

Computational, experimental, and clinical evidence for a specific but peculiar evolutionary nature of (COVID-19) SARS-CoV-2

Gerard Kian-Meng Goh,^{1,*} A. Keith Dunker,² James A. Foster,^{3,4} and Vladimir N. Uversky^{5,6}

¹ Goh's BioComputing, Singapore, Republic of Singapore

² Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana, USA.

³ Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA.

⁴ Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho, USA

⁵ Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA.

⁶ Laboratory of New Methods in Biology, Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Moscow region, Russia

*Corresponding author

Email address:

GKMG: gohsbiocomputing@yahoo.com

Abstract

SARS-CoV-2 was empirically and computationally found to be of a specific but peculiar evolution. Shell disorder models found that the outer shell (M protein) of SARS-CoV-2 to be among the hardest in its CoV family. The hard outer shell (low M percentage of disorder (PID)) is likely to be related to the SARS-CoV-2 resistance to the antimicrobial enzymes in saliva and mucus, and be responsible for the high-level of viral shedding which has been observed clinically. Experimental studies have also shown that SARS-CoV-2 is more resilient in the environment than many other CoVs, including SARS-CoV-1. Another aspect of the shell disorder models predicts that SARS-CoV-1 is more virulent than SARS-CoV-2 because of higher inner shell disorder (N PID) that helps SARS-CoV-1 replicate faster in vital organs despite being of lesser viral loads in the saliva and mucus, unlike SARS-CoV-2. This has been reaffirmed experimentally, where higher levels (50 folds) of infectious particles were detected in the SARS-CoV-1 samples in comparison with those of SARS-CoV-2. The hard outer shell of SARS-CoV-2 has been found to be associated with burrowing animals, particularly pangolins, which are often in contact with buried feces. For these reasons, the M protein is highly conserved among close relatives of SARS-CoV-2. The phylogenetic tree using M, unlike the genome-wide one, shows that pangolin-CoVs are more closely related to SARS-CoV-2 than bat-RaTG13. Previous phylogenetic studies may have been confused by recombinations that are usually poorly handled. According to the shell disorder models based on the N PID, an attenuated COVID-19 strain is likely to have entered humans via pangolins in 2017 or before, which provides the virus enough time to adapt to humans. This could explain why the SARS-CoV-2 S protein is highly adapted to the human ACE-2. The specific but peculiar evolution has a wide range of clinical, immunological, and epidemiological implications.

Keywords: pangolin; intrinsic; disorder; protein; nucleocapsid; virulence; shell; covid; coronavirus; vaccine; immune; antibody; shell; nucleoprotein; matrix; attenuate; severe acute respiratory

Introduction: the lingering mysteries of SARS-CoV-2

Some characteristics of COVID-19 and SARS-CoV-2

The first observed coronavirus disease 2019 (COVID-19) outbreak occurred in December 2019 in Wuhan, China [1]. Because the initial cases are mainly associated with a Wuhan wet market, Huanan Seafood Wholesale Market, where live animals were sold, the possibility of a zoonotic transmission occurring there was suspected [2]. The virus, labelled severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was sequenced and was found to be a betacoronavirus that has about 70% homology to the 2003 SARS-CoV (SARS-CoV-1). A search in the sample archives yielded a 2013 bat sample from Yunnan province, RaTG13, that had a 96% sequence similarity to SARS-CoV-2 [3-5]. Furthermore, CoV samples were retrieved from smuggled pangolins in Guangxi and Guangdong in 2017-8 and 2019 respectively, and they were found to have approximately 90% homology to SARS-CoV-2 [6-9].

Natural or unnatural origin?

Right from the beginning, there were questions. Where did this virus come from? When and how did it enter humans? Since the Huanan Seafood Wholesale Market market is just a few miles from the Wuhan Institute of Virology, was the virus man-made [10]? To address some of the concerns, Andersen *et al.* [11] examined the S protein, which plays an important role in the viral entry into the cell by binding to the host ACE-2 (angiotensin converting enzyme 2) receptor. The group attempted to “design” the virus using computer software and found that the actual S protein binds using a different set of residues from the computer suggested ones, which implies that the virus is simply too novel to be designed in the laboratory. The team mentioned that it is possible that the virus has been in humans for a long period of time. Other researchers have found evidence to suggest this as well. For example, computational studies have found that SARS-CoV-2 S protein binds to ACE-2 with an affinity that is 20-30 times that of SARS-CoV-1 to ACE-2 [11, 12]. Another study showed that the S protein has a binding affinity for the human ACE-2 that is greater than that of all of the other animal ACE-2 studied [13]. This begs the question: if SARS-CoV-2 entered humans some

time ago, why didn't the medical community notice it then, instead only noticing it recently, in December 2019? Could a precursor have entered the human population as an attenuated strain before mutating to its current virulent form? In fact, our shell disorder models suggest that a pangolin-CoV strain did enter humans as an attenuated SARS-CoV-2 precursor in 2017 or before. Related to the virus' high adaptation to humans is the question of its contagiousness. Why is it so contagious that it exceeds SARS-CoV-1? [2] Some scientists believe that the S protein is responsible [14]. While not necessarily contradicting this theme, we found peculiarities of SARS-CoV-2 that are likely to contribute to its infectivity.

Parent shell disorder model foresaw the successful development of COVID-19 vaccines

The parent shell disorder model that studied the intrinsic disorder of the various protein shells of viruses was initiated in 2005, when it was discovered that the human immunodeficiency virus (HIV) is likely to have abnormally disordered outer shell that is believed to be responsible for the failures in the search for an effective HIV vaccine [15-17]. Conversely, the outer shell disorder was found to not be similar to HIV, but resembled classical viruses, for which vaccines have been developed [18, 19]. This lead us to correctly foresee the successful development of effective vaccines for COVID-19 months before their approval by the Food and Drug Administration (FDA) [20, 21]. The model involves the use of artificial intelligence (AI) to detect and calculate the levels of intrinsic disorder in query proteins. In 2014, there was a spinoff of this model that detected positive correlations between inner shell disorder and virulence among variety of viruses, such as Ebola (EBOV) and Nipah (NiV) viruses [22-25]. This allows the shell disorder model to detect attenuated strains.

Shell disorder models: contagiousness, virulence, and immunity

In 2011, before the MERS-CoV outbreak, there was an initiation of yet another spinoff model, the

CoV shell disorder model [26] to predict the levels of fecal-oral and respiratory transmissions of the various CoV depending on the levels of disorder at mainly the inner shell (N). The model predicted SARS-CoV-1 to be of intermediate fecal-oral and respiratory potentials. Upon the various outbreaks, the model was further validated with MERS-CoV being placed in a category of CoVs with lower respiratory but higher fecal-oral transmission potentials [27], whereas SARS-CoV-2, like SARS-CoV-1, has both intermediate fecal-oral and respiratory transmission potentials. SARS-CoV-2 [28, 29]. However, something else has been observed in SARS-CoV-2 but not SARS-CoV-1: SARS-CoV-2 has one of the hardest outer shells (low M disorder) among CoVs [28, 29]. This provides the initial evidence for the resilience of the virus, especially in body fluids, where it is exposed to harmful enzymes [30, 31]. As a result, heavy viral shedding occurs.

