259; PCC_{lung} = 0.052) in contrast to the MERS-CoV ($PCC_{housekeeping}$ = 0.349; PCC_{lung} = 0.172) and the SARS-CoV1 ($PCC_{housekeeping}$ = 0.401; PCC_{lung} = 0.213) with respect to both the housekeeping and the lung-specific genes (Figure 4). The correlation values suggest that both the SARS-CoV1 and MERS-CoV are better codon-adapted with housekeeping as well as lung-specific gene, compared to theSARS-CoV2, which has just started adapting to its novel host.

4. Discussion

The host-pathogen coevolution has been a popular area of research for ages, where the hosts are exploited by the pathogens for their survival and reproduction, causing several ailments including life-threatening diseases. Although, host-pathogen interaction is better explainable under their coevolutionary dynamics, it is difficult to portray due to the complex nature of the interactions at the cellular, molecular, and biochemical levels, and/or the availability of only a snapshot of these interactions at any point of time. Of particular importance, viruses, utilize their host cellular machinery to synthesize the proteins required for survival, reproduction, immune evasion and disease progression (89). For a deadly virus, like SARS-CoV2, it is not easy to establish a sustainable coevolutionary relationship with the novel host, as its deadly effects often kill the host, hampering the propagation of the virus. However, with time, molecular interactions between host and virus usually get stabilized, forming a stable PHPPI network, resulting in coexistence. In this study, we compared the virus-host interactome of the novel coronavirus SARS-CoV2 and the other major human coronaviruses, namely SARS-CoV1 and MERS-CoV to understand the adaptation and establishment of SARS-CoV2 in humans. Most viruses target host proteins with a high degree and betweenness centrality. These proteins play important roles in the host cellular processes and assist pathogens to gain control over the host PPI network. However, theSARS-CoV2-interacting human proteins are less central than those interacting with other coronaviruses, which are presumably more adapted to their host, having relatively long-term coevolutionary relationships. This suggests that the novel host-pathogen protein interactions may have been initiated by chance. However, the random nonspecific protein interactions are often attributed by the intrinsically disordered stretches present in its structure (4). The SARS-CoV2-human protein interaction network also revealed more interaction with the intrinsically disordered proteins with a high number of disordered protein binding sites promoting low-affinity interactions. The human proteins attacked by SARS-CoV2

are mostly localized in cell membrane, Golgi apparatus and endoplasmic reticulum, indicating that the virus utilizes the cellular transport system (Figure 3). Moreover, we also observed that the protein length is also a major determinant of host-virus interaction, as longer proteins have a higher probability of harboring intrinsically disordered residues and disordered protein binding motifs. So, their abundance, but not density, is the key determinant, and this promotes the host-pathogen interactions 'by chance'. Furthermore, to survive and multiply in the host's body and utilize the host cellular machinery, the virus needs to adapt to host codon usage preferences. The host tRNA pool and codon usage remain optimized for the expression of the host proteins (88). For a novel host-virus association, this is merely a chance factor that the host tRNA pool will support the codons used in the viral proteome. However, with time, viruses may adapt to host codon usage preferences for maintaining optimum efficiency in their multiplication. We used the codon usage pattern of the human housekeeping genes and the genes specifically enriched in the lung tissues and correlated with the individual coronavirus CDSs. The housekeeping genes need a steady expression in all tissues and the required cognate tRNAs should be maintained at a steady level for their expression. Whereas, genes expressed in the lung, one of the sensitive sites for colonization and infection of the coronaviruses, should maintain a steady cognate tRNA pool. A lower correlation (PCC) in the codon usage pattern ofSARS-CoV2 with both the human housekeeping and lung-specific proteome was observed than that of MERS-CoV and SARS-CoV1 suggesting a better adaptation in the evolutionarily 'older' viruses to host's codons and tRNA pool in contrast to the novel coronavirus.

5. Conclusion

This study is aimed at the understanding of viral evolution in a novel host from the perspective of pathogen-host PPI network, using human coronaviruses. The novel coronavirus SARS-CoV2 was found to bear early signatures of pathogen-host PPI network, where the pathogen utilizes host proteins with higher connectivity, in particular the longer ones bearing intrinsically

disordered stretches, compared to the other human coronaviruses. The SARS-CoV2 shows lowest codon adaptation to human, further suggesting that this virus is in its early stage of host-association. Together, these results indicate that SARS-CoV2, a novel and deadly human coronavirus, has recently been adapted to its human host, utilizing the intrinsic disordered residues. The virus has not yet fully adapted to the human cellular and molecular network and that of codon usage pattern, but preferentially exploiting the central proteins, particularly bottlenecks for its deadly pathogenicity and disease progression.

CRediT authorship statement

Debarun Acharya (Conceptualization, Investigation, Formal analysis, Writing- original draft), and Tapan K. Dutta (Conceptualization, Supervision, Writing- review & editing)

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest statement

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Funding

No funding was received by the authors.

Acknowledgement

The authors thank Bose Institute, Kolkata.

