Exploring the Human-Coronavirus protein-protein interaction network from the perspective of a novel host-virus association

Authors:

Debarun Acharya[1, 2]* (debarunacharya@gmail.com)

Tapan K Dutta[1]* (tapan@jcbose.ac.in)

Affiliations:

[1]Department of Biological Sciences,
Bose Institute
EN-80, Sector V, Salt Lake, Kolkata - 700091, West Bengal, India

[2]Institute of Health Sciences,
Presidency University (2nd Campus)
Plot No. DG/02/02, Premises No. 14-0358, Action Area – ID, New Town, Kolkata - 700156

*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 association of severe acute respiratory syndrome (SARS) coronaviruses (CoVs) with their human host from the genomic and evolutionary perspectives based on a comparative analysis of SARS and MERS coronaviruses. We observed that human proteins that are part of the SARS-CoV2-human association are enriched in hubs and bottlenecks. Again, these proteins take part in more protein complexes and show slower evolutionary rates compared to the human proteins associated with the two other coronaviruses, SARS-CoV and MERS-CoV. Moreover, the human proteins involved in the interaction with SARS-CoV2 are mostly longer proteins harboring long intrinsically disordered stretches and a higher level of disordered protein binding sites. Codon usage analysis revealed that the novel coronavirus is least adapted to codons used in housekeeping and lung-specific genes, compared to the other two coronaviruses. We conclude that the signatures showed by the SARS-CoV2-human protein interaction network revealed the virus's association with vital human proteins and pathways, via interactions mediated by protein complexes and intrinsically disordered protein binding sites, which may have assisted the higher infectivity of SARS-CoV2 in its human host than the other two less-virulent human coronaviruses, despite having a lower optimization to its host's codons.

Keywords: Host-pathogen interaction; SARS-CoV2; Protein-protein interaction network; Protein intrinsic disorder; Codon usage.

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 first occurrence of the SARS-CoV2 infection in 2019, a plethora of research articles have been published to explore the origin and the nature of the virus (Hu et al., 2021; Li et al., 2020; Pagani et al., 2023; Rasmussen, 2021; Singh & Yi, 2021), the molecular basis underlying the pathogenesis and infection (Jackson et al., 2022; Lamers & Haagmans, 2022; McCallum et al., 2021; Niu et al., 2021; Shang et al., 2020), the epidemiology of the disease (Alteri et al., 2021; Dhar et al., 2021; Esper et al., 2021), the impact on human population (Aburto et al., 2021; Aburto et al., 2022; Jin et al., 2021), the protection strategies from the virus (Abdoul-Azize & El Gamil, 2021; Overbaugh, 2020; Zheng et al., 2021), and the genetic variations and evolution (Li et al., 2020; Markov et al., 2023; Roemer et al., 2023) among the many aspects of SARS-CoV2 infection. The high infectivity and mortality rates among SARS-CoV2 patients influenced these studies and accelerated the development of therapeutic and prophylactic measurements against SARS-CoV2 throughout the globe (Andreano & Rappuoli, 2021; Hotez et al., 2021). This pandemic situation further initiated the discovery of efficient treatments like convalescent plasma therapy (Chen et al., 2020; Focosi et al., 2020), the use of monoclonal antibodies (Lloyd et al., 2021; Marovich et al., 2020; Wang et al., 2021), the development of vaccines (Krammer, 2020; Ndwandwe & Wiysonge, 2021; Tregoning et al., 2020), and the repurposing of drugs (Consortium, 2021; Jang et al., 2021; Li & De Clercq, 2020; Tarighi et al., 2021; Wang et al., 2020; Y. Zhou et al., 2020). About 300 repurposed drugs (Menestrina et al., 2021) and 50 vaccines (https://covid19.trackvaccines.org/vaccines/approved/#vaccine-list) were approved globally to combat the deadly virus as of June 2025. However, the continuous accumulation of mutations in SARS-CoV-2 has posed significant challenges to therapeutic strategies, contributing to the emergence of multiple variants— from Alpha and Beta in December 2020 to Omicron in November 2021 and more recently to subvariants such as JN.1, KP.3, KP.3.1.1, LB.1, XEC, LP.8.1, and NB.1.8.1 observed throughout 2023–2025—many of which exhibit increased transmissibility or immune evasion potential, showing a high adaptability of this virus (Lippi et al., 2022; Roemer et al., 2023; Singh et al., 2021; Who, 2025).

The horseshoe bats (*Rhinolophus* spp.) are the reservoir hosts of coronaviruses, including the SARS-CoV and SARS-CoV2 (Ruiz-Aravena et al., 2021; Wang et al., 2006; P. Zhou et al., 2020), with a few reports indicating pangolins as the intermediate host (Lam et al., 2020; Liu et

al., 2020; Zhang & Holmes, 2020). The high genomic sequence identity (~99.99%) of SARS-CoV2 viruses isolated from different patients suggests a very recent host shift to humans. Although both SARS-CoV and SARS-CoV2 share their entry receptor, angiotensin-converting enzyme 2 (ACE2), that binds to the spike protein of both coronaviruses, the affinity for ACE2 is 10-20 times higher for the SARS-CoV2 spike protein than the SARS-CoV counterpart explaining the higher infectivity of the former (Abdolmaleki et al., 2022).

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 (Fieldman, 2024). 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 (Lippi et al., 2022; Singh et al., 2021).

The novel coronavirus SARS-CoV2 is the newest member of the *Coronaviridae* family, belonging to the *Sarbecovirus* subgenus (genus *Betacoronavirus*) (Lefkowitz et al., 2018). 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 (Sironi et al., 2020). 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 (Liu et al., 2021). 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 (Omrani et al., 2015). 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, as the recent pandemic caused by SARS-CoV2 led to over 7 million deaths and about 780 million infected persons globally as of June 2025. The

SARS-CoV2 has 79.7% sequence similarity with the SARS-CoV of 2003 (Y. Zhou et al., 2020), 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-CoVs and MERS-CoV may control the severity of these infections, influencing the biological function of the human proteins and their associated pathways, via the protein-protein interaction (PPI) network. Moreover, other preventive measures of the disease such as passive immunization or vaccination certainly help to stop the disease transmission or may imply further selection pressure to generate vaccine-resistant variants (Luo et al., 2021). The underlying cause of rapid propagation of SARS-CoV2 across human populations suggests that SARS-CoV2 may have an upper hand in regulating the human PPI network, via the establishment of a stable pathogen-host PPI network. Comprehensive host-pathogen protein-protein interaction network analysis for human pathogens have suggested that pathogen proteins often interact with human proteins with a high degree centrality, which is consistent for viral, bacterial, protozoan, and fungal pathogens (Acharya & Dutta, 2021; Crua Asensio et al., 2017; Dyer et al., 2008; Nithya et al., 2021). Considering the coronaviruses, the novel coronavirus SARS-CoV2 has recently shown a global outbreak, and a more rapid adaptation than the other two coronaviruses (MERS-CoV and SARS-CoV) in a very short span. As a result, this rapid adaptation of the SARS-CoV2 should be evident in the pathogen-human protein-protein interaction (PHPPI) network involving the human and SARS-CoV2 proteins. Using high-throughput affinity purification mass spectrometry, Gordon et al. first mapped the interactions between 26 SARS-CoV2 proteins and 332 human proteins, revealing key host targets involved in translation, vesicle trafficking, and immune signaling (Gordon et al., 2020). Such analyses help identify functional protein modules and prioritize druggable host factors, offering potential therapeutic targets beyond viral components. Integration of SARS-CoV-2 PPIs with host systems biology—such as protein complexes and signaling pathways—further enables understanding of the broader cellular impact and network perturbations caused by infection (Stukalov et al., 2021). On the contrary, the MERS-CoV has mostly circulated among humans from camels and SARS-CoV circulated among humans for a very short time (Omrani et al., 2015) and their association with human proteins in the PHPPI network should reflect a weaker association. Concerning the higher severity of SARS-CoV2 infections, SARS-CoV2 may have a higher control over host cellular systems, despite being novel, due to its rapid adaptability and propagation within human populations. In this study, we have analyzed the recent human

coronavirus interactome data for MERS-CoV, SARS-CoV and SARS-CoV2 from the perspective of protein-protein interactions between human and coronaviruses. The study aims to compare these three human coronaviruses from a network biology perspective accounting for the roles of human proteins, their network topology and structural features that facilitate these interspecies interactions.