Computational, experimental, and clinical evidence points to a specific but peculiar evolution of SARS-CoV-2 that has important clinical and epidemiological implications

The shell disorder models do not stop there. A later search reveals that the hardness of viral shell is associated with burrowing animals, such as rabbits and pangolins that often come in contact with buried feces [32]. A more careful study of pangolin-CoV using shell disorder models, including the inner shell-virulence one, provides evidence that the 2017 strain is attenuated. Therefore, the shell disorder models do not only support the natural evolution of SARS-CoV-2, but also point to a peculiar but specific evolution that has important clinical and epidemiological implications including immunity, virulence, and infectivity. The evidence put forth thus far was computational and empirically based and involved the use of artificial intelligence (AI). In this paper, however, we will also point to published experimental and clinical results that support important aspects of the shell disorder models pertaining to the SARS-CoV-2, extending even to the finer details of the models' predictions. The results point to a specific but peculiar natural evolution of SARS-CoV-2. The WHO mission report [10] underscores the importance of studies such as this in the attempt to

uncover the origin of the virus.

The Viral Shell Disorder Models

Protein intrinsic disorder and AI tools

An important concept that was used in the mentioned shell disorder models is protein intrinsic disorder, which can be defined as entire proteins or portions of a protein that have no unique 3D structure [33-36]. It has long been known that disordered proteins have important functions [33-41]. Consequently, many tools used to predict disorder have been built. One of these is PONDR® VLXT (<https://www.pondr.com>), which is known to be among the most sensitive in the detection of protein-protein/RNA/DNA/sugar/ lipid interactions [42-46], and has been successful in the study of structural proteins of a large number of viruses, including influenza A virus, rabies virus, HIV, smallpox virus, herpes simplex virus (HSV), hepatitis C virus (HCV), and yellow fever virus (YFV) [15-17, 32, 47, 48]. PONDR® VLXT is a neural network that is trained using sequences of known ordered and disordered protein [49-51]. The predictor is fed with the sequence of a protein and the output is the prediction of order or disorder predisposition for each residue. An important measure that is frequently used in disorder status studies is the PID (Percentage of Intrinsic Disorder), which is defined as the number of residues predicted to be disordered, divided by the total number of residues, and multiplied by 100. The PID provides a gauge for the level of disorder in a protein chain.

The sequences are available at UniProt (<https://www.uniprot.org>) and GenBank-NCBI (<https://www.ncbi.nlm.nih.gov/protein>). The sequence were entered into PONDR® VLXT (<https://www.pondr.com>). Both sequences and PONDR® VLXT results were automatically added to a MySQL server using a JAVA program [16]. Measurements of sequence homology were obtained using BLASTP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). Phylogenetic trees were constructed using EMBI-EBI Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Figures

were drawn using GIMP (<https://www.gimp.org/>) and OpenOffice Draw (<https://www.openoffice.org/download/>) . Calculations necessary for correlation, regression, and multivariate analyses were conducted using R statistical package [52].

The viral shell disorder models: three closely related projects

There are three closely related and highly interlinked shell disorder-based models. One involves the parent project, i.e. HIV vaccine mystery/Viral shapeshifter. The others are spinoffs (CoV transmission shell disorder model and the virulence-inner shell disorder model). What is common among them is the study of viral protein shell disorder. Each model has related but different implications.

Parent project: the viral shapeshifters and the HIV vaccine mystery

In 2005, while compiling a database of viral shell proteins, something unusual was observed. The outer shell protein matrix of many HIV-1 strains was found to be abnormally disordered, which had not been seen in any of the other viruses before then. The results was first published in Virology Journal in 2008 [16, 17], and this unusual characteristic can be traced to the absence of an effective HIV vaccine despite a search that spanned approximately 40 years [15, 18, 19]. As more data became available, two other viruses, HSV and HCV, were added to the list of viruses with highly disordered outer shells, while virtually all other viruses (and especially those with effective vaccines available) have low or very low disorder levels of their outer shells.

Figure 1 illustrates the basic virion physiology with HIV and CoV as examples. We are able to see the differences and similarities between them in terms of shell proteins. HIV has three shells, the matrix, capsid, and the nucleocapsid, whereas CoV has two shells, the membrane (M) and the nucleocapsid (N). Similar virion physiology can be found in other viruses, even if they have a single shell layer or multiple shell layers, as in the cases of poliovirus and smallpox (variola) virus,

respectively [15, 53].

As a shell disorder database was being built, a common characteristic of viruses that emerges is the tendency of their outer shell to be more highly ordered, i.e. have a low PID [16, 17]. This can be attributed to the need of the virus to protect itself from environmental damage, as the outermost shell plays the greatest role in protecting the virion (as seen in **Figure 2C**). When the shells of HIV-1 were inspected, something abnormal was observed. Many of HIV-1 strains have high disorder in the matrix (outer shell), not seen in virtually all viruses [15-17]. As the database of viral shell disorder grew to encompass a much larger variety of viruses, two other viruses were seen as having similar characteristics as HIV [18, 20], namely HCV and HSV. More interestingly, HCV and HSV have strong association with sexual transmission, and effective vaccines have never been found for these viruses as well. These are also characteristics of HIV. On the other hand, there is a complete absence of high disorder levels at the outer shells among the classical viruses such as YFV, poliovirus, and smallpox virus, for which effective vaccines have been found (as seen in **Figure 2B**).

The mechanism of the “viral shapeshifting” immune evasion can be seen in **Figure 2A**. The highly disordered matrix (outer shell) allows for motions that increase the motions of the viral surface glycoprotein. The high level of motions at the surface therefore prevents strong binding of antibodies to the virus. The association of a soft disordered outer shell with sexual transmission is related to the fact that sexual transmission does not require the virus to remain in harsh non-physiological environment for a long time, and therefore a more protective outer shell is usually not necessary. As seen in **Figure 2**, the outer shell (M) PID looks nothing like those of HIV, HSV, or HCV, for which no effective vaccine has ever been found. We first made the accurate prediction of the feasibility of a COVID-19 vaccine in a paper published in Oct 2020, months before the first COVID-19 vaccine was approved by the US Food and Drug Administration [20, 21, 54].

CoV Shell Disorder-Transmission model

A spinoff from the above mentioned parent project came in 2011 [26], which took place before the 2012 MERS-CoV outbreak. The N and M CoV PIDs were collected as an extension of our shell disorder database. N and M proteins are the inner and outer proteins of CoVs, as seen in **Figure 1B**. Since human coronaviruses (HCoVs) were previously only associated with mild colds, prior to the 2003 SARS outbreak medical research of CoVs was sorely neglected, but knowledge of animal CoVs was plentiful due to their constant threat to the farming industry [15, 27, 28, 55]. We applied knowledge of animal CoVs, particularly those of porcine CoVs, to our database of CoV shell disorder and discovered that CoVs clustered easily into three groups based mainly on the N PID [26-29]. As shown in **Table 1**, Group A is made up of CoVs with lower fecal-oral but higher respiratory transmissions, group B consists of CoVs with intermediate fecal-oral and respiratory transmission potentials, whereas group C includes viruses Group C with lower levels of respiratory transmission and higher levels of fecal-oral transmission.

When the model was first built, it was determined that based on its N PID (50%), the 2003 SARS-CoV-1 fell into group B, which includes viruses that have intermediate fecal-oral and respiratory transmission potentials. The model was first published in 2012 [26]. When the MERS-CoV outbreak came shortly after the publication, it became an opportunity to validate the model. MERS-CoV was assigned to group C given an N PID of 43% [27]. The fact that MERS-CoV is categorized as having higher fecal but lower respiratory potentials is consistent with what we now know about the virus [10]. The virus is not easily spread among humans but spreads easily among camels, which are farmed in the Middle East [56].