Abbreviations

CDS: coding sequence

CoV: coronavirus

GO: gene ontology

HK: housekeeping

IBV: infectious bronchitis virus

MERS: Middle East respiratory syndrome

MHV: mouse hepatitis virus

MoRF: molecular recognition feature

PCC: Pearson correlation coefficient

PHPPI: pathogen-host protein-protein interaction

PPI: protein-protein interaction

SARS: severe acute respiratory syndrome

tRNA: transfer-ribonucleic acid

References:

1. **Abdoul-Azize, H. T., and R. El Gamil.** 2021. Social protection as a key tool in crisis management: learnt lessons from the COVID-19 pandemic. Glob Soc Welf **8:**107-116.
2. **Aburto, J. M., R. Kashyap, J. Schöley, C. Angus, J. Ermisch, M. C. Mills, and J. B. Dowd.** 2021. Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. J Epidemiol Community Health **75:**735-740.
3. **Aburto, J. M., J. Schöley, I. Kashnitsky, L. Zhang, C. Rahal, T. I. Missov, M. C. Mills, J. B. Dowd, and R. Kashyap.** 2022. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. Int J Epidemiol **51:**63-74.
4. **Acharya, D., and T. K. Dutta.** 2021. Elucidating the network features and evolutionary attributes of intra- and interspecific protein–protein interactions between human and pathogenic bacteria. Sci Rep **11:**190.
5. **Acharya, D., and T. C. Ghosh.** 2016. Global analysis of human duplicated genes reveals the relative importance of whole-genome duplicates originated in the early vertebrate evolution. BMC Genomics **17:**1-14.
6. **Alteri, C., V. Cento, A. Piralla, V. Costabile, M. Tallarita, L. Colagrossi, S. Renica, F. Giardina, F. Novazzi, and S. Gaiarsa.** 2021. Genomic epidemiology of SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy. Nat Commun **12:**1-13.
7. **Andreano, E., and R. Rappuoli.** 2021. SARS-CoV-2 escaped natural immunity, raising questions about vaccines and therapies. Nat Med **27:**759-761.
8. **Andres, A. D., Y. Feng, A. R. Campos, J. Yin, C.-C. Yang, B. James, R. Murad, H. Kim, A. J. Deshpande, and D. E. Gordon.** 2020. SARS-CoV-2 ORF9c is a membrane-associated protein that suppresses antiviral responses in cells. bioRxiv.
9. **Armstrong, L. A., S. M. Lange, V. de Cesare, S. P. Matthews, R. S. Nirujogi, I. Cole, A. Hope, F. Cunningham, R. Toth, and R. Mukherjee.** 2020. Characterization of protease activity of Nsp3 from SARS-CoV-2 and its in vitro inhibition by nanobodies. BioRxiv.
10. **Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers.** 2012. GenBank. Nucleic Acids Res **41:**D36-D42.
11. **Breuer, K., A. K. Foroushani, M. R. Laird, C. Chen, A. Sribnaia, R. Lo, G. L. Winsor, R. E. W. Hancock, F. S. L. Brinkman, and D. J. Lynn.** 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res **41:**D1228-D1233.
12. **Brockbank, S. M. V., J. Soden, R. Faba-Rodriguez, L. R. Ribeiro, C. Geh, H. Thomas, J. Delight, L. Coverley, W. M. Abbott, and J. Freeth.** 2021. SARS-CoV-2 comprehensive receptor profiling: mechanistic insight to drive new therapeutic strategies. BioRxiv.
13. **Bujanic, L., O. Shevchuk, N. von Kügelgen, A. Kalinina, K. Ludwik, D. Koppstein, N. Zerna, A. Sickmann, and M. Chekulaeva.** 2022. The key features of SARS-CoV-2 leader and NSP1 required for viral escape of NSP1-mediated repression. RNA **28:**766-779.
14. **Cai, T., Z. Yu, Z. Wang, C. Liang, and S. Richard.** 2021. Arginine Methylation Regulates SARS-CoV-2 Nucleocapsid Protein Function and Viral Replication. bioRxiv.
15. **Chakraborty, S., B. Kahali, and T. C. Ghosh.** 2010. Protein complex forming ability is favored over the features of interacting partners in determining the evolutionary rates of proteins in the yeast protein-protein interaction networks. BMC Syst Biol **4:**1-9.