2. Materials and methods

2.1. Sequence data

The coding sequences (CDS) for MERS and SARS coronaviruses were obtained using the NCBI GenBank (Benson et al., 2012) (NCBI Reference Sequences NC_019843.3, NC_004718.3 and NC_045512 for MERS-CoV, SARS-CoV and SARS-CoV2, respectively). The Human CDS sequences were obtained using the Biomart interface of Ensembl (Yates et al., 2020). Human protein sequences were obtained from the UniProt (UniProt, 2015) [proteome ID: UP000005640]. We have selected the 'reviewed' proteome only, which leads to a list of 20410 proteins and their corresponding amino acid sequences.

2.2. Protein-protein interaction data

The coronavirus-human protein-protein interaction data for MERS-CoV, SARS-CoV, and SARS-CoV2 were obtained from H2V (Zhou et al., 2021) and BioGRID (Oughtred et al., 2021) database. In BioGRID, we have filtered the interaction data and included only experimentally validated PPI data involving large-scale experiments (involving >10 PPI pairs) that show direct evidence of interactions, by incorporating the following experimental methods: Affinity Capture-MS, Affinity Capture-Western, Biochemical Activity, Co-crystal Structure, Cross-Linking-MS, PCA, Proximity Label-MS, Reconstituted Complex and Two-hybrid. We have integratred the filtered human-virus interacting pairs obtained from H2V and BioGRID databases and removed the redundant interacting pairs. The final list of Human-coronavirus PPI data contains 694 Human-MERS-CoV, 1451 Human-SARS-CoV and 12134 Human-SARS-CoV2 binary protein-protein interaction pairs, involving 572, 1142 and 3168 human proteins, respectively. The proteins interacting with both SARS-CoV and SARS-CoV2 (N= 490) were also considered, but the proteins interacting with all coronaviruses (MERS-CoV, SARS-CoV and SARS-CoV2), and those interacting with MERS-CoV and any SARS coronavirus (MERS-CoV and SARS-CoV, MERS-CoV and SARS-CoV2) were discarded to remove any influence of MERS-CoV in the

results. The proteins without any involvement of interaction of any of the coronaviruses under study were considered as a 'Noninteracting' group (N= 16566).

The human protein-protein interaction network was obtained from the supplementary data of Cheng *et al.* (Cheng et al., 2018), which provides a comprehensive list of human protein-protein interaction networks collected from established PPI databases. With this data, we have included the updated human binary protein-protein interaction data from BioGRID (version 4.4.245) (Oughtred et al., 2021), which includes only the validated PPIs from experiments showing direct evidence of interaction, as mentioned earlier. The Entrez gene IDs were converted to corresponding UniProt IDs using UniProt's ID conversion tool (UniProt, 2015). The final data contains 372895 interactions involving 16358 human proteins. The human PPI network was analyzed using the NetworkAnalyzer plugin of Cytoscape 3.10.1 (Shannon et al., 2003) to determine the degree and betweenness centrality of human proteins. The top 20% proteins having the highest degree and betweenness centrality were considered hubs and bottlenecks, respectively (Acharya & Dutta, 2021).

2.3. Protein complex association

The human protein complexes were obtained from the Corum database (release 5.1) (Tsitsiridis et al., 2023). The human protein complex data provides a unique number for each protein complex and the UniProt ID of the participants of the complex. The unique number of complexes involving a protein was considered as its complex number, which was calculated using in-house Python script. We have obtained a list of 5366 human protein complexes involving 5051 human proteins that have matched with 4799 proteins in total from the five groups of our study.

2.4. Protein intrinsic disorder and disordered protein binding sites

The intrinsic disorder of human proteins was predicted using the IUPred2A (Mészáros et al., 2018), 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 long-disordered regions. An in-house PERL script was used to determine such long-disordered regions of human proteins, with a consecutive stretch of 30 or more amino acids having a disorder score of ≥ 0.5 for all the amino acids in the stretch (Acharya & Dutta, 2021). The long-disordered protein binding regions were calculated using the ANCHOR2 under IUPred2A, an additional tool designed to predict disordered binding regions within intrinsically disordered

regions. For each protein sequence, ANCHOR2 provides a per-residue score indicating the likelihood of that residue participating in a disordered protein-protein interaction interface. Disordered stretches within a protein containing short stretches of amino acid residues with ANCHOR2 scores ≥ 0.5 were classified as the putative disordered protein binding regions. Two quantitative metrics were extracted from each protein's ANCHOR2 profile: the ANCHOR regions: the total number of regions with ≥ 5 residues with ANCHOR2 scores ≥ 0.5 within the disordered stretches and total ANCHOR residues per proteins that fall within the disordered residues and have ANCHOR2 scores ≥ 0.5 (Dosztányi et al., 2009). In addition, the Molecular Recognition Features (MoRFs) of human disordered proteins were obtained from the fMoRFpred webserver (Yan et al., 2016), where each amino acid in a disordered protein is classified as MoRF or non-MoRF residue. Similar to ANCHOR, the consecutive stretches of ≥ 5 MoRF residues were considered as regions responsible for disordered protein binding (MoRF regions) and the number of such regions in each human protein, along with the total MoRF residues in a protein was calculated using an in-house PERL script.

2.5. Evolutionary Rate

The evolutionary rates of human proteins were obtained using the Ensemblbiomart (Yates et al., 2020), using one-to-one orthologs of human proteins with four other species: Chimpanzee (*Pan troglodytes*), Macaque (*Macacamulatta*), Pig (*Susscrofa*), and Mouse (*Musmusculus*). We obtained the nonsynonymous nucleotide substitutions per nonsynonymous sites (dN) and synonymous nucleotide substitutions per synonymous sites (dS) of the orthologous pairs from Ensembl biomart. The dN/dS ratio for each human protein, with their 1:1 orthologous proteins in all the species considered, were used to measure the evolutionary rate. The mutation saturation was controlled by removing all proteins having dS values >3 (Acharya & Ghosh, 2016).

2.6. Pathway enrichment analysis

The pathway enrichment analyses were conducted using the clusterProfiler R package(Yu et al., 2012). The UniProt IDs were converted to Entrez Gene IDs using the org.Hs.eg.db annotation database in R (Carlson et al., 2019). The resulting gene sets were grouped based on viral interaction status (MERS-CoV, SARS-CoV and SARS-CoV2), producing three distinct input lists for comparative analysis. The pathway enrichment analysis was conducted using Gene Ontology (GO) enrichment for the Biological Process (BP) ontology (Gene Ontology, 2019) using the enrichGO function, with the Benjamini–Hochberg method for multiple testing correction. We have also performed KEGG Pathway (Kanehisa et al., 2025) enrichment using the enrichKEGG function and adjusted p-values via the Benjamini–Hochberg method.

Enrichment results for both GO terms and KEGG pathways were visualized using a dotplot from the enrichplot package. Dot plots were faceted by the virus group to enable direct visual comparison of the top ten significantly enriched terms.

2.7. Housekeeping and lung-specific genes

Human housekeeping genes were obtained from the HRTAtlas 1.0 database (Hounkpe et al., 2021), 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 (Uhlén et al., 2015). We have selected a final list of 190 protein-coding genes that are selectively elevated in the lungs.

2.8. Codon usage

The frequency of each codon was calculated using the ‘cusp’ function of the EMBOSS, an open-source software package (Rice et al., 2000). 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 of the 61 codons in the provided input sequence, per 1000 bases (i.e., codon frequency) was 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-CoV, HK-SARS-CoV, HK-SARS-CoV2 and Lung-MERS-CoV, Lung-SARS-CoV, Lung-SARS-CoV2).

2.9. Statistical Packages and Software

The statistical analyses have been carried out using IBM SPSS Statistics 22 (Meyers et al., 2013), R packages. Statistical analyses were conducted using non-parametric methods due to the non-normal distribution of the data. For comparisons between two independent groups, the Mann-Whitney U test was applied to evaluate differences in central tendency. For comparisons involving more than two independent groups, the Kruskal-Wallis H test was used to assess whether group medians differed significantly. Jonckheere-Terpstra test was conducted to detect ordered differences across groups. All statistical tests were two-tailed, and a p-value of less than 0.05 was considered indicative of statistical significance. The codon usage correlation

analyses have been carried out using the Pearson correlation coefficient. In house PERL, Python and R scripts were also used in data analyses.