Yet another chance to validate the model came with the COVID-19 outbreak. This time SARS-CoV-2 had to be placed in group B, alongside SARS-CoV-1. There is, however, something irregular about this virus. When the shell disorder analysis was first performed on the SARS-CoV-2, the virus was predicted to have one of the hardest outer shell (i.e. lowest M PID) within its CoV family. Since the outer shells of viruses play a vital role in protecting the virion from damage, suspicion was immediately cast on this factor as the culprit for the greater contagiousness of SARS-CoV-2. A

harder shell basically means that the virus will be more resistant to the anti-microbial enzymes found in body fluids [28, 30, 31]. A search for CoVs with similarly hard outer shells revealed that this hard shell is associated with CoVs of burrowing animals, such as pangolins and rabbits, even if the CoVs are not closely related to each other or to SARS-CoV-2, as in the case of rabbit-CoV [29]. We will present both clinical and experimental evidence that SARS-CoV-2 has a hard shell and is shed in large quantities through the nose, nasal cavity, throat, and mouth.

Virulence-inner shell disorder model

In 2014, yet another spinoff of the model was initiated. In this case, correlations between inner shell disorder and virulence were detected in a variety of viruses including flaviviruses, filoviruses, and the Nipah virus [22-25]. **Figure 3** shows some example of the correlations. **Figure 3a** represents the correlation between EBOV virulence and inner shell, NP, whereas **Figure 3b** illustrates the same correlation for DENV CFRs and inner shell (C) disorder. A strong correlation has been found between EBOV virulence (CFR) and inner shell (NP) disorder with a correlation coefficient (r) of 0.92 ($p < 0.001$) [24]. The DENV yielded a strong correlation of 0.9 ($p < 0.001$) [22]. It is interesting to note that only a modest correlation ($r = 0.6$, $p < 0.05$) can be obtained when the data encompassed a large variety of flaviviruses (e.g. DENV, YFV, ZIKV) but the correlation jumps to 0.9 ($p < 0.001$) when the outer shell (membrane, M) disorder is included [23].

It would, however, be a mistake to assume that the outer membrane always has a correlation with virulence, as no correlation could be detected between DENV virulence and M disorder.

Furthermore, EBOV membrane protein (vp40) shows a negative relationship ($r = -0.9$, $p < 0.05$) between virulence and outer shell disorder (Goh 2015). The confusing trend is the result of the intricately complex relationships between the inner and outer shells, which are rather different in different viruses. In the case of EBOV, matrix and nucleocapsid disorder has an inverse relationship because the shells play compensating roles in protecting the virion. Therefore, when the inner shell is more disordered, the outer shell compensates for the shortcoming by being more ordered, and

vice-versa. In the case of flaviviruses, the outer shell disorder often assists in the virulence by providing for better penetration into vital organs via greater efficiency in protein-protein/DNA/RNA/lipid interaction, as in the case of the highly pathogenic YFV. In the case of ZIKV, in which it has relatively high M disorder but low inner shell (C, capsid) disorder, the M disorder is manifested as greater fetal morbidity, since this allows for greater penetration of the placenta [20]. SARS-CoV-1/2 do not penetrate the placenta easily as revealed thus far by clinical studies [57, 58], which is consistent with the fact that SARS-CoV1/2 have relatively hard outer shells (low M PIDs).

“Trojan horse” immune evasion

The reason for the correlations between inner shell disorder and virulence has to do with an immune evasion strategy that is described as “Trojan horse” [15, 20, 22-25]. This involves a strategy where the virus attempts to replicate rapidly before the host immune system notices the presence of the virus. In the process of doing so, however, the resulting high viral loads in vital organs often overwhelm the host, thereby killing it. It therefore often backfires on the host [15]. Since the inner shell proteins play important roles in viral replication, intrinsic disorder enhances the efficiency of these proteins by allowing protein-protein/RNA/DNA/lipid bindings [33, 34, 36, 39, 42-46, 59].

The inner shell proteins of viruses have similar, though not identical, functions in their roles in viral replication [53, 60-63]. They are commonly involved in packaging of viral particles before their release [53, 60]. The coronavirus N recruits RNA and other proteins to regions near the endoplasmic reticulum (ER) and Golgi apparatus, where they are assembled and packaged [63]. Similarly, the precursor of the inner shell, C (capsid) protein of the flavivirus migrates to the ER [53]. Upon doing so, it then becomes embedded into the ER membrane, where it binds to viral mRNA and other viral proteins towards assembly. In the case of ebolavirus, its inner shell protein (NP, nucleoprotein) assists in building a tube-like structure, which is involved in the transportation of proteins that are assembled and budded as viral particles. As for the Nipah virus, measles virus,

and other related viruses, the N (Nucleocapsid) binds to the P and L proteins to become the RNA polymerase, which is crucial for the replication of the RNA [64]. All these require protein-protein/RNA/lipid interactions. Greater protein intrinsic disorder enhances the efficiency of such interactions by providing better molecular fitting.

Disorder models: a specific evolutionary nature of SARS-CoV-2

Hard shell (Low M PID) and burrowing animals

We have already mentioned that SARS-CoV-2 differs from most of the other coronaviruses by having the hardest shell. **Figure 4** shows that SARS-CoV-2 has one of the hardest outer shells (i.e. a lowest M PID) among a representative selection of CoVs. CoVs with similarly hard shells can only be found among CoVs of burrowing animals, such as pangolins and rabbits. Interestingly, rabbit-CoV is not closely related to SARS-CoV-2. It is therefore likely that the hard outer shells arose evolutionarily from the fact that burrowing animals are likelier to be in contact with fecal materials that have been buried for months, or even years. Interestingly, pangolins also have feeding habits that may enhance the chances of ingesting buried fecal materials. Pangolins have strong arms that allow them to dig for subterranean ants and termites. While their sticky tongues allow termites and ants to be trapped as food before swallowing, accidental ingestion of feces and soil is inevitable. The fecal-oral route is not only supported by the M PID data as in **Figure 4** and **Table 2**, but is also supported by the CoV-transmission model, as seen in **Figure 5**.

Computational evidence suggests that SARS-CoV-2 entered humans via pangolins in 2017 or earlier as an attenuated strain

Table 2 and **Figure 5** show that there are not just differences in the N PIDs of the various pangolin-CoVs, but stepwise differences as the collection dates become farther from 2019, which suggests an interesting but peculiar evolution. The N PID of one of the 2017 pangolin-CoV (marked 'XX' in

Figure 5) is the lowest when compared to other pangolin-CoV samples. When we apply disorder analysis using CoV-transmission models, we discover that most of the pangolin-CoVs samples falls into C, the group that contains CoVs with higher fecal-oral transmission but lower respiratory potentials. When we apply the virulence-inner shell disorder model, however, it reveals that most pangolin-CoVs, especially the 2017 one, are likely attenuated versions of SARS-CoV-2. An exception is the 2019 pangolin-CoV, which has similar N PID to SARS-CoV-2. Apparently, the virus could have been attenuated by the natural fecal-oral behaviors of pangolins.

Figure 5 shows that there is a relationship between virulence and N PID in SARS-CoV-1/2. The broad SARS-CoV-2 CFR range given is merely a reflection of estimates commonly found in literatures, and is based on the current number of cases and death [65, 66]. More importantly, there is a consensus within the scientific community that SARS-CoV-1 was generally more lethal than SARS-CoV-2 [66, 67].