16. **Chatr-Aryamontri, A., B.-J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, C. Stark, A. Breitkreutz, N. Kolas, and L. O'Donnell.** 2015. The BioGRID interaction database: 2015 update. Nucleic Acids Res **43:**D470-D478.
17. **Chen, L., J. Xiong, L. Bao, and Y. Shi.** 2020. Convalescent plasma as a potential therapy for COVID-19. Lancet Infect Dis **20:**398-400.
18. **Chen, Z., C. Wang, X. Feng, L. Nie, M. Tang, H. Zhang, Y. Xiong, S. K. Swisher, M. Srivastava, and J. Chen.** 2021. Interactomes of SARS-CoV-2 and human coronaviruses reveal host factors potentially affecting pathogenesis. EMBO J **40:**e107776.
19. **Cheng, F., R. J. Desai, D. E. Handy, R. Wang, S. Schneeweiss, A.-L. Barabási, and J. Loscalzo.** 2018. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun **9:**2691.
20. **Consortium, W. H. O. S. T.** 2021. Repurposed antiviral drugs for Covid-19—interim WHO solidarity trial results. N Engl J Med **384:**497-511.
21. **Cowley, M. J., M. Pinese, K. S. Kassahn, N. Waddell, J. V. Pearson, S. M. Grimmond, A. V. Biankin, S. Hautaniemi, and J. Wu.** 2012. PINA v2. 0: mining interactome modules. Nucleic Acids Res **40:**D862-D865.
22. **Crua Asensio, N., E. Munoz Giner, N. S. De Groot, and M. Torrent Burgas.** 2017. Centrality in the host–pathogen interactome is associated with pathogen fitness during infection. Nat Commun **8:**1-6.
23. **Davies, J. P., K. M. Almasy, E. F. McDonald, and L. Plate.** 2020. Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus nonstructural proteins identifies unique and shared host-cell dependencies. ACS Infect Dis **6:**3174-3189.
24. **Dhar, M. S., R. Marwal, R. Vs, K. Ponnusamy, B. Jolly, R. C. Bhoyar, V. Sardana, S. Naushin, M. Rophina, and T. A. Mellan.** 2021. Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. Science **374:**995-999.
25. **Disfani, F. M., W.-L. Hsu, M. J. Mizianty, C. J. Oldfield, B. Xue, A. K. Dunker, V. N. Uversky, and L. Kurgan.** 2012. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics **28:**i75-i83.
26. **Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon.** 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics **21:**3433-3434.
27. **Dosztányi, Z., B. Mészáros, and I. Simon.** 2009. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics **25:**2745-2746.
28. **Drummond, D. A., J. D. Bloom, C. Adami, C. O. Wilke, and F. H. Arnold.** 2005. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci USA **102:**14338-14343.
29. **Durmuş Tekir, S. D., and K. Ö. Ülgen.** 2013. Systems biology of pathogen-host interaction: networks of protein-protein interaction within pathogens and pathogen-human interactions in the post-genomic era. Biotechnology J **8:**85-96.
30. **Dyer, M. D., T. M. Murali, and B. W. Sobral.** 2008. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog **4:**e32.
31. **Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini.** 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics **10:**1-7.
32. **Esper, F. P., Y.-W. Cheng, T. M. Adhikari, Z. J. Tu, D. Li, E. A. Li, D. H. Farkas, G. W. Procop, J. S. Ko, and T. A. Chan.** 2021. Genomic epidemiology of SARS-CoV-2 infection during the initial pandemic wave and association with disease severity. JAMA network open **4:**e217746-e217746.

33. **Focosi, D., A. O. Anderson, J. W. Tang, and M. Tuccori.** 2020. Convalescent plasma therapy for COVID-19: state of the art. Clin Microbiol Rev **33:**e00072-00020.
34. **Gene Ontology, C.** 2019. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res **47:**D330-D338.
35. **Gordon, D. E., G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, and D. L. Swaney.** 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature**:**1-13.
36. **Gu, Y., J. Cao, X. Zhang, H. Gao, Y. Wang, J. Wang, J. Zhang, G. Shen, X. Jiang, and J. Yang.** 2020. Interaction network of SARS-CoV-2 with host receptome through spike protein. BioRxiv.
37. **Gupta, M., C. M. Azumaya, M. Moritz, S. Pourmal, A. Diallo, G. E. Merz, G. Jang, M. Bouhaddou, A. Fossati, and A. F. Brilot.** 2021. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. Res Sq.
38. **Hotez, P. J., C. Batista, Y. B. Amor, O. Ergonul, J. P. Figueroa, S. Gilbert, M. Gursel, M. Hassanain, G. Kang, and D. C. Kaslow.** 2021. Global public health security and justice for vaccines and therapeutics in the COVID-19 pandemic. EClinicalMedicine **39:**101053.
39. **Hounkpe, B. W., F. Chenou, F. de Lima, and E. V. De Paula.** 2021. HRT Atlas v1. 0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Res **49:**D947-D955.
40. **Hu, B., H. Guo, P. Zhou, and Z.-L. Shi.** 2021. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol **19:**141-154.
41. **Jang, W. D., S. Jeon, S. Kim, and S. Y. Lee.** 2021. Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay. Proc Natl Acad Sci USA **118**.
42. **Jin, J., N. Agarwala, P. Kundu, B. Harvey, Y. Zhang, E. Wallace, and N. Chatterjee.** 2021. Individual and community-level risk for COVID-19 mortality in the United States. Nat Med **27:**264-269.
43. **Kim, D.-K., B. Weller, C.-W. Lin, D. Sheykhkarimli, J. J. Knapp, N. Kishore, M. Sauer, A. Rayhan, V. Young, and N. Marin-de la Rosa.** 2021. A map of binary SARS-CoV-2 protein interactions implicates host immune regulation and ubiquitination. bioRxiv.
44. **Kotani, N., and T. Nakano.** 2020. Candidate screening of host cell membrane proteins involved in SARS-CoV-2 entry. bioRxiv.
45. **Krammer, F.** 2020. SARS-CoV-2 vaccines in development. Nature **586:**516-527.
46. **Kruse, T., C. Benz, D. H. Garvanska, R. Lindqvist, F. Mihalic, F. Coscia, R. Inturi, A. Sayadi, L. Simonetti, and E. Nilsson.** 2021. Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. Nat Commun **12:**1-13.
47. **Lam, T. T.-Y., N. Jia, Y.-W. Zhang, M. H.-H. Shum, J.-F. Jiang, H.-C. Zhu, Y.-G. Tong, Y.-X. Shi, X.-B. Ni, and Y.-S. Liao.** 2020. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature **583:**282-285.
48. **Laurent, E. M. N., Y. Sofianatos, A. Komarova, J.-P. Gimeno, P. S. Tehrani, D.-K. Kim, H. Abdouni, M. Duhamel, P. Cassonnet, and J. J. Knapp.** 2020. Global BioID-based SARS-CoV-2 proteins proximal interactome unveils novel ties between viral polypeptides and host factors involved in multiple COVID19-associated mechanisms. BioRxiv.
49. **Lefkowitz, E. J., D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell, and D. B. Smith.** 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res **46:**D708-D717.
50. **Li, C., Y. Yang, and L. Ren.** 2020. Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. Infect Genet Evol **82:**104285.