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, resulting in 2302 human proteins interacting with SARS-CoV2 only, 480 proteins interacting with SARS-CoV only, 490 proteins interacting with both the SARS-CoVs, 146 proteins interacting with MERS-CoV only, and 16566 non-interacting proteins which imply no reported interaction with any of these coronaviruses (see materials and methods). The central proteins in the human protein-protein interaction (PPI) network were identified from the constructed human high-throughput PPI data, from where the hubs and the bottlenecks were classified using degree and betweenness centrality values, respectively (see materials and methods). Our comparative analysis revealed that the human proteins that do not interact with any coronavirus proteins (noninteracting group) contain the lowest proportion of hubs and bottlenecks (Figure 1). Among the coronavirus-interacting groups, human proteins interacting with SARS-CoV2 contain a higher proportion of hubs and bottlenecks, followed by MERS-CoV and SARS-CoV. The result indicates two observations: firstly, the human coronaviruses mostly target host central proteins with higher connectivity, as revealed in earlier studies with other pathogens (Calderwood et al., 2007; Dyer et al., 2008). Secondly, the SARS-CoV2-interacting human proteins have a higher proportion of hubs and bottlenecks than those interacting with other two coronaviruses. This suggests that the SARS-CoV2 may have a better opportunity to hijack the host cellular network more efficiently compared to the other two coronaviruses under study, which may be helpful for the optimization of viral propagation and virulence, despite its recent zoonotic origin.

3.2. Protein complex association

The human cellular interaction network is composed of many proteins that often form complexes that perform particular cellular functions with the help of its participants. The protein complex association reflects how a protein is integrated into various macromolecular complexes to carry out the specific functions of those complexes. The protein's participation in multiple complexes, referred to as the protein complex number (PCN) (Chakraborty & Ghosh, 2013), serves as an

appropriate measure of the number of functions (multifunctionality) in which the protein is involved. Proteins that are involved in many complexes, are important molecular players of the cell. Here, the protein complex association of human proteins interacting with different coronaviruses revealed that the complex number of human proteins follows the following trend regarding the five groups under study: SARS-CoV2 (PCN= 4.149) > SARS-CoV_and_SARS-CoV2 (PCN= 3.836) > MERS-CoV (PCN= 3.658) > SARS-CoV (PCN= 3.341) > Noninteracting (PCN= 3.363). To assess the significance of this directional trend in protein complex number across the ordered viral interaction groups, the Jonckheere-Terpstra test was employed to confirm a statistically significant increasing trend (P= 1.43 × 10^{-2}, Jonckheere-Terpstra test). This suggests that the virus-interacting human proteins have a higher protein complex association. Moreover, among these viruses, SARS-CoV2 proteins can hijack proteins participating in more complexes within the human PPI network. Nevertheless, to survive within the human host, the interactions between these host proteins and viral proteins are crucial, which warrants the exploration of the structural features of host proteins that are assisting in such interactions and the relevance of such features considering human-coronavirus interactions.

3.3. Intrinsic disorder and disordered protein binding regions

Protein intrinsic disorder plays an important role in the host-pathogen protein-protein interaction network as it provides flexibility to protein structure to promote low-affinity interactions between proteins (Acharya & Dutta, 2021; Panda et al., 2017). The increased flexibility of the host's disordered proteins allows host-pathogen protein interactions, which otherwise would remain infeasible (Acharya & Dutta, 2021). For a detailed understanding of this scenario, human-coronavirus protein-protein interactions were examined using the IUPred2A (Mészáros et al., 2018) protein intrinsic disorder prediction algorithm with regards to the coronaviruses interacting and noninteracting groups. Interestingly, the human proteins interacting with SARS-CoV2 showed a significantly higher number of long disordered stretches and intrinsically disordered residues than those interacting with the other two coronaviruses, followed by the non-interacting group (Table 1). The SARS-CoV2 interacting proteins were also found to contain a higher number of long disordered protein binding regions, and residues that promote disordered protein binding than that of the other two coronaviruses, as revealed by ANCHOR2 program under IUPred2A (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) than those interacting with the SARS-CoV, MERS-CoV and noninteracting groups (P= 3.677× 10^{-110},

Kruskal-Wallis test). These longer human proteins having more intrinsically disordered residues harbor disordered protein binding sites that may be utilized by the SARS-CoV2 proteins, assisting in the low-affinity protein-protein interactions with the human host.

3.4. Molecular Recognition Features

The protein-protein interactions involving the intrinsically disordered proteins are often mediated by disordered protein binding regions that undergo disorder-to-order transition during the physical interaction with their binding partners. These regions, better known as Molecular Recognition Features (MoRFs), contain amino acid residues that are important for protein-protein interactions (Disfani et al., 2012; Van Der Lee et al., 2014). 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 (Panda et al., 2017), including the pathogen-host protein-protein interactions during infections (Acharya & Dutta, 2021).The abundance of MoRF residues in human-disordered proteins interacting with coronaviruses was calculated using the fMoRFpred webserver (Yan et al., 2016). Similar to our ANCHOR2 results, it was observed that the SARS-CoV2-interacting human disordered proteins contain a higher number of MoRF regions and residues, which may contribute to their interaction with viral proteins (Table 1). This also suggests that the SARS-CoV2 proteins involved in human-virus interactions utilize the disordered protein-binding MoRF regions of host proteins more extensively than those of other coronaviruses, thereby facilitating their association by promoting low-affinity interactions.

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 (Drummond et al., 2005). 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. To explore further, we have studied the evolutionary rate of human proteins interacting with different coronaviruses, using 1:1 orthologs of four species: Chimpanzee, Macaque, Pig and Mouse (see materials and methods). Among all the groups, the noninteracting group was found to evolve most rapidly, and among the coronaviruses, SARS-CoV-interacting human proteins were found to evolve faster than SARS-CoV2 (Figure 2, $P < 0.05$ for all 1:1 orthologous group, Kruskal-Wallis test), suggesting that SARS-CoV2 interacts with human proteins encoded by

evolutionarily more conserved genes, and therefore has a better opportunity for long-term association with its human host.

3.6. Pathway enrichment analysis

During the viral infection, the host proteins that interact with the viral proteins can offer insights into the virus-utilized host pathways and functions. To explore the functional and pathway enrichments in the three coronaviruses under study, the R package clusterProfiler (Yu et al., 2012) was used to compare human proteins interacting only with one of the three coronaviruses. Enrichment analysis using the Gene Ontology Biological Process suggests that the proteins interacting with SARS-CoV2 were significantly enriched in processes like ribosome biogenesis, organelle localization, and RNA metabolism-related processes, including RNA splicing, and RNA processing, highlighting a strong association with RNA regulatory machinery. Furthermore, enrichment of chromosome segregation and spindle organization profiles suggest a possible influence of SARS-CoV2 in host cell cycle regulation (Figure 3A). In contrast, SARS-CoV showed enrichment in processes linked to chaperone-mediated protein folding and protein glycosylation, implicating modulation of host protein quality control and post-translational modification systems (Figure 3A). On the other hand, proteins interacting with MERS-CoV (MERS) were enriched in ribosome biogenesis, mitochondrial translation, and protein localization to the nucleus, indicating a functional preference for components of the host translational machinery and nuclear transport. In addition, they are also involved in viral transport and DNA structure alteration (Figure 3A).

KEGG pathway enrichment further differentiated the host functional landscapes targeted by each of the viruses. The SARS-CoV2-interacting human proteins showed enrichment in nucleocytoplasmic transport, endocytosis, cell cycle and lysosome, highlighting a role in the cellular transport system (Figure 3B). In contrast, SARS-CoV-associated proteins showed enrichment in protein processing and disease-related pathways such as Alzheimer's disease and prion disease suggesting potential overlap with host neurodegenerative mechanisms (Figure 3B). While MERS-CoV-interacting proteins were enriched in pathways including nucleocytoplasmic transport, chemical carcinogenesis and non-homologous end joining, pointing to a possible engagement with host DNA damage response and oncogenic signaling routes (Figure 3B). Together, these enrichment patterns suggest that SARS-CoV2 engages extensively with host intracellular transport and regulatory machinery, potentially facilitating efficient viral replication and immune evasion, and may reflect its broader host adaptability and fitness compared to SARS-CoV and MERS-CoV.