The phylogenetic tree of the M protein provides evidence for a closer relationship between SARS-CoV-2 and pangolins-CoV

Phylogenetic studies have been done to examine the relationship between pangolin-CoV and SARS-CoV-2. Most of the studies available are based on the virus' genome-wide analysis. In fact, at least one study discounted the likelihood of pangolin-CoV, particularly the 2019 one, as being a direct ancestor of the SARS-CoV-2 [7]. Because it is genome-wide, the chances of recombinations occurring in at least one of the regions increases, which could lead to a mistake in the phylogenetic analysis [68, 69]. One solution is to constrain the phylogenetic analysis to a highly conserved proteins, such as M. The M protein is an ideal target, as seen in **Table 2**, which shows that pangolin-CoV M has a homology of 97-98% to SARS-CoV-2, compared to approximately 91% for genome-wide sequence similarity [9]. Indeed, we will see that the result for M phylogenetic analysis is different from what we have seen in previous studies.

Because the M and N proteins are crucial to our research, further investigation revealed a specific pattern of evolution of pangolin-CoVs. **Figure 6** shows the phylogenetic trees of N and M proteins. While much of the two phylogenetic trees resembles each other, important differences should be noticed as shown in the regions shaded in green. Previous phylogenetic studies, especially those using the entire CoV genome, paints similar relationships among SARS-CoV-2, pangolin-CoV, and bat-RaTG13 as seen in **Figure 6a**, which uses the N protein. The phylogenetic tree based on M shows, however, that RaTG13 is not closer to SARS-CoV-2 than Pangolin-CoVs are to SARS-CoV-2, despite the greater sequence similarity between RaTG13 and SARS-CoV-2. This is also in sharp contrast to the phylogenetic studies based on the entire CoV genome or proteins other than M where RaTG13 is depicted as having the closest relationship to SARS-CoV-2.

There are important reasons for this discrepancy. The fact that CoVs closely related to SARS-CoV-2 all have the hardest outer shell (lowest M PID) in the family adds to the reason that the M is likely to be very conserved among the related CoVs. Because the M proteins are highly structured and conserved, only mutations, not recombinations, are likely to have occurred, and this provides for a more accurate phylogenetic snapshot of the ancestral tree of CoVs closely related to SARS-CoV-2. Further support of this can be found in **Table 2**. The sequence similarity of M proteins of pangolins-CoVs is in the range of 97-98.6%, whereas RaTG13 has a sequence homology of 99%, which is in contrast to approximately 90% and 96% homologies for pangolin-CoV and RaTG13 genome to SARS-CoV-2, respectively. In summary, the chances of pangolin-CoV being the intermediary increase dramatically using M for our phylogenetic analysis. It should also be noted that recombinations are notorious for causing confusion in previous studies, as most phylogenetic algorithms are not designed to handle them. This was therefore likely the case in all previous pangolin-CoV studies, and the choice of M in our phylogenetic study sidesteps this hurdle.

The specific evolution of SARS-CoV-2 via pangolins

Figure 7 illustrates the specific evolution of SARS-CoV-2. More importantly, it points to an evolution with implications that have thus far been seen as unique. Because of pangolin association with fecal-oral transmission, the ancestral strains of SARS-CoV-2 were likely attenuated and had greater fecal-oral transmission potentials. There is evidence that other viruses could have been attenuated by animals that have fecal-oral transmission behaviors.

One example is the case of NiV (see **Figure 7**). It was first discovered in Malaysia when pigs ate virus-laden fruits that had fallen on the ground after being consumed by bats. Farm workers were then infected by the pigs (which are farm animals that are bred in close contact with each other, and therefore fecal-oral transmission is an important factor) [70]. The case-fatality rate was about 38%, which is in sharp contrast to the outbreaks in Bangladesh-India after 2001 that have CFRs of 70-80% [71]. Disorder analysis revealed differences in inner shell disorder (N PID) 41-42% vs 43-44%.

No such attenuation could however be seen in the case of the 2003 SARS-CoV-2 and civet cat. A search for signs of attenuation among civet-CoV came out empty. This necessarily leads to the scenario seen in **Figure 7** as denoted by “SARS1”. It is therefore likely that the virus entered the human population as a virulent strain that alerted the medical community almost immediately, unlike “SAR2”, i.e. SARS-CoV-2.

A curious but baffling trend that is worth mentioning is the stepwise increase in pangolin-CoV N PID for each year (**Figure 5**). This intriguing pattern is indeed very difficult to account for. The only plausible explanation is the possibility of cross-species transmission from humans back to pangolins for each subsequent year, since the first introduction of pangolin-CoV to humans is illustrated in **Figure 6**, where arrows point to the possibilities of viruses moving back to pangolins and bats. It is, however, not difficult to imagine the common situation when humans dispose of feces together with sweet and fruity foodstuff that attracts ants and other insects, and would eventually be scavenged by pangolins and bats.

It is also plausible that the stepwise N PIDs are actually capturing a snapshot of the evolutionary

gain of function for the human versions. There was an obvious evolutionary pressure towards greater respiratory transmission potentials among its human variants that forced the virus towards becoming more virulent, as respiratory transmission requires an adequate viral load at the saliva and mucus for it to be viable and, likewise, an overwhelming viral load at vital organs can cause death [15, 20, 22-25]. The involved mechanisms of action have already been discussed above.

Experimental and clinical evidence

Strong experimental evidence of SARS-CoV-2 resilience is consistent with the shell disorder model

We have seen computational evidence of the hard outer shell (low M DID) of SARS-CoV-2 arising from being evolutionarily intertwined with a burrowing animal, which raises the question: is there experimental evidence for this? As it turns out, the answer is “yes”.

Both SARS-CoV-1 and SARS-CoV-2 were tested on hard surfaces like plastic and steel, but the investigators, van Doremalen *et al.* [72], were unable to find any significant differences in stability of the two viruses, as infectious particles could be collected up to 72 hours. They were, however, able to detect hints of stronger resilience for SARS-CoV-2 on cardboard, as this virus remained viable for 24 hours, in contrast to SARS-CoV-1, which lasted only 8 hours. Van Doremalen *et al.* were unfortunately unable to reaffirm their result, owing to high variability [72].

SARS-CoV-2 resilience is many times that of other CoVs: experimental data

Riddell *et al.* [73] conducted a similar experiment but, this time, in the dark, devoid of harmful UV light. They also took note of the temperature and humidity. The group revealed that at 20°C and 50% RH (relative humidity), SARS-CoV-2 remained viable even after 28 days. This is in sharp contrast to MHV (murine hepatitis virus) and TGEV (transmissible gastroenteritis virus), which

were found by another group, Casanova *et al.* [74] to last only for 3 and 5 days under the same conditions, respectively.

The stark results are astonishingly consistent with the CoV shell disorder model with respect to the hardness of the outer shell (M PID). The M PIDs of SARS-CoV-2, MHV, and TGEV are 6%, 8%, and 14% just as the viruses lasted >28, 5, and 3 days, respectively under, 20°C and 50% RH. The correlation concurs with the basic shell disorder tenet that stipulates that shells, especially the outer one, protect the virion from damage. While SARS-CoV-1 was not studied by Rill *et al.* or Casanova *et al.*, the M PID of SARS-CoV-1 (8.6%) is very close to that of MHV (8%) [26-29] and the necessary interpolation and inference can be made.

A hard outer shell protects the virus from the environment and anti-microbial enzymes found in body fluid

The fact that greater evidence of resilience arose when the experiments were conducted away from any light is likely a reflection of the nature of the environmental protection that the hard outer shell confers. This complements the fact that ancestral coronaviruses were likely to have been buried in the soil covered by fecal materials away from UV light for a long time while awaiting contact with the next pangolin. Furthermore, it also highlights the environment that the virus is exposed to in the nasal regions, throat, and mouth, physiologically devoid of UV light but exposed to anti-microbial enzymes in body fluids. The results therefore illuminate the ability of the virus to resist the antimicrobial enzymes, thus giving rise to heavy viral shedding that provides for greater contagiousness.