51. **Li, G., and E. De Clercq.** 2020. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov **19:**149-150.
52. **Licata, L., L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, and E. Santonico.** 2012. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res **40:**D857-D861.
53. **Lippi, G., C. Mattiuzzi, and B. M. Henry.** 2022. Updated picture of SARS-CoV-2 variants and mutations. Diagnosis **9:**11-17.
54. **Liu, D. X., J. Q. Liang, and T. S. Fung.** 2021. Human coronavirus-229e,-oc43,-nl63, and-hku1 (coronaviridae). Encyclopedia of Virology**:**428.
55. **Liu, P., J.-Z. Jiang, X.-F. Wan, Y. Hua, L. Li, J. Zhou, X. Wang, F. Hou, J. Chen, and J. Zou.** 2020. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog **16:**e1008421.
56. **Lloyd, E. C., T. N. Gandhi, and L. A. Petty.** 2021. Monoclonal antibodies for COVID-19. Jama **325:**1015-1015.
57. **Luo, R. B., A. Delaunay-Moisan, K. Timmis, and A. Danchin.** 2021. SARS-CoV-2 biology and variants: anticipation of viral evolution and what needs to be done. Environmental Microbiology **23:**2339-2363.
58. **Marovich, M., J. R. Mascola, and M. S. Cohen.** 2020. Monoclonal antibodies for prevention and treatment of COVID-19. Jama **324:**131-132.
59. **McCallum, M., A. C. Walls, K. R. Sprouse, J. E. Bowen, L. E. Rosen, H. V. Dang, A. De Marco, N. Franko, S. W. Tilles, and J. Logue.** 2021. Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. Science **374:**1621-1626.
60. **Menestrina, L., C. Cabrelle, and M. Recanatini.** 2021. COVIDrugNet: a network-based web tool to investigate the drugs currently in clinical trial to contrast COVID-19. Sci Rep **11:**1-15.
61. **Nambou, K., M. Anakpa, and Y. S. Tong.** 2022. Human genes with codon usage bias similar to that of the nonstructural protein 1 gene of influenza A viruses are conjointly involved in the infectious pathogenesis of influenza A viruses. Genetica **150:**97-115.
62. **Ndwandwe, D., and C. S. Wiysonge.** 2021. COVID-19 vaccines. Curr Opin Immunol **71:**111-116.
63. **Nithya, C., M. Kiran, and H. A. Nagarajaram.** 2021. Comparative analysis of Pure Hubs and Pure Bottlenecks in Human Protein-protein Interaction Networks. bioRxiv.
64. **Niu, S., J. Wang, B. Bai, L. Wu, A. Zheng, Q. Chen, P. Du, P. Han, Y. Zhang, and Y. Jia.** 2021. Molecular basis of cross-species ACE2 interactions with SARS-CoV-2-like viruses of pangolin origin. EMBO J **40:**e107786.
65. **Orchard, S., M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, and N. Del-Toro.** 2014. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res **42:**D358-D363.
66. **Overbaugh, J.** 2020. Understanding protection from SARS-CoV-2 by studying reinfection. Nat Med **26:**1680-1681.
67. **Panda, A., D. Acharya, and T. C. Ghosh.** 2017. Insights into human intrinsically disordered proteins from their gene expression profile. Mol BioSyst **13:**2521-2530.
68. **Peri, S., J. D. Navarro, T. Z. Kristiansen, R. Amanchy, V. Surendranath, B. Muthusamy, T. K. B. Gandhi, K. N. Chandrika, N. Deshpande, and S. Suresh.** 2004. Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res **32:**D497-D501.
69. **Pushker, R., C. Mooney, N. E. Davey, J.-M. Jacqué, and D. C. Shields.** 2013. Marked variability in the extent of protein disorder within and between viral families. PloS one **8:**e60724.
70. **Rasmussen, A. L.** 2021. On the origins of SARS-CoV-2. Nat Med **27:**9-9.

71. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European molecular biology open software suite. Trends Genet **16:**276-277.
72. **Ruiz-Aravena, M., C. McKee, A. Gamble, T. Lunn, A. Morris, C. E. Snedden, C. K. Yinda, J. R. Port, D. W. Buchholz, and Y. Y. Yeo.** 2021. Ecology, evolution and spillover of coronaviruses from bats. Nat Rev Microbiol**:**1-16.
73. **Samavarchi-Tehrani, P., H. Abdouni, J. D. R. Knight, A. Astori, R. Samson, Z.-Y. Lin, D.-K. Kim, J. J. Knapp, J. St-Germain, and C. D. Go.** 2020. A SARS-CoV-2–host proximity interactome. BioRxiv.
74. **Shang, J., G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li.** 2020. Structural basis of receptor recognition by SARS-CoV-2. Nature **581:**221-224.
75. **Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker.** 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13:**2498-2504.
76. **Shilts, J., T. W. M. Crozier, A. Teixeira-Silva, I. Gabaev, P. P. Gerber, E. J. D. Greenwood, S. J. Watson, B. M. Ortmann, C. M. Gawden-Bone, and T. Pauzaite.** 2023. LRRC15 mediates an accessory interaction with the SARS-CoV-2 spike protein. PLoS biology **21:**e3001959.
77. **Singh, D., and S. V. Yi.** 2021. On the origin and evolution of SARS-CoV-2. Exp Mol Med **53:**537-547.
78. **Singh, J., S. A. Rahman, N. Z. Ehtesham, S. Hira, and S. E. Hasnain.** 2021. SARS-CoV-2 variants of concern are emerging in India. Nat Med **27:**1131-1133.
79. **Sironi, M., S. E. Hasnain, B. Rosenthal, T. Phan, F. Luciani, M.-A. Shaw, M. A. Sallum, M. E. Mirhashemi, S. Morand, and F. González-Candelas.** 2020. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. Infect Genet Evol **84:**104384.
80. **St-Germain, J. R., A. Astori, P. Samavarchi-Tehrani, H. Abdouni, V. Macwan, D.-K. Kim, J. J. Knapp, F. P. Roth, A.-C. Gingras, and B. Raught.** 2020. A SARS-CoV-2 BioID-based virus-host membrane protein interactome and virus peptide compendium: new proteomics resources for COVID-19 research. BioRxiv.
81. **Stukalov, A., V. Girault, V. Grass, O. Karayel, V. Bergant, C. Urban, D. A. Haas, Y. Huang, L. Oubraham, and A. Wang.** 2021. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. Nature **594:**246-252.
82. **Tarighi, P., S. Eftekhari, M. Chizari, M. Sabernavaei, D. Jafari, and P. Mirzabeigi.** 2021. A review of potential suggested drugs for coronavirus disease (COVID-19) treatment. Eur J Pharmacol **895:**173890.
83. **Tregoning, J. S., E. S. Brown, H. M. Cheeseman, K. E. Flight, S. L. Higham, N. M. Lemm, B. F. Pierce, D. C. Stirling, Z. Wang, and K. M. Pollock.** 2020. Vaccines for COVID-19. Clin Exp Immunol **202:**162-192.
84. **Uhlén, M., L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, and A. Asplund.** 2015. Tissue-based map of the human proteome. Science **347:**1260419.
85. **UniProt, C.** 2015. UniProt: a hub for protein information. Nucleic Acids Res **43:**D204-D212.
86. **Van Der Lee, R., M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, and D. T. Jones.** 2014. Classification of intrinsically disordered regions and proteins. Chem Rev **114:**6589-6631.
87. **Victor, M. P., D. Acharya, T. Begum, and T. C. Ghosh.** 2019. The optimization of mRNA expression level by its intrinsic properties—insights from codon usage pattern and structural stability of mRNA. Genomics **111:**1292-1297.

88. **Victor, M. P., D. Acharya, S. Chakraborty, and T. C. Ghosh.** 2020. The combined influence of codon composition and tRNA copy number regulates translational efficiency by influencing synonymous nucleotide substitution. Gene **745:**144640.
89. **Walsh, D., and I. Mohr.** 2011. Viral subversion of the host protein synthesis machinery. Nat Rev Microbiol **9:**860-875.
90. **Wang, L.-F., Z. Shi, S. Zhang, H. Field, P. Daszak, and B. T. Eaton.** 2006. Review of bats and SARS. Emerg Infect Dis **12:**1834.
91. **Wang, M., R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, and G. Xiao.** 2020. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res **30:**269-271.
92. **Wang, N., Y. Sun, R. Feng, Y. Wang, Y. Guo, L. Zhang, Y.-Q. Deng, L. Wang, Z. Cui, and L. Cao.** 2021. Structure-based development of human antibody cocktails against SARS-CoV-2. Cell Res **31:**101-103.
93. **Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, and R. Bennett.** 2020. Ensembl 2020. Nucleic Acids Res **48:**D682-D688.
94. **Zhang, Y.-Z., and E. C. Holmes.** 2020. A genomic perspective on the origin and emergence of SARS-CoV-2. Cell **181:**223-227.
95. **Zheng, D., Q. Luo, and B. W. Ritchie.** 2021. Afraid to travel after COVID-19? Self-protection, coping and resilience against pandemic 'travel fear'. Tour Manag **83:**104261.
96. **Zhou, N., J. Bao, and Y. Ning.** 2021. H2V: a database of human genes and proteins that respond to SARS-CoV-2, SARS-CoV, and MERS-CoV infection. BMC Bioinformatics **22:**1-10.
97. **Zhou, P., X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, and C.-L. Huang.** 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature **579:**270-273.
98. **Zhou, Y., Y. Hou, J. Shen, Y. Huang, W. Martin, and F. Cheng.** 2020. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov **6:**14.

Legends to the figures

Figure 1. The percentage of A. hubs and B. bottlenecks in human proteins grouped according to the coronavirus interacting status. All the differences are statistically significant except the marked ones.

Figure 2. Evolutionary rate of human proteins grouped according to the coronavirus interacting status. dN and dN/dS values with 1:1 mouse orthologs were used to calculate the evolutionary rate of proteins.

Figure 3. Functional enrichment analysis of human proteins interacting with SARS-CoV2 compared to Other-CoVs, based on Gene Ontology (GO) enrichment in the GOrilla webserver.