3.7. Codon usage

The function of genes relies on their encoded proteins, which fold into the desired three-dimensional structure attributed to the interactions between its constituent amino acids. However, the desired expression level of a protein is crucial for its function, often represented as the 'translational efficiency'. 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 (Victor et al., 2019). Viral pathogens, being dependent on host cellular machinery and the host's aminoacylated tRNAs for translating of their proteins, should prefer a higher translation rate to facilitate their multiplication and propagation in new host cells/tissues (Nambou et al., 2022). This can be facilitated by using a similar codon usage pattern with host proteins. Here we used the EMBOSS tool to determine the frequency of each codon from MERS-CoV, SARS-CoV, and SARS-CoV2. The codon usage pattern was correlated with human housekeeping [HK] (N= 2070) and lung-specific [Lung] (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_{HK}= 0.259; PCC_{lung}= 0.052) in contrast to the MERS-CoV (PCC_{HK}= 0.349; PCC_{lung}= 0.172) and the SARS-CoV (PCC_{HK}= 0.401; PCC_{lung}= 0.213) with both the housekeeping ($P < 0.05$ for PCC), but not the lung-specific gene groups ($P > 0.05$ for PCC) (Figure 4). This suggests that both the SARS-CoV and MERS-CoV show better codon adaptation with HK genes but show an insignificant trend with Lung genes, compared to the SARS-CoV2.

Although SARS-CoV2 exhibits widespread interaction with human proteins, its lower codon adaptation to human housekeeping and lung-specific genes suggests that it has not yet undergone extensive evolutionary optimization for translational efficiency in the human host. This discrepancy may stem from the virus's recent zoonotic spillover, giving it less evolutionary time to fine-tune its codon usage in alignment with host tRNA pools. In contrast, SARS-CoV and MERS-CoV, which have either circulated longer or undergone more cycles of host adaptation, demonstrate higher codon usage correlation with human genes, potentially facilitating more efficient translation.

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 the availability of only a snapshot of these interactions at any point in time. Of particular importance are viruses that utilize their host cellular machinery to synthesize the proteins required for survival, reproduction, immune evasion, and disease progression (Walsh & Mohr, 2011). 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 may kill the host, hampering the propagation of the virus. However, with time, molecular interactions between host and virus usually get stabilized, forming a stable pathogen-host protein-protein interaction (PHPPI) network, resulting in coexistence. In this study, we compared the virus-host interactome of the novel coronavirus SARS-CoV2 and the other two human coronaviruses, namely SARS-CoV and MERS-CoV from the perspective of the PHPPI network. The central proteins were previously shown to be the prime targets of pathogens and are involved in the PHPPI network (Acharya & Dutta, 2021; Crua Asensio et al., 2017; Dyer et al., 2008). These proteins play important roles in the host cellular processes and assist pathogens to gain control over the host protein-protein interaction (PPI) network. The higher infectivity of SARS-CoV2 may have shaped their host-adaptation better than the closely related SARS-CoV. Here we observed that the SARS-CoV2 interacts with human proteins that are more central, containing a higher proportion of network hubs and bottlenecks than those interacting with other coronaviruses, namely MERS-CoV and SARS-CoV, which cause milder infections. By interacting with central proteins in the host PPI network, SARS-CoV2 increases its evolutionary fitness by gaining control over host cellular systems. Delving further, we observed that the human proteins interacting with SARS-CoV2 are associated with a higher number of protein complexes, allowing them to interact with additional functional modules in the human PPI network. However, previous studies suggested that pathogens could utilize intrinsically disordered proteins or intrinsically disordered residues to facilitate low-affinity interactions (Acharya & Dutta, 2021; Blundell et al., 2020). Therefore, we explored the presence of disordered stretches present within human proteins interacting with different coronaviruses. The PHPPI network revealed a higher involvement of intrinsically disordered proteins in the SARS-CoV2-interacting human proteins, with a higher number of disordered protein binding sites promoting such low-affinity interactions, compared to the other two coronaviruses (Table 1). Moreover, we also observed that the protein length is also a major determinant of host-virus

interaction, as longer proteins have a probability of harboring more intrinsically disordered residues and disordered protein binding motifs (Table 1). So, their abundance, but not density, is the key determinant, and their presence may facilitate the host-pathogen interactions. Considering the protein evolutionary rates, human proteins interacting with SARS-CoV2 are more evolutionarily conserved than the two other coronaviruses, suggesting that its PHPPI network will favour a long-term host association (Brito & Pinney, 2017).

In addition, to survive and multiply in the host's body, the viral pathogens must utilize the host's cellular machinery. The comparative pathway-level analysis using three coronaviruses indicates that SARS-CoV2-interacting human proteins are associated with crucial biological pathways in humans, including ribosome biogenesis, organelle localization, RNA metabolism-related processes, chromosome segregation, spindle organization, cell cycle and nucleocytoplasmic transport that are central to host cellular functions. Furthermore, another crucial component of viral adaptation to the host comes from the host codon usage preferences (Bahir et al., 2009; Nambou et al., 2022). The host tRNA pool and codon usage remain optimized for the desired expression of the host proteins (Victor et al., 2020). For a novel host-virus association, this is merely a chance factor that the host tRNA pool will support the codons used by the viral proteome. However, with time, viruses may adapt to host codon usage preferences to maintain optimum efficiency in their multiplication (Bahir et al., 2009). We correlated the coronavirus codon usage pattern with that of the human housekeeping genes and the genes enriched in the lung tissues for each of the coronaviruses. 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 for the optimized expression of lung-specific genes. A lower correlation (PCC) in the codon usage pattern of SARS-CoV2 with the human housekeeping genes was observed, followed by MERS-CoV and SARS-CoV. Moreover, SARS-CoV2's broad tissue tropism and rapid transmissibility may be driven by mechanisms beyond codon optimization, such as its ability to modulate host metabolism and immune evasion pathways to compensate for suboptimal translational compatibility.

However, this study has several inherent limitations that should be acknowledged. First, the protein-protein interaction (PPI) data used in this study are representative static snapshots from heterogeneous sources, limiting the dynamic representation of host-pathogen coevolution. The interactome data are also disproportionately enriched for SARS-CoV2 due to its recent global

impact and a huge number of studies in comparison to the limited datasets for SARS-CoV and MERS-CoV, potentially biasing comparative analyses. Furthermore, this study does not include strains/variants in the viral PPI datasets or temporal PPI dynamics for each virus due to the lack of specific information. Nevertheless, within the circumstantial limitations, this study highlights the importance of human-coronavirus protein-protein interaction and the role of protein structural features that can assist SARS-CoV2 with greater transmission and host-cellular control, despite sharing close genomic sequence identity with the less-virulent SARS-CoV.

5. Conclusion

This comparative analysis of host-pathogen protein-protein interactions (PPIs) among three major human coronaviruses reveals critical insights into viral strategies for host exploitation and adaptation. Our findings demonstrate that SARS-CoV2 preferentially targets central and functionally versatile human proteins, engages with a higher number of protein complexes, and utilizes intrinsically disordered regions and longer host proteins to form dynamic, low-affinity interactions. These characteristics enhance its capacity to interface with multiple host pathways and functional modules. Despite showing the lowest codon adaptation to the host's translational machinery among the three viruses, SARS-CoV2 compensates by targeting essential host pathways, including those involved in RNA processing, cell cycle control, and nucleocytoplasmic transport—thereby optimizing its replication and propagation. This contrast between structural interaction patterns and translational compatibility reflects a unique adaptation strategy that may support the high transmissibility of SARS-CoV2, compared to SARS-CoV and MERS-CoV.

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.

Declaration of Competing Interest

The authors declare no competing interests.

Data availability

Data will be made available on request.

Acknowledgement

The authors thank the Director, Bose Institute, Kolkata, India for encouragement and support.