Clinical evidence of heavy shedding provides clues to transmissibility and viral resilience in body fluids

While the evidence of viral resilience implies that the virus is resistant to the antimicrobial enzymes

found in body fluids and thereby suggests greater viral shedding and transmissivity, the question is then: is there any clinical evidence of such? The answer again is: “yes”. Wölfel *et al.* performed a study of patients and found that COVID-19 patients typically shed viral particles in volumes of 1,000 times those of SARS patients [75]. The shedding begins when the patients start showing the mildest symptoms, and continues even after the symptoms are over.

Strong experimental evidence supporting the disorder models of infection and virulence: SARS-CoV-1 vs. SARS-CoV-2

A group of Eric J. Snijder in the Netherlands [76] infected separate sets of Vero E6 cells with SARS-CoV-1 and SARS-CoV-2. The presence of infectious particles was detected and quantified using antibodies and electron microscopy. The number of infectious particles of SARS-CoV-1 was found to be greater than that of SARS-CoV-2 by a factor of approximately 50. The levels of intracellular RNA were also quantified using polymerase chain reaction (PCR). Conversely, a higher level of RNA was observed for SARS-CoV-2.

The experiment reproduced the predictions of the shell disorder models

The results are astonishing, since Orgando *et al.* [76] have basically reproduced the predictions made by the shell disorder models. The models stipulate that SARS-CoV-1 (CFR = 2%) is more virulent than SARS-CoV-2 (CFR = 10%) because SARS-COV-1 (N PID = 50%) 1 has a higher N disorder than SARS-CoV-2 (N PID = 48%). The reason for the greater virulence is the higher viral load at vital organs and the greater ability of the virus to replicate more rapidly via a disordered inner shell before the host immune system is able to detect its presence. This begs the question: if SARS-CoV-1 replicates faster than SARS-CoV-2, why isn't the virus shedding in greater quantities as in the case of SARS-CoV-1, as we have seen above? The only possible explanation is that it is not as resistant to the anti-microbial enzymes found in the saliva and mucus as highlighted by the

shell disorder model.

Clinical and experimental validation of the shell disorder models: viral load in the viral organ (virulence) vs. body fluids (respiratory transmission)

While the experiment [76] shows that SARS-CoV-1 is more aggressive than SARS-CoV-2 due to producing a greater number of virus copies, Wölfel *et al.* [75] have shown that SARS-CoV-2 is more contagious than SARS-CoV-1 by shedding much more infectious particles. A superficial analysis of this matter would lead us to believe that the experimental and clinical evidence is contradictory. A closer look, however, reveals that the only plausible way that the different sets of data are actually consistent with each other is the scenario described by the CoV transmission-shell disorder model, which tells us that the reason that SARS-CoV-2 is more contagious is attributed to its hard outer shell that resists the antimicrobial actions of the enzymes found in the saliva and mucus [30, 31].

There are also hints in the results of Wölfel *et al.* that this is the case. The viral loads in the saliva and mucus of COVID-19 patients could reach 1000 times that of SARS patients, whereas the viral loads from lung samples are similar between the different sets of patients [75]. While according to the shell disorder models, SARS-CoV-1 patients should have higher viral loads in the lungs than those of SARS-CoV-2, the clinical results pertaining to the lung samples could reflect the differences in the life-cycles of SARS and COVID-19, which could affect the timing of the release of viral particles. Nevertheless, the difference in the viral loads in body fluids of the two types of patients is startling and very telling.

COVID-19 vaccine success: yet another validation of the shell disorder models pertaining to SARS-CoV-2

While the world sighs with relief at the arrival of effective COVID-19 vaccines, we must not forget

that the search for effective HIV, HCV, and HSV vaccines has taken approximately 40, 30, and 100 years, and has thus far been met with complete failure [15, 18, 19, 77]. The world is gripped by the nightmare scenario that this could happen to the COVID-19 vaccine search. From the beginning, the shell disorder models have plenty to say about this, and if any were proven wrong, their main tenet would have been totally invalidated. **Figure 2b** shows that the outer shell of SARS-CoV-2 does not resemble those of HIV, HSV, and HCV in terms of outer shell disorder. Instead it resembles the classical viruses, for which effective vaccines have been found. In fact, given the samples of coronaviruses we have examined, none of their outer shell disorder resembles that of HIV. This is consistent with what we know about CoVs. CoVs have strong fecal-oral transmission potentials, in contrast to HIV's, HCV's, and HSV's association with sexual transmission [15, 18, 19, 32]. The rapidity of the discovery of COVID-19 should also not be surprising, given its hard (low M PID) shell among CoVs and other viruses, as can be seen in **Figures 2, 3C, and 4A**. We actually foresaw the successful development of COVID-19 vaccines months before their approvals, as seen in our previous publications.

Addressing SARS-CoV-2's mysterious adaptation to humans

S protein and furin cleavage sites: Do the S protein enigmas contradict the shell disorder theories? No!

While shell disorder models are based mainly on the N and M proteins that offer a set of evolutionary implications, most other studies have been centered on the S (spike) protein. The SARS-CoV-2 S protein facilitates viral entry by attaching to the ACE-2 receptor. As mentioned above, the SARS-CoV-2 S protein binds to ACE-2 with an affinity that is 20-30 times that of SARS-CoV-1 in a computational study [11, 12]. SARS-CoV-2 has also been found to have the greatest affinity to human ACE-2 by another computational study [13].

Furthermore, SARS-CoV-2 S protein has a unique furin cleavage site that is seldom seen among

betacoronaviruses, including its close relatives, SARS-CoV-1, pangolin-CoVs, and bat-RATG13 [2, 11][14]. The furin site plays a role in the cleavage of the S protein into S1 and S2, which is an important process prior to viral entry. The SARS-CoV-2 furin cleavage site is somewhat unique by being polybasic, which allows for more efficient cleavage [11, 14]. Because of all these established characteristics of the S protein, it has become a somewhat common belief that the S protein is responsible for the greater transmissibility of SARS-CoV-2 [14]. If so, the question is then: is the S protein or the N/M proteins responsible for the greater contagiousness of SARS-CoV-2? Or is it both?

While the S protein may still contribute to the transmissibility, we will see in the next paragraph that published experimental and clinical data support greater contagiousness arising from order and disorder of M and N proteins, respectively. Furthermore, we will see that the shell disorder theories provide a pathway by which SARS-CoV-2 could evolve with humans for a long period of time, as suggested by many studies on the S protein.