Figure 4. Codon usage correlation of MERS-CoV, SARS-CoV1 and SARS-CoV2 with the human housekeeping (HK) and the lung-specific (Lung) genes, using Pearson correlation coefficient.

Table 1. The status of protein intrinsic disorder and disordered protein binding regions of coronavirus-interacting human proteins.

Interaction status of human proteins	Mean long disordered regions	Average residues in long disordered regions	Average number of disordered residues	Mean long disorder-binding regions ANCHOR	Mean MoRF regions	Protein length
Noninteracting	1.019	86.179	131.702	0.888	7.10	517.62
SARS-CoV2	1.405	116.288	178.951	1.239	7.46	688.72
Other CoV	1.159	98.856	151.905	1.052	6.11	571.77
Both (SARS-CoV2 and Other CoV interacting)	1.191	92.898	152.430	0.997	7.25	696.25
P- value (Kruskal-Wallis)	8.684×10^{-24}	1.057×10^{-21}	1.946×10^{-42}	1.48×10^{-20}	1.83×10^{-2}	7.89×10^{-152}

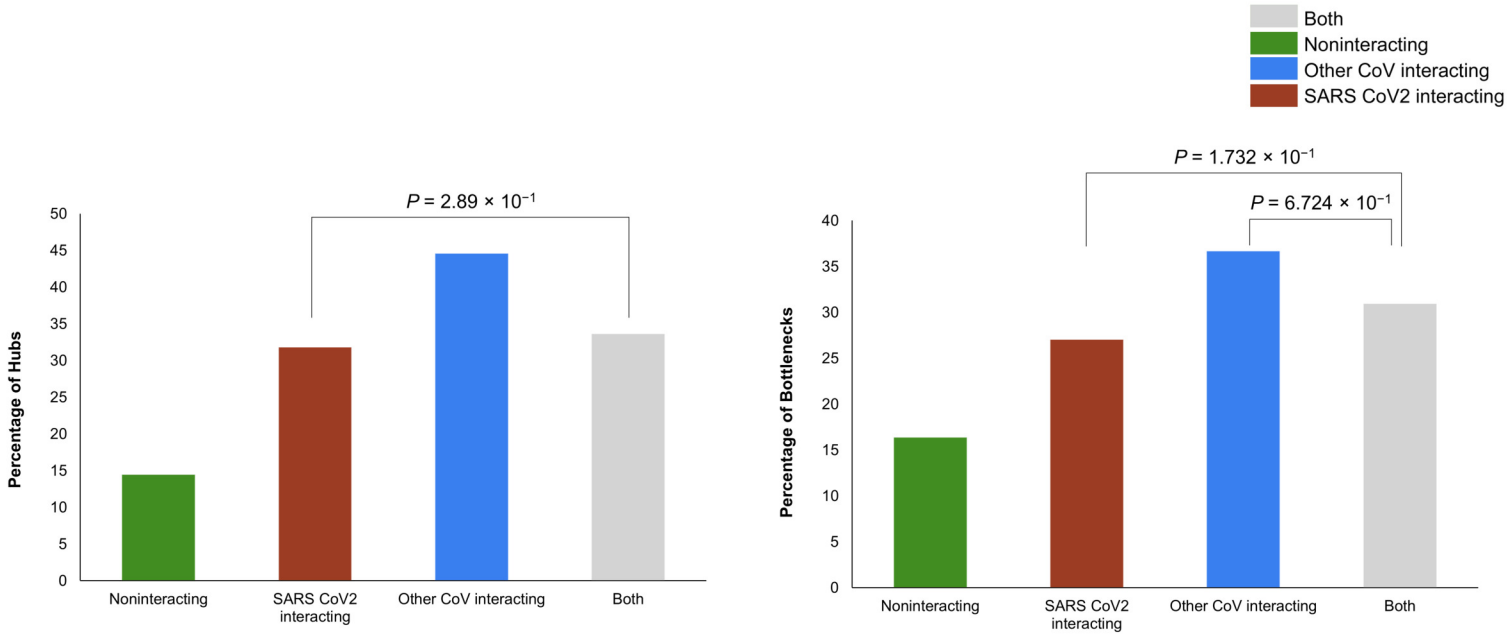
Both
Noninteracting
Other CoV interacting
SARS CoV2 interacting
P = 2.89 × 10−1
Percentage of Hubs
50
45
40
35
30
25
20
15
10
5
0
Noninteracting
SARS CoV2 interacting
Other CoV interacting
Both
P = 1.732 × 10−1
P = 6.724 × 10−1
Percentage of Bottlenecks
40
35
30
25
20
15
10
5
0
Noninteracting
SARS CoV2 interacting
Other CoV interacting
Both