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:

Abdolmaleki, G., Taheri, M. A., Paridehpour, S., Mohammadi, N. M., Tabatabaei, Y. A., Mousavi, T., & Amin, M. (2022). A comparison between SARS-CoV-1 and SARS-CoV2: an update on current COVID-19 vaccines. *DARU Journal of Pharmaceutical Sciences*, *30*(2), 379-406.

Abdoul-Azize, H. T., & El Gamil, R. (2021). Social protection as a key tool in crisis management: learnt lessons from the COVID-19 pandemic. *Glob Soc Welf*, *8*(1), 107-116.

Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., & Dowd, J. B. (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*(8), 735-740.

Aburto, J. M., Schöley, J., Kashnitsky, I., Zhang, L., Rahal, C., Missov, T. I., Mills, M. C., Dowd, J. B., & Kashyap, R. (2022). Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int J Epidemiol*, *51*(1), 63-74.

Acharya, D., & Dutta, T. K. (2021). Elucidating the network features and evolutionary attributes of intra- and interspecific protein–protein interactions between human and pathogenic bacteria. *Sci Rep*, *11*(1), 190. https://doi.org/10.1038/s41598-020-80549-x

Acharya, D., & Ghosh, T. C. (2016). Global analysis of human duplicated genes reveals the relative importance of whole-genome duplicates originated in the early vertebrate evolution [journal article]. *BMC Genomics*, *17*(1), 1-14. https://doi.org/10.1186/s12864-016-2392-0

Alteri, C., Cento, V., Piralla, A., Costabile, V., Tallarita, M., Colagrossi, L., Renica, S., Giardina, F., Novazzi, F., & Gaiarsa, S. (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), 1-13.

Andreano, E., & Rappuoli, R. (2021). SARS-CoV-2 escaped natural immunity, raising questions about vaccines and therapies. *Nat Med*, *27*(5), 759-761.

Bahir, I., Fromer, M., Prat, Y., & Linial, M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol*, *5*, 311. https://doi.org/10.1038/msb.2009.71

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Res*, *41*(D1), D36-D42.

Blundell, T. L., Gupta, M. N., & Hasnain, S. E. (2020). Intrinsic disorder in proteins: Relevance to protein assemblies, drug design and host-pathogen interactions. *Progress in Biophysics and Molecular Biology*, *156*, 34-42.

Brito, A. F., & Pinney, J. W. (2017). Protein–protein interactions in virus–host systems. *Frontiers in microbiology*, *8*, 1557.

Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., Ewence, A. E., Li, N., Hirozane-Kishikawa, T., & Hill, D. E. (2007). Epstein–Barr virus and virus human protein interaction maps. *Proceedings of the National Academy of Sciences*, *104*(18), 7606-7611.

Carlson, M., Falcon, S., Pages, H., & Li, N. (2019). org. Hs. eg. db: Genome wide annotation for Human. *R package version*, *3*(2), 3.

Chakraborty, S., & Ghosh, T. C. (2013). Evolutionary rate heterogeneity of core and attachment proteins in yeast protein complexes. *Genome Biol Evol*, *5*(7), 1366-1375. https://doi.org/10.1093/gbe/evt096

Chen, L., Xiong, J., Bao, L., & Shi, Y. (2020). Convalescent plasma as a potential therapy for COVID-19. *Lancet Infect Dis*, *20*(4), 398-400.

Cheng, F., Desai, R. J., Handy, D. E., Wang, R., Schneeweiss, S., Barabási, A.-L., & Loscalzo, J. (2018). Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*, *9*(1), 2691. https://doi.org/10.1038/s41467-018-05116-5

Consortium, W. H. O. S. T. (2021). Repurposed antiviral drugs for Covid-19—interim WHO solidarity trial results. *N Engl J Med*, *384*(6), 497-511.

Crua Asensio, N., Munoz Giner, E., De Groot, N. S., & Torrent Burgas, M. (2017). Centrality in the host–pathogen interactome is associated with pathogen fitness during infection. *Nat Commun*, *8*(1), 1-6.

Dhar, M. S., Marwal, R., Vs, R., Ponnusamy, K., Jolly, B., Bhoyar, R. C., Sardana, V., Naushin, S., Rophina, M., & Mellan, T. A. (2021). Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science*, *374*(6570), 995-999.

Disfani, F. M., Hsu, W.-L., Mizianty, M. J., Oldfield, C. J., Xue, B., Dunker, A. K., Uversky, V. N., & Kurgan, L. (2012). MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics*, *28*(12), i75-i83. https://doi.org/10.1093/bioinformatics/bts209

Dosztányi, Z., Mészáros, B., & Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, *25*(20), 2745-2746.

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., & Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA*, *102*(40), 14338-14343.

Dyer, M. D., Murali, T. M., & Sobral, B. W. (2008). The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog*, *4*(2), e32.

Esper, F. P., Cheng, Y.-W., Adhikari, T. M., Tu, Z. J., Li, D., Li, E. A., Farkas, D. H., Procop, G. W., Ko, J. S., & Chan, T. A. (2021). Genomic epidemiology of SARS-CoV-2 infection during the initial pandemic wave and association with disease severity. *JAMA network open*, *4*(4), e217746-e217746.

Fieldman, T. (2024). Evolutionary principles for modifying pathogen virulence. *Critical Reviews in Microbiology*, *50*(4), 385-396.

Focosi, D., Anderson, A. O., Tang, J. W., & Tuccori, M. (2020). Convalescent plasma therapy for COVID-19: state of the art. *Clin Microbiol Rev*, *33*(4), e00072-00020.

Gene Ontology, C. (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res*, *47*(D1), D330-D338.

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., & Swaney, D. L. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, *583*(7816), 459-468.

Hotez, P. J., Batista, C., Amor, Y. B., Ergonul, O., Figueroa, J. P., Gilbert, S., Gursel, M., Hassanain, M., Kang, G., & Kaslow, D. C. (2021). Global public health security and justice for vaccines and therapeutics in the COVID-19 pandemic. *EClinicalMedicine*, *39*, 101053.

Hounkpe, B. W., Chenou, F., de Lima, F., & De Paula, E. V. (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*(D1), D947-D955.

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*, *19*(3), 141-154.

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature reviews Molecular cell biology*, *23*(1), 3-20.

Jang, W. D., Jeon, S., Kim, S., & Lee, S. Y. (2021). Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay. *Proc Natl Acad Sci USA*, *118*(30).

Jin, J., Agarwala, N., Kundu, P., Harvey, B., Zhang, Y., Wallace, E., & Chatterjee, N. (2021). Individual and community-level risk for COVID-19 mortality in the United States. *Nat Med*, *27*(2), 264-269.

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. *Nucleic acids research*, *53*(D1), D672-D677.

Krammer, F. (2020). SARS-CoV-2 vaccines in development. *Nature*, *586*(7830), 516-527.

Lam, T. T.-Y., Jia, N., Zhang, Y.-W., Shum, M. H.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., & Liao, Y.-S. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, *583*(7815), 282-285.

Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis. *Nature reviews microbiology*, *20*(5), 270-284.

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res*, *46*(D1), D708-D717.

Li, C., Yang, Y., & Ren, L. (2020). Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. *Infect Genet Evol*, *82*, 104285.

Li, G., & De Clercq, E. (2020). Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat Rev Drug Discov*, *19*(3), 149-150.

Lippi, G., Mattiuzzi, C., & Henry, B. M. (2022). Updated picture of SARS-CoV-2 variants and mutations. *Diagnosis*, *9*(1), 11-17.

Liu, D. X., Liang, J. Q., & Fung, T. S. (2021). Human coronavirus-229e,-oc43,-nl63, and-hku1 (coronaviridae). *Encyclopedia of Virology*, 428.

Liu, P., Jiang, J.-Z., Wan, X.-F., Hua, Y., Li, L., Zhou, J., Wang, X., Hou, F., Chen, J., & Zou, J. (2020). Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog*, *16*(5), e1008421.

Lloyd, E. C., Gandhi, T. N., & Petty, L. A. (2021). Monoclonal antibodies for COVID-19. *Jama*, *325*(10), 1015-1015.

Luo, R. B., Delaunay-Moisan, A., Timmis, K., & Danchin, A. (2021). SARS-CoV-2 biology and variants: anticipation of viral evolution and what needs to be done. *Environ Microbiol*, *23*(5), 2339-2363. https://doi.org/10.1111/1462-2920.15487

Markov, P. V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N. I., & Katzourakis, A. (2023). The evolution of SARS-CoV-2. *Nature reviews microbiology*, *21*(6), 361-379.