Revisiting the experiment: definitive answers and some hints

The experiments [76] provided both definitive answers and some hints for the S protein enigma. Since it has been shown that the SARS-COV-2 has an S protein that is more adapted to the human ACE-2 receptors in many ways when compared to SARS-CoV-1, we could reasonably expect SARS-CoV-2 to produce more infectious particles if we claim that the S protein is responsible for its greater contagiousness. The experiment, however, clearly shows that this is not the case. The number of copies of infectious particles of SARS-CoV-1 exceeds those of SARS-CoV-2 by about 50 folds. Therefore, greater efficiency of the S protein in cleavage and binding to the ACE-2 does not lead to greater contagiousness with respect to the increase in viral loads, even in body fluids. Nevertheless, there are important signs of more efficient viral entry for SARS-CoV-2 seen in the results of the experiment reported by Ogando et al. [76]. Even though there were lesser SARS-CoV-

2 infectious particles, there were, conversely, also greater quantities of viral RNA copies in the case of SARS-CoV-2. The greater presence of viral RNA is likely an evidence of greater ease of viral entry. Such evidence is not only consistent with the shell disorder models, but also suggests how the models could explain the peculiarities of the data. The greater presence of RNA but lower presence of viral particles suggest that plenty of RNA has been replicated but is awaiting packaging and release. The virus is, however, unable to release more particles, as in the case of SARS-CoV-1, because its N protein is not sufficiently disordered. Apparently, the data points to the greater efficiency in the packaging, assembly, and budding of SARS-CoV-1. This is consistent with what is understood about the role of protein disorder in assembly, packaging, and viral budding as already described above. There is, however, a hint on how the excessive RNA could affect the transmissibility. Excessive RNA may mean that there are more RNA waiting in line within the cell in the queue to be assembled as part to the infectious particle. This could mean that there is a longer infectivity duration at the cellular level. There are actually clinical evidence for such in the case of SARS-CoV-2, in which some patients still shed infectious particles even months after recovery. Nevertheless, further research needs to be done.

S protein: Evidence of SARS-CoV-2 evolution with humans

We have seen compelling evidence that points to the high adaptation of the SARS-CoV-2 to human ACE-2. There is yet further evidence for this. For example, computational studies have found that the SARS-CoV-2 S protein is more adapted to the human ACE-2 than those of a variety of animals. We have also seen that the polybasic furin cleavage site is found only in SARS-CoV-2 and not even among its closest relatives including pangolin-CoVs, RATG13, and SARS-CoV-1. If that is the case, how did SARS-CoV-2 acquire its polybasic characteristic? An Indian group found a matching sequence in an unrelated human enzyme, eNac [78]. This sequence has the closest match to the human eNac, compared to those of many other animals.

If SARS-CoV-2 has been present in humans for a long time, how did it escape the notice of the medical community for all these years?

Assuming that all this plausible evidence that the virus has been in humans for a long period of time is correct, how did SARS-CoV-2 remain in humans for so long without the notice of the medical community? While it is possible that the virus had been in its virulent form within humans for a long time without being noticed by doctors, it is logically difficult to imagine this to be the case. This becomes a paradox for the S protein enigmas. This is where the shell disorder models come in as a complementary theory, not a competing one. The data presented by the shell disorder models imply that SARS-CoV-2 entered the human population via pangolins as an attenuated strain a few years ago, perhaps in 2017, or even earlier. This provides ample time for the virus to adapt to humans. In the past, attenuated viruses and vaccines have mutated into virulent strains. An example is the Sabin polio vaccine [15, 32].

SARS-CoV-1 vs SARS-CoV-2: replication cycle

Figure 8 summarizes the differences in replication cycle of the SARS-CoV-1 (SARS1) and SARS-CoV-2 (SARS2). Upon entry, both viruses undergo replication. However, the SARS-CoV-1 is more efficient in replicating and releasing the infectious particles, as shown in the experiment by Ogando *et al.* [76] and predicted by the shell disorder models. Because SARS-CoV-1 produces more infectious particles, especially at vital organs, it is more virulent than SARS-CoV-1. While the experiment reaffirms the presence of a greater number of infectious particles upon the initial infection of the cell, clinical evidence suggests that large quantities of SARS-CoV-2 are shed. This can only be possible given the information from Ogando *et al.* [76] if SARS-CoV-2 is more resistant to the antimicrobial enzymes found in the saliva and mucus because of its harder outer shell as detected by the shell disorder models. The large quantities of the virus found in the saliva and mucus contribute to the contagiousness of COVID-19.

There is also the S protein enigma that hypothesizes that the S protein contributes to the transmissibility of the virus. The results from the experiments by Ogando *et al.* [76] suggest that the S proteins contribute to the transmissibility in other ways. The presence of large amounts of viral RNA detected in SARS-CoV-2 implies the possibility that huge amounts of incomplete viral particles have been found. This could account for the clinical evidence that some people are still infectious months after recovery [79], and it is represented by a longer timeline for SARS2 in **Figure 8**. This phenomenon points to the likelihood that greater protein intrinsic disorder in SARS-CoV-1 leads to greater efficiency in the packaging and release of the particles, and not necessarily in RNA replication. This is consistent with the tenet that protein intrinsic disorder provides for more efficient protein-protein/RNA/lipid interactions, and that the N protein is certainly involved in the packing, assembly, and release of the virus.

The China conundrum and pangolin factor

A curious statistics that has not escaped the scientific community has to do with the low rate of COVID-19 infection in China, especially when we compare the number of cases in the USA and Europe. As of today, the total number of cases in China is close to 91,000, whereas the number of cases in the USA is above 32 million [65]. The numbers are simply astonishing, since the populations of China and USA are over one billion and 300 million, respectively. Nobody has yet been able to provide a satisfactory answer to this puzzle. A highly plausible answer yet to be suggested is that an attenuated strain has entered the population in China many years ago as a mild cold that provided immunity. If this is the case, it is then likely that a pangolin-CoV entered the human population many years before 2017. It will be, however, exceptionally difficult, if not impossible, to prove an attenuated strain had passed through China's population or otherwise, because studies have shown that many people who were infected have COVID-19 antibodies that begin to drop to undetectable levels after 6 months [80, 81]. Nevertheless, an infected patient may still be protected regardless of the presence of antibodies, as immune system cells such as the T-cells and memory B cells may provide long-term protection [53, 82,

83]. We need to keep in mind, however, that the idea that there is already mass immunity among people in China is currently a speculative but compelling observation that warrants further investigation.

Summary and conclusion

Computational, experimental, and clinical evidence for the specific evolution of SARS-CoV-2

We have seen that the shell disorder models elucidated a specific but peculiar evolutionary pathway. The hint of this pathway was first noticed when the shell disorder models found that the SARS-CoV-2 has among the hardest outer shell disorder among CoVs, and it was later discovered that this hard outer shell is associated with burrowing animals, particularly pangolins. This characteristic is validated by at least one experimental study. The hard outer-shell nature of SARS-CoV-2 has wide implications, including those that are of clinical, epidemiological, and immunological importance. The volumes of virus shed by COVID-19 patients is attributed by computational, clinical, and experimental studies to the hard protective outer shell that provides the virus' resistance to anti-microbial enzymes found in the saliva and mucus.

The shell disorder models not only accurately predicted the feasibility of COVID-19 vaccine development based on its M PID, but also the existence of an attenuated strain that had entered the human population from pangolins several years ago. The latter was done using N-M shell disorder analysis and phylogenetic analysis based on the M protein, which reveals a much closer relationship between SARS-CoV-2 and pangolin-CoV than previously thought. The same shell disorder analysis also predicts that SARS-CoV-1 is more virulent because of its higher N disorder that allows for more efficient replication of infectious particles, especially at vital organs, but has a lower viral load. This is not only reaffirmed by the experimental study of Ogando *et al.* [76] by showing larger quantities of infectious particles found in SARS-CoV-1, but the experimental results hint at the possibility that the shell disorder models are actually complementary to studies showing that S protein higher adaptation to

human ACE2 contributes to transmissibility. More specifically, the results showing larger amount of viral RNA, not infectious particles in cells infected by SARS-CoV-2, suggest that the S protein may be prolonging the infectious cycle. This is supported by the observation that many patients still shed infectious particles months after recovery.