Figure 1

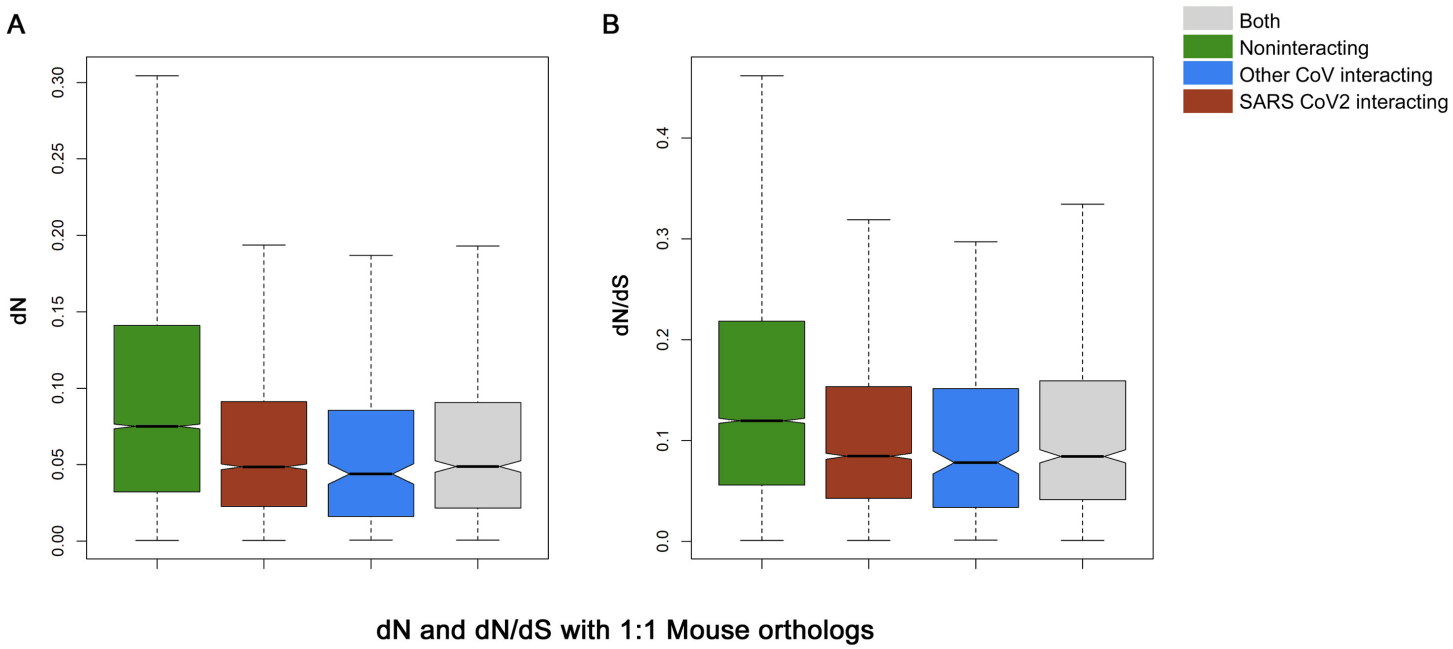
A
B
dN
0.00
0.05
0.10
0.15
0.20
0.25
0.30
dN/dS
0.0
0.1
0.2
0.3
0.4
Both
Noninteracting
Other CoV interacting
SARS CoV2 interacting
dN and dN/dS with 1:1 Mouse orthologs

Figure 2

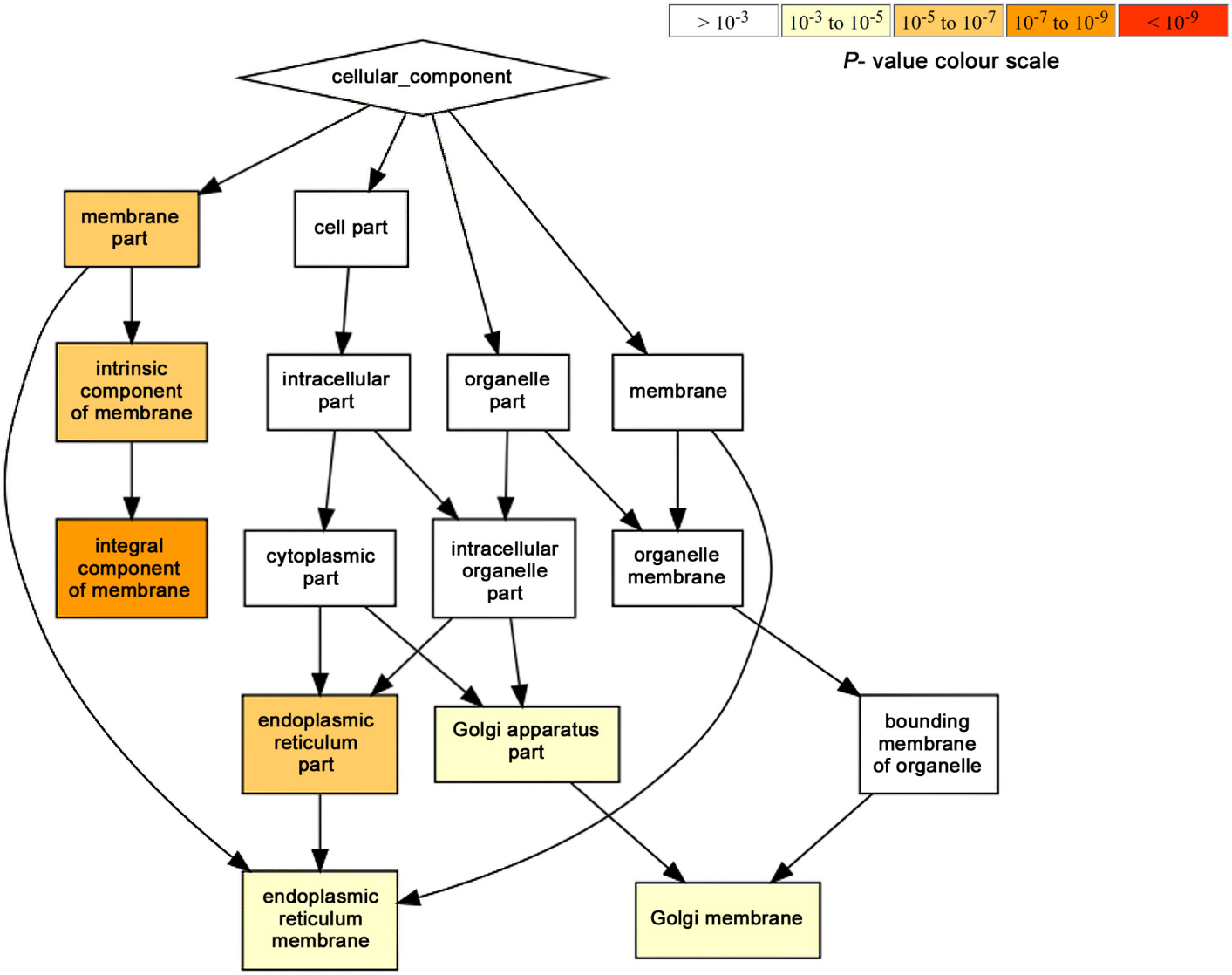
> 10-3
10-3 to 10-5
10-5 to 10-7
10-7 to 10-9
< 10-9
P- value colour scale
cellular_component
membrane part
cell part
intrinsic component of membrane
intracellular part
organelle part
membrane
integral component of membrane
cytoplasmic part
intracellular organelle part
organelle membrane
endoplasmic reticulum part
Golgi apparatus part
bounding membrane of organelle
endoplasmic reticulum membrane
Golgi membrane