Marovich, M., Mascola, J. R., & Cohen, M. S. (2020). Monoclonal antibodies for prevention and treatment of COVID-19. *Jama*, *324*(2), 131-132.

McCallum, M., Walls, A. C., Sprouse, K. R., Bowen, J. E., Rosen, L. E., Dang, H. V., De Marco, A., Franko, N., Tilles, S. W., & Logue, J. (2021). Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science*, *374*(6575), 1621-1626.

Menestrina, L., Cabrelle, C., & Recanatini, M. (2021). COVIDrugNet: a network-based web tool to investigate the drugs currently in clinical trial to contrast COVID-19. *Sci Rep*, *11*(1), 1-15.

Mészáros, B., Erdős, G., & Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research*, *46*(W1), W329-W337.

Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2013). *Performing data analysis using IBM SPSS*. John Wiley & Sons.

Nambou, K., Anakpa, M., & Tong, Y. S. (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*(2), 97-115.

Ndwandwe, D., & Wiysonge, C. S. (2021). COVID-19 vaccines. *Curr Opin Immunol*, *71*, 111-116.

Nithya, C., Kiran, M., & Nagarajaram, H. A. (2021). Comparative analysis of Pure Hubs and Pure Bottlenecks in Human Protein-protein Interaction Networks. *bioRxiv*.

Niu, S., Wang, J., Bai, B., Wu, L., Zheng, A., Chen, Q., Du, P., Han, P., Zhang, Y., & Jia, Y. (2021). Molecular basis of cross-species ACE2 interactions with SARS-CoV-2-like viruses of pangolin origin. *EMBO J*, *40*(16), e107786.

Omrani, A. S., Al-Tawfiq, J. A., & Memish, Z. A. (2015). Middle East respiratory syndrome coronavirus (MERS-CoV): animal to human interaction. *Pathogens and global health*, *109*(8), 354-362.

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., & Zhang, F. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science*, *30*(1), 187-200.

Overbaugh, J. (2020). Understanding protection from SARS-CoV-2 by studying reinfection. *Nat Med*, *26*(11), 1680-1681.

Pagani, I., Ghezzi, S., Alberti, S., Poli, G., & Vicenzi, E. (2023). Origin and evolution of SARS-CoV-2. *The European Physical Journal Plus*, *138*(2), 157.

Panda, A., Acharya, D., & Ghosh, T. C. (2017). Insights into human intrinsically disordered proteins from their gene expression profile. *Mol BioSyst*, *13*(12), 2521-2530.

Rasmussen, A. L. (2021). On the origins of SARS-CoV-2. *Nat Med*, *27*(1), 9-9.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet*, *16*(6), 276-277.

Roemer, C., Sheward, D. J., Hisner, R., Gueli, F., Sakaguchi, H., Frohberg, N., Schoenmakers, J., Sato, K., O'Toole, Á., & Rambaut, A. (2023). SARS-CoV-2 evolution in the Omicron era. *Nature microbiology*, *8*(11), 1952-1959.

Ruiz-Aravena, M., McKee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C. E., Yinda, C. K., Port, J. R., Buchholz, D. W., & Yeo, Y. Y. (2021). Ecology, evolution and spillover of coronaviruses from bats. *Nat Rev Microbiol*, 1-16.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221-224.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, *13*(11), 2498-2504.

Singh, D., & Yi, S. V. (2021). On the origin and evolution of SARS-CoV-2. *Exp Mol Med*, *53*(4), 537-547.

Singh, J., Rahman, S. A., Ehtesham, N. Z., Hira, S., & Hasnain, S. E. (2021). SARS-CoV-2 variants of concern are emerging in India. *Nat Med*, *27*(7), 1131-1133.

Sironi, M., Hasnain, S. E., Rosenthal, B., Phan, T., Luciani, F., Shaw, M.-A., Sallum, M. A., Mirhashemi, M. E., Morand, S., & González-Candelas, F. (2020). SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. *Infect Genet Evol*, *84*, 104384.

Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., Haas, D. A., Huang, Y., Oubraham, L., & Wang, A. (2021). Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature*, *594*(7862), 246-252.

Tarighi, P., Eftekhari, S., Chizari, M., Sabernavaei, M., Jafari, D., & Mirzabeigi, P. (2021). A review of potential suggested drugs for coronavirus disease (COVID-19) treatment. *Eur J Pharmacol*, *895*, 173890.

Tregoning, J. S., Brown, E. S., Cheeseman, H. M., Flight, K. E., Higham, S. L., Lemm, N. M., Pierce, B. F., Stirling, D. C., Wang, Z., & Pollock, K. M. (2020). Vaccines for COVID-19. *Clin Exp Immunol 202*(2), 162-192.

Tsitsiridis, G., Steinkamp, R., Giurgiu, M., Brauner, B., Fobo, G., Frishman, G., Montrone, C., & Ruepp, A. (2023). CORUM: the comprehensive resource of mammalian protein complexes–2022. *Nucleic acids research*, *51*(D1), D539-D545.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., & Asplund, A. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419.

UniProt, C. (2015). UniProt: a hub for protein information. *Nucleic Acids Res*, *43*(D1), D204-D212.

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., & Jones, D. T. (2014). Classification of intrinsically disordered regions and proteins. *Chem Rev*, *114*(13), 6589-6631.

Victor, M. P., Acharya, D., Begum, T., & Ghosh, T. C. (2019). The optimization of mRNA expression level by its intrinsic properties—insights from codon usage pattern and structural stability of mRNA. *Genomics*, *111*(6), 1292-1297.

Victor, M. P., Acharya, D., Chakraborty, S., & Ghosh, T. C. (2020). The combined influence of codon composition and tRNA copy number regulates translational efficiency by influencing synonymous nucleotide substitution. *Gene*, *745*, 144640.

Walsh, D., & Mohr, I. (2011). Viral subversion of the host protein synthesis machinery. *Nat Rev Microbiol*, *9*(12), 860-875.

Wang, L.-F., Shi, Z., Zhang, S., Field, H., Daszak, P., & Eaton, B. T. (2006). Review of bats and SARS. *Emerg Infect Dis*, *12*(12), 1834.

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., & Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*, *30*(3), 269-271.

Wang, N., Sun, Y., Feng, R., Wang, Y., Guo, Y., Zhang, L., Deng, Y.-Q., Wang, L., Cui, Z., & Cao, L. (2021). Structure-based development of human antibody cocktails against SARS-CoV-2. *Cell Res*, *31*(1), 101-103.

Who, W. H. O. (2025). Tracking SARS-CoV-2 variants. *World Health Organization*.

Yan, J., Dunker, A. K., Uversky, V. N., & Kurgan, L. (2016). Molecular recognition features (MoRFs) in three domains of life. *Molecular BioSystems*, *12*(3), 697-710.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., & Bennett, R. (2020). Ensembl 2020. *Nucleic Acids Res*, *48*(D1), D682-D688.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, *16*(5), 284-287.

Zhang, Y.-Z., & Holmes, E. C. (2020). A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell*, *181*(2), 223-227.

Zheng, D., Luo, Q., & Ritchie, B. W. (2021). Afraid to travel after COVID-19? Self-protection, coping and resilience against pandemic 'travel fear'. *Tour Manag*, *83*, 104261.

Zhou, N., Bao, J., & Ning, Y. (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), 1-10.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., & Huang, C.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270-273.

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov*, *6*(1), 14. https://doi.org/10.1038/s41421-020-0153-3

Legends to the figures

Figure 1. Percentage of **A.** hubs and **B.** bottlenecks in human proteins, grouped according to the coronavirus interacting status. All the differences are statistically significant ($P < 0.05$).