Specific but natural evolution

Conspiracy theories will persist as the precise virus that first entered humans from an animal intermediary is likely to be extinct, especially if the virus had entered humans at least a few years ago. Given that it is likely to be extinct, we need to rely on genetic and proteomic analyses such as the shell disorder models. As mentioned, the shell disorder model has plenty to say about the evolutionary nature of SAR-CoV-2. The models explain that the likely reason that S protein is so adapted to the human ACE-2 is that it first enter humans as an attenuated strain a few years ago, not that it was engineered in a laboratory [32]. The silent spread, helped by its attenuated nature, escaped the notice of the medical communities since it could easily have been mistaken for a common cold. Furthermore, if the virus was engineered in a laboratory, how was anyone able to acquire the knowledge, which was previously unheard of, pertaining to the virus that we have just mentioned, including the contribution of its hard outer shell to its contagiousness?

There are also questions as to whether the pangolin-CoV samples were actually from pangolin-CoVs, since the samples were obtained from caged animals that were confiscated from smugglers, and the pangolins could have been infected by other caged animals [84]. For this, our data support the likelihood that the samples are those of pangolin-CoV, since evidence based on our CoV search and phylogenetic tree point to the fact that CoVs with the hardest outer shell are associated with unrelated burrowing animals such as rabbits and pangolins.

Implications are far-reaching

The scientific implications are aplenty and simply too many to be listed in this paper. As already mentioned, the immunological significance includes the ease of vaccine development and the existence of attenuated strains in nature. We have also seen how the outer shell affects the contagiousness of the virus. This should provide further hints on the control that would be of interests to epidemiologists. The envisaged way, if proven true, that the S protein affects transmissibility via prolonged infectious periods, could provide options for treatments. The intricate evolutionary pathway that ancestral strains of SARS-CoV-2 have to undergo provides us with insights and possible patterns on what to look out for, especially when we monitor for new potential pandemics and possible zoonotic or non-zoonotic COVID-19 reemergence after the current mass vaccination effort.

Author Contributions

GKMG conceived the idea, collected, and analyzed the data, and wrote the first draft. VNU helped with the collection and analysis of literature data, reviewed, and revised the draft. AKD and JAF reviewed the manuscript and provided the resources necessary for the research.

Conflicts of Interest

GKMG is an independent researcher and the owner of Goh's BioComputing, Singapore.

GKMG has also written a book ("Viral Shapeshifters: Strange Behaviors of HIV and Other Viruses") on a related subject. The authors have no other potential conflict of interests.

References

1. WHO. *Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19)* ([https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19))). 2019.
2. Hu, B., et al., *Characteristics of SARS-CoV-2 and COVID-19*. Nat Rev Microbiol, 2021. **19**(3): p. 141-154.
3. Li, X., et al., *Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2*. J Med Virol, 2020. **92**(6): p. 602-611.

4. Zhou, P., et al., *Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **588**(7836): p. E6.
5. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **579**(7798): p. 270-273.
6. Xiao, K., et al., *Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins*. Nature, 2020. **583**(7815): p. 286-289.
7. Liu, P., et al., *Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?* PLoS Pathog, 2020. **16**(5): p. e1008421.
8. Fan, H.H., et al., *Repurposing of clinically approved drugs for treatment of coronavirus disease 2019 in a 2019-novel coronavirus-related coronavirus model*. Chin Med J (Engl), 2020. **133**(9): p. 1051-1056.
9. Lam, T.T., et al., *Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins*. Nature, 2020. **583**(7815): p. 282-285.
10. WHO. *Middle Eastern respiratory syndrome coronavirus (MERS-CoV)* [https://www.who.int/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-\(mers-cov\)](https://www.who.int/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov)) 2015.
11. Andersen, K.G., et al., *The proximal origin of SARS-CoV-2*. Nat Med, 2020. **26**(4): p. 450-452.
12. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2*. Nature, 2020. **581**(7807): p. 221-224.
13. Paplani, S., et al., *In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus*. ArXiv, 2020. **2005**: p. 06199.
14. Hoffmann, M., H. Kleine-Weber, and S. Pohlmann, *A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells*. Mol Cell, 2020. **78**(4): p. 779-784 e5.
15. Goh, G.K., *Viral Shapeshifters: Strange Behaviors of HIV and Other Viruses*. 2017, Singapore: Simplicity Research Institute.
16. Goh, G.K., A.K. Dunker, and V.N. Uversky, *Protein intrinsic disorder toolbox for comparative analysis of viral proteins*. BMC Genomics, 2008. **9 Suppl 2**: p. S4.
17. Goh, G.K., A.K. Dunker, and V.N. Uversky, *A comparative analysis of viral matrix proteins using disorder predictors*. Virol J, 2008. **5**: p. 126.
18. Goh, G.K., et al., *HIV Vaccine Mystery and Viral Shell Disorder*. Biomolecules, 2019. **9**(5).
19. Goh, G.K. and V.N. Uversky, *Shell disorder and the HIV vaccine mystery: lessons from the legendary Oswald Avery*. J Biomol Struct Dyn, 2021: p. 1-10.
20. Goh, G.K., et al., *A Novel Strategy for the Development of Vaccines for SARS-CoV-2 (COVID-19) and Other Viruses Using AI and Viral Shell Disorder*. J Proteome Res, 2020. **19**(11): p. 4355-4363.
21. Goh, G.K., et al., *Feasibility of the vaccine development for SARS-CoV-2 and other viruses using the shell disorder analysis*. Pac Symp Biocomput, 2021. **26**: p. 143-153.
22. Goh, G.K., A.K. Dunker, and V.N. Uversky, *Correlating Flavivirus virulence and levels of intrinsic disorder in shell proteins: protective roles vs. immune evasion*. Mol Biosyst, 2016. **12**(6): p. 1881-91.
23. Goh, G.K., et al., *Zika and Flavivirus Shell Disorder: Virulence and Fetal Morbidity*. Biomolecules, 2019. **9**(11).
24. Goh, G.K., A.K. Dunker, and V.N. Uversky, *Detection of links between Ebola nucleocapsid and virulence using disorder analysis*. Mol Biosyst, 2015. **11**(8): p. 2337-44.
25. Goh, G.K., et al., *Nipah shell disorder, modes of infection, and virulence*. Microb Pathog, 2020. **141**: p. 103976.
26. Goh, G.K., A.K. Dunker, and V.N. Uversky, *Understanding Viral Transmission Behavior via Protein Intrinsic Disorder Prediction: Coronaviruses*. J Pathog, 2012. **2012**: p. 738590.
27. Goh, G.K., A.K. Dunker, and V. Uversky, *Prediction of Intrinsic Disorder in MERS-CoV/HCoV-EMC Supports a High Oral-Fecal Transmission*. PLoS Curr, 2013. **5**.
28. Goh, G.K., et al., *Rigidity of the Outer Shell Predicted by a Protein Intrinsic Disorder Model*