Figure 3

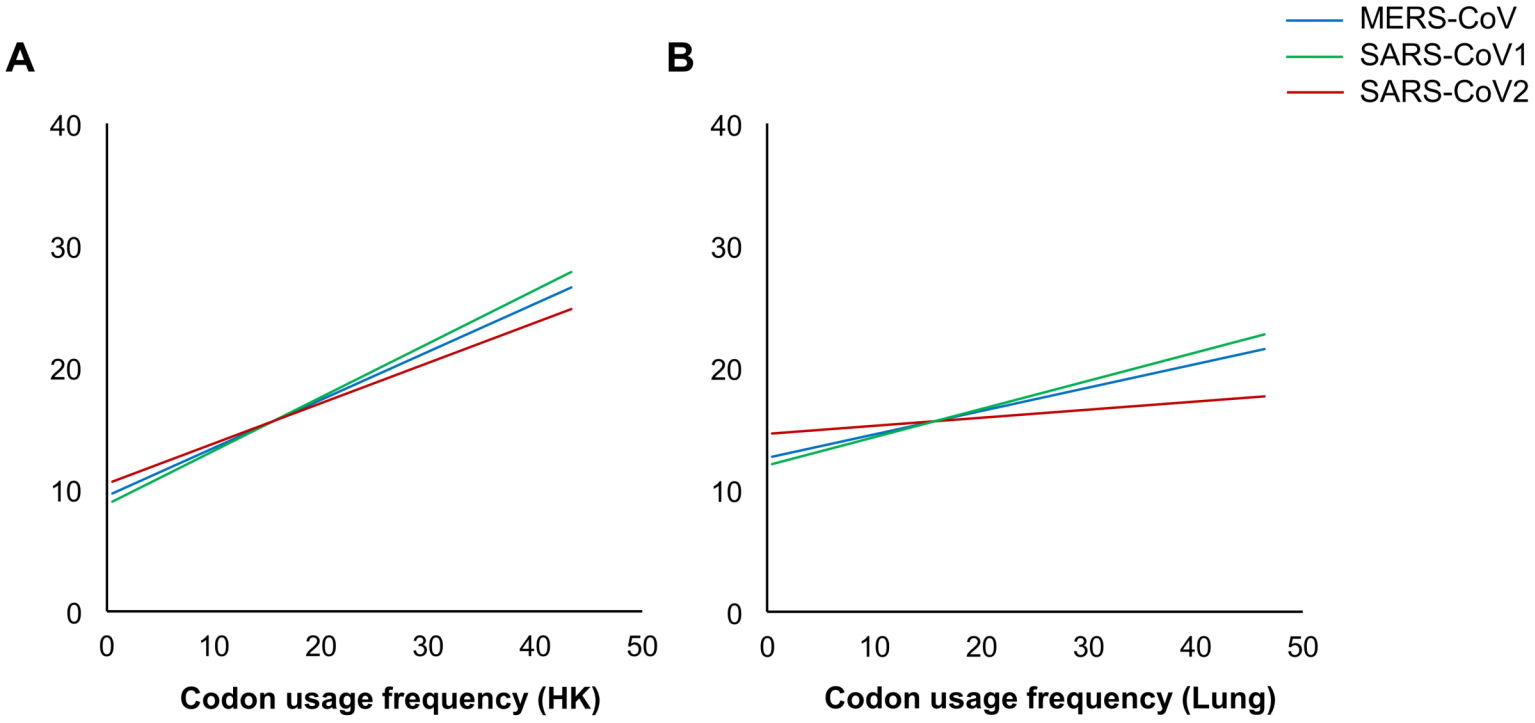
A
B
MERS-CoV
SARS-CoV1
SARS-CoV2
40
30
20
10
0
0
10
20
30
40
50
Codon usage frequency (HK)
Codon usage frequency (Lung)

Figure 4