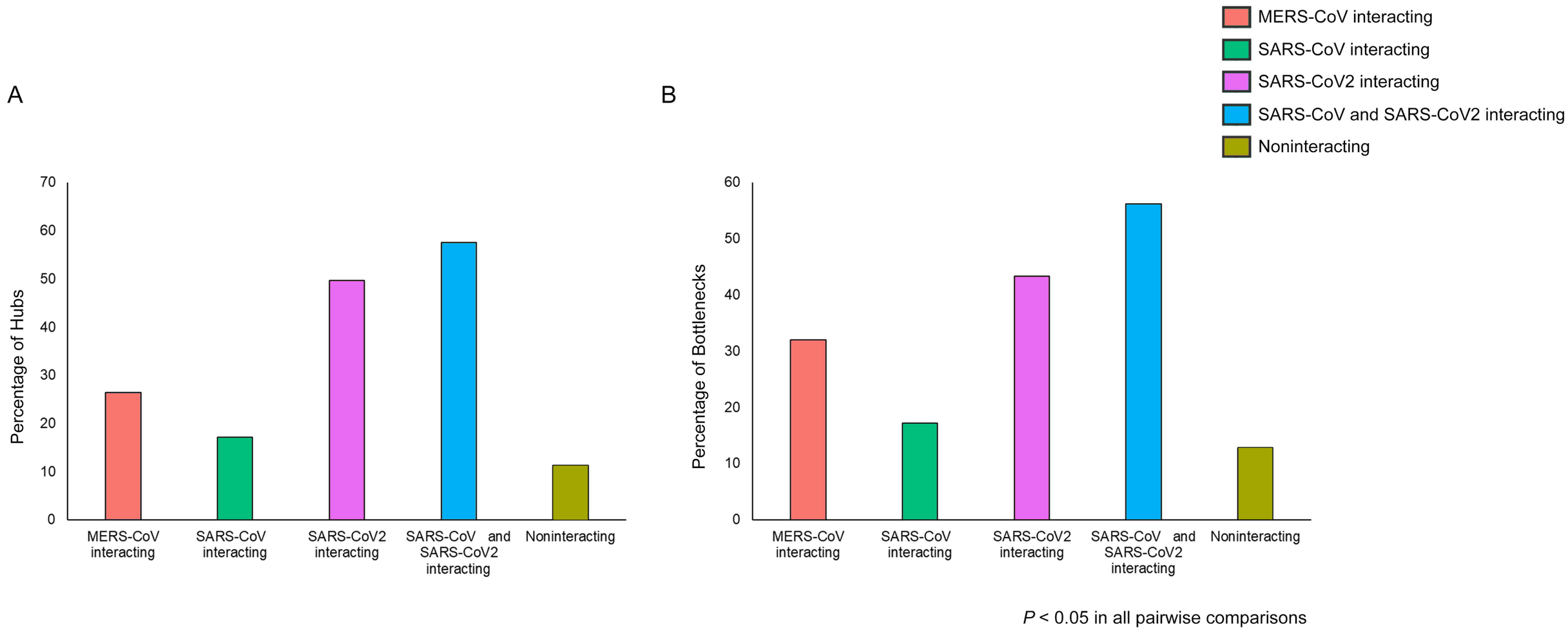
Figure 2. Evolutionary rate of human proteins, grouped according to the coronavirus interacting status. dN/dS values with 1:1 orthologs considering **A.** Mouse, **B.** Pig, **C.** Macaque and **D.** Chimpanzee were used to calculate the evolutionary rate of proteins.

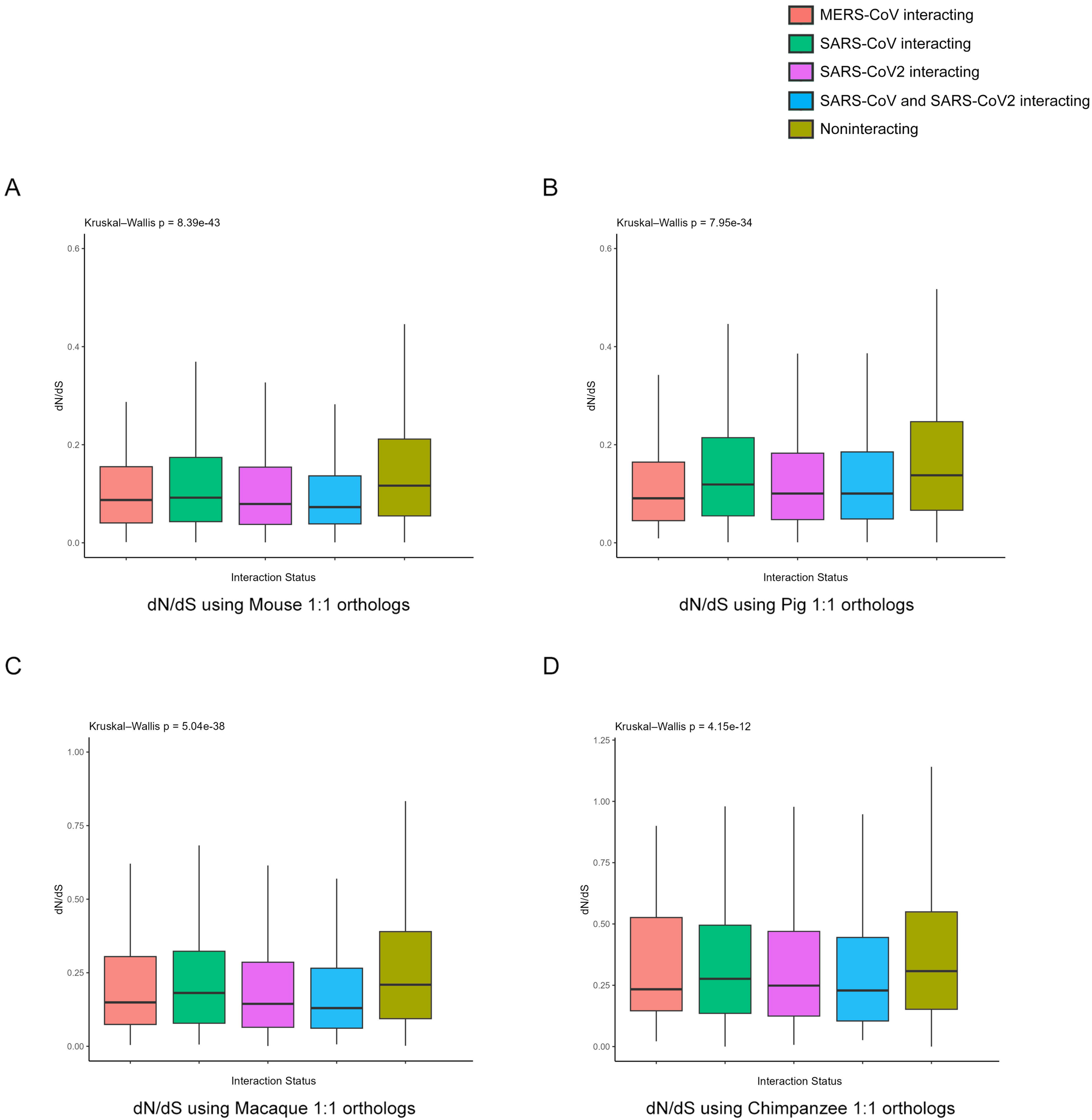
Figure 3. Pathway analysis of human genes interacting with the three coronaviruses: MERS-CoV, SARS-CoV and SARS-CoV2 using **A.** Gene Ontology (GO) Biological Process and **B.** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Dot plots display the top 10 significantly enriched pathways, with Benjamini–Hochberg adjusted P-values represented by colour gradients.

Figure 4. Codon usage correlation of MERS-CoV, SARS-CoV and SARS-CoV2 with the human genes **A.** housekeeping genes and **B.** lung-specific genes, using Pearson correlation coefficient. MERS-CoV is represented in blue while SARS CoV in black and SARS-CoV2 in red.

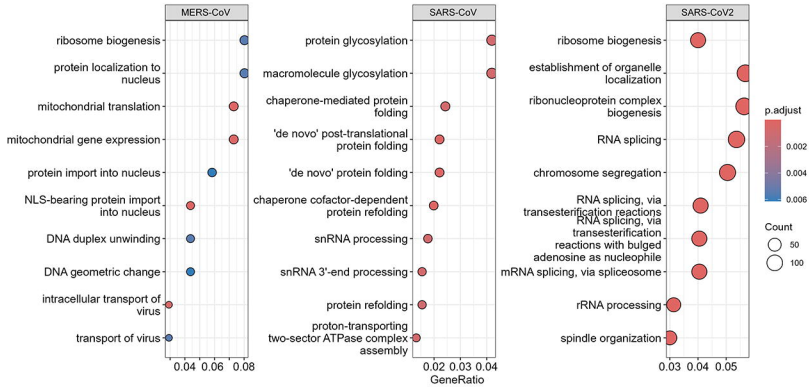
Table 1. The status of protein intrinsic disorder and disordered protein binding regions of coronavirus-interacting human proteins.

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 ANCHOR2	Average No of all disordered binding residues	Mean MoRF regions	Average number of MoRF residues	Protein length
MERS-CoV interacting	1.123	99.712	148.651	2.870	90.192	0.943	13.486	566.25
SARS-CoV interacting	1.177	98.650	154.635	2.854	91.860	0.941	13.137	643.79
SARS-CoV2 interacting	1.626	139.521	206.838	3.836	128.163	1.021	14.433	725.61
SARS-CoV and SARS-CoV2 interacting	1.371	119.716	179.506	3.233	108.345	0.929	13.256	686.71
Noninteracting	1.017	84.755	130.925	2.376	80.017	0.924	12.585	523.76
P- value (Kruskal-Wallis)	1.356×10^{-55}	2.420×10^{-55}	1.595×10^{-73}	1.298×10^{-53}	6.298×10^{-50}	4.295×10^{-2}	4.00×10^{-6}	3.677×10^{-110}


MERS-CoV interacting
SARS-CoV interacting
SARS-CoV2 interacting
SARS-CoV and SARS-CoV2 interacting
Noninteracting
A
Percentage of Hubs
70
60
50
40
30
20
10
0
MERS-CoV interacting
SARS-CoV interacting
SARS-CoV2 interacting
SARS-CoV and SARS-CoV2 interacting
Noninteracting
B
Percentage of Bottlenecks
60
50
40
30
20
10
0
MERS-CoV interacting
SARS-CoV interacting
SARS-CoV2 interacting
SARS-CoV and SARS-CoV2 interacting
Noninteracting
P < 0.05 in all pairwise comparisons

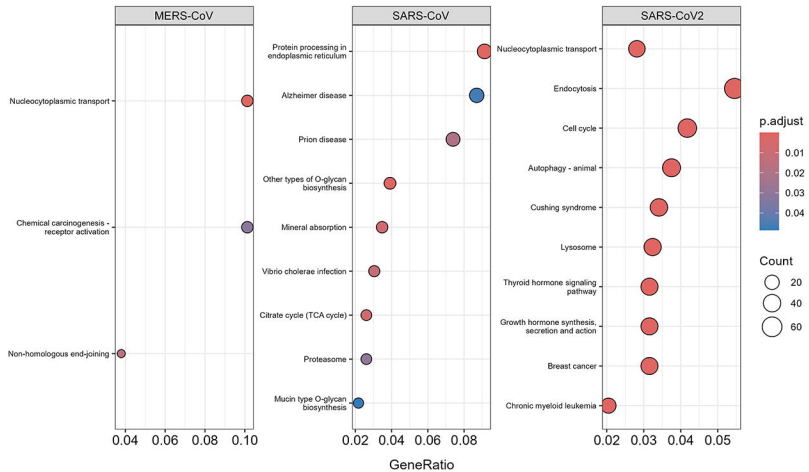

MERS-CoV interacting
SARS-CoV interacting
SARS-CoV2 interacting
SARS-CoV and SARS-CoV2 interacting
Noninteracting
A
Kruskal–Wallis p = 8.39e-43
dN/dS
0.0
0.2
0.4
0.6
Interaction Status
dN/dS using Mouse 1:1 orthologs
B
Kruskal–Wallis p = 7.95e-34
dN/dS
0.0
0.2
0.4
0.6
Interaction Status
dN/dS using Pig 1:1 orthologs
C
Kruskal–Wallis p = 5.04e-38
dN/dS
0.00
0.25
0.50
0.75
1.00
Interaction Status
dN/dS using Macaque 1:1 orthologs
D
Kruskal–Wallis p = 4.15e-12
dN/dS
0.00
0.25
0.50
0.75
1.00
1.25
Interaction Status
dN/dS using Chimpanzee 1:1 orthologs

A

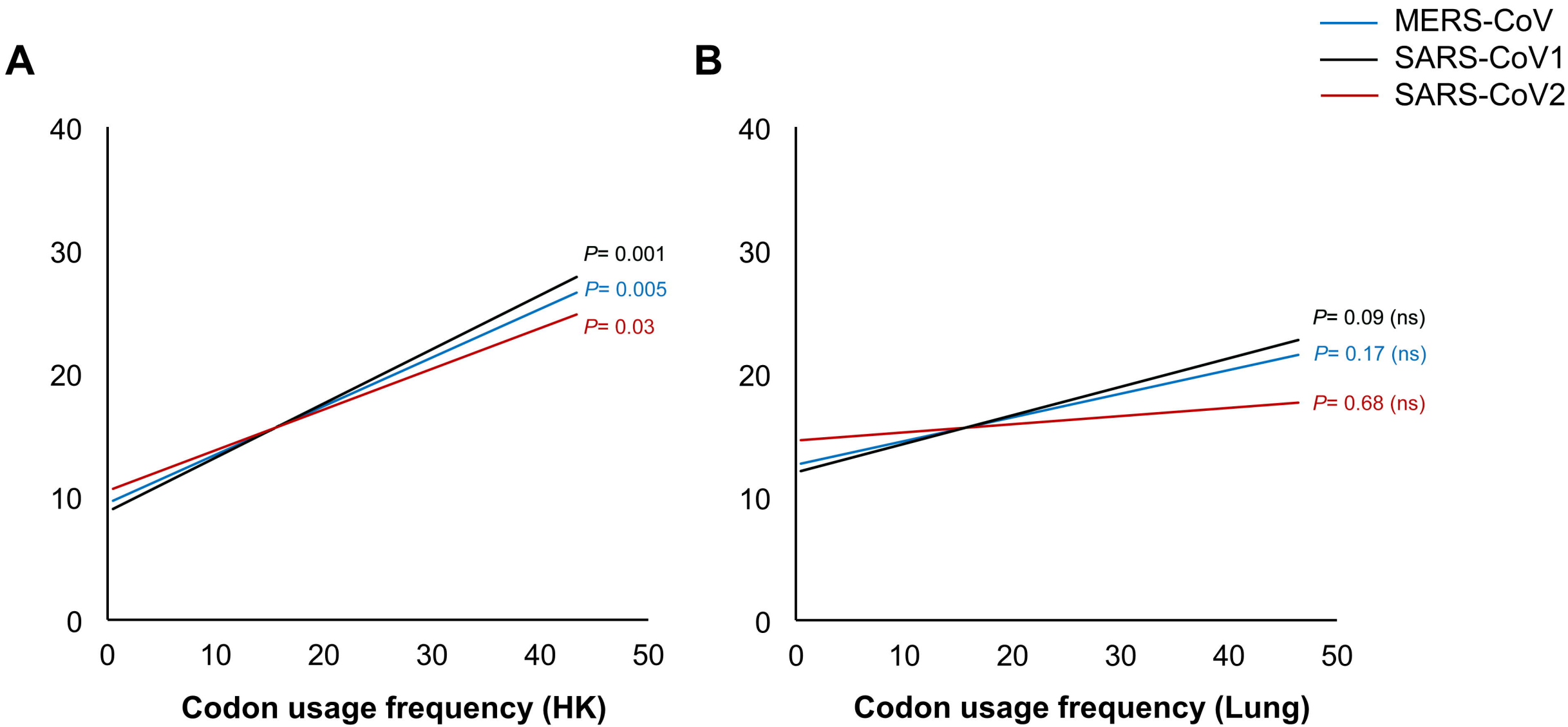


Pathway enrichment using Gene Ontology: Biological Process

B



Pathway enrichment using KEGG


A
B
MERS-CoV
SARS-CoV1
SARS-CoV2
40
30
20
10
0
0
10
20
30
40
50
P= 0.001
P= 0.005
P= 0.03
Codon usage frequency (HK)
P= 0.09 (ns)
P= 0.17 (ns)
P= 0.68 (ns)
Codon usage frequency (Lung)