- Sheds Light on the COVID-19 (Wuhan-2019-nCoV) Infectivity*. Biomolecules, 2020. **10**(2).
29. Goh, G.K., et al., *Shell disorder analysis predicts greater resilience of the SARS-CoV-2 (COVID-19) outside the body and in body fluids*. Microb Pathog, 2020. **144**: p. 104177.
 30. Malamud, D., et al., *Antiviral activities in human saliva*. Adv Dent Res, 2011. **23**(1): p. 34-7.
 31. Cole, A.M., P. Dewan, and T. Ganz, *Innate antimicrobial activity of nasal secretions*. Infect Immun, 1999. **67**(7): p. 3267-75.
 32. Goh, G.K., et al., *Shell Disorder Analysis Suggests That Pangolins Offered a Window for a Silent Spread of an Attenuated SARS-CoV-2 Precursor among Humans*. J Proteome Res, 2020. **19**(11): p. 4543-4552.
 33. Dunker, A.K., et al., *Intrinsically disordered protein*. J Mol Graph Model, 2001. **19**(1): p. 26-59.
 34. Tompa, P., *Intrinsically unstructured proteins*. Trends Biochem Sci, 2002. **27**(10): p. 527-33.
 35. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-27.
 36. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm*. J Mol Biol, 1999. **293**(2): p. 321-31.
 37. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins*. Chem Rev, 2014. **114**(13): p. 6589-631.
 38. Uversky, V.N. and A.K. Dunker, *Understanding protein non-folding*. Biochim Biophys Acta, 2010. **1804**(6): p. 1231-64.
 39. Dunker, A.K., et al., *Function and structure of inherently disordered proteins*. Curr Opin Struct Biol, 2008. **18**(6): p. 756-64.
 40. Dunker, A.K., et al., *Flexible nets. The roles of intrinsic disorder in protein interaction networks*. FEBS J, 2005. **272**(20): p. 5129-48.
 41. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling*. J Mol Recognit, 2005. **18**(5): p. 343-84.
 42. Yan, J., et al., *Molecular recognition features (MoRFs) in three domains of life*. Mol Biosyst, 2016. **12**(3): p. 697-710.
 43. Cheng, Y., et al., *Mining alpha-helix-forming molecular recognition features with cross species sequence alignments*. Biochemistry, 2007. **46**(47): p. 13468-77.
 44. Vacic, V., et al., *Characterization of molecular recognition features, MoRFs, and their binding partners*. J Proteome Res, 2007. **6**(6): p. 2351-66.
 45. Mohan, A., et al., *Analysis of molecular recognition features (MoRFs)*. J Mol Biol, 2006. **362**(5): p. 1043-59.
 46. Oldfield, C.J., et al., *Coupled folding and binding with alpha-helix-forming molecular recognition elements*. Biochemistry, 2005. **44**(37): p. 12454-70.
 47. Xue, B., et al., *Viral disorder or disordered viruses: do viral proteins possess unique features?* Protein Pept Lett, 2010. **17**(8): p. 932-51.
 48. Goh, G.K., A.K. Dunker, and V.N. Uversky, *Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses*. Virol J, 2009. **6**: p. 69.
 49. Romero, P., et al., *Sequence complexity of disordered protein*. Proteins, 2001. **42**(1): p. 38-48.
 50. Garner, E., et al., *Predicting Binding Regions within Disordered Proteins*. Genome Inform Ser Workshop Genome Inform, 1999. **10**: p. 41-50.
 51. Li, X., et al., *Predicting Protein Disorder for N-, C-, and Internal Regions*. Genome Inform Ser Workshop Genome Inform, 1999. **10**: p. 30-40.
 52. Team, R.C. R: A language and environment for statistical computing. 2016 [cited 2021; Available from: <https://www.R-project.org/>].
 53. Acheson, N.H., *Fundamentals of Molecular Virology*. 2007, Hoboken, NJ, USA: Wiley.
 54. Skowronski, D.M. and G. De Serres, *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine*. N Engl J Med, 2021. **384**(16): p. 1576-1577.
 55. Lin, C.M., et al., *Antigenic relationships among porcine epidemic diarrhea virus and transmissible gastroenteritis virus strains*. J Virol, 2015. **89**(6): p. 3332-42.

56. Ferguson, N.M. and M.D. Van Kerkhove, *Identification of MERS-CoV in dromedary camels*. Lancet Infect Dis, 2014. **14**(2): p. 93-4.
57. Wadman, M., *COVID-19 unlikely to cause birth defects, but doctors await fall births*. Science, 2020. **369**(6504): p. 607.
58. Komine-Aizawa, S., K. Takada, and S. Hayakawa, *Placental barrier against COVID-19*. Placenta, 2020. **99**: p. 45-49.
59. Macossay-Castillo, M., et al., *The Balancing Act of Intrinsically Disordered Proteins: Enabling Functional Diversity while Minimizing Promiscuity*. J Mol Biol, 2019. **431**(8): p. 1650-1670.
60. McBride, R., M. van Zyl, and B.C. Fielding, *The coronavirus nucleocapsid is a multifunctional protein*. Viruses, 2014. **6**(8): p. 2991-3018.
61. Masters, P.S., *The molecular biology of coronaviruses*. Adv Virus Res, 2006. **66**: p. 193-292.
62. Watanabe, S., T. Noda, and Y. Kawaoka, *Functional mapping of the nucleoprotein of Ebola virus*. J Virol, 2006. **80**(8): p. 3743-51.
63. Mortola, E. and P. Roy, *Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system*. FEBS Lett, 2004. **576**(1-2): p. 174-8.
64. Longhi, S. and M. Oglesbee, *Structural disorder within the measles virus nucleoprotein and phosphoprotein*. Protein Pept Lett, 2010. **17**(8): p. 961-78.
65. Worldometer. *Pandemic COVID-19 coronavirus pandemic* (<https://www.worldometers.info/coronavirus/>). 2021.
66. Rajgor, D.D., et al., *The many estimates of the COVID-19 case fatality rate*. Lancet Infect Dis, 2020. **20**(7): p. 776-777.
67. Abraham, T., *Twenty-first century plague: The story of SARS*. 2004, Baltimore, Maryland, USA: The Johns Hopkins University Press.
68. Flores-Alanis, A., et al., *The receptor binding domain of SARS-CoV-2 spike protein is the result of an ancestral recombination between the bat-CoV RaTG13 and the pangolin-CoV MP789*. BMC Res Notes, 2020. **13**(1): p. 398.
69. Posada, D., *How does recombination affect phylogeny estimation?* Trends Ecol Evol, 2000. **15**(12): p. 489-490.
70. Lo Presti, A., et al., *Origin and evolution of Nipah virus*. J Med Virol, 2016. **88**(3): p. 380-8.
71. Mire, C.E., et al., *Pathogenic Differences between Nipah Virus Bangladesh and Malaysia Strains in Primates: Implications for Antibody Therapy*. Sci Rep, 2016. **6**: p. 30916.
72. van Doremalen, N., et al., *Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1*. N Engl J Med, 2020. **382**(16): p. 1564-1567.
73. Riddell, S., et al., *The effect of temperature on persistence of SARS-CoV-2 on common surfaces*. Virol J, 2020. **17**(1): p. 145.
74. Casanova, L.M., et al., *Effects of air temperature and relative humidity on coronavirus survival on surfaces*. Appl Environ Microbiol, 2010. **76**(9): p. 2712-7.
75. Wolfel, R., et al., *Virological assessment of hospitalized patients with COVID-2019*. Nature, 2020. **581**(7809): p. 465-469.
76. Ogando, N.S., et al., *SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology*. J Gen Virol, 2020. **101**(9): p. 925-940.
77. Uversky, V.N., *On the irrationality of rational design of an HIV vaccine in light of protein intrinsic disorder*. Arch Virol, 2021. **166**(5): p. 1283-1296.
78. Anand, P., et al., *SARS-CoV-2 strategically mimics proteolytic activation of human ENaC*. Elife, 2020. **9**.
79. Mahase, E., *Covid-19: Past infection provides 83% protection for five months but may not stop transmission, study finds*. BMJ, 2021. **372**: p. n124.
80. Lumley, S.F., et al., *Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers*. N Engl J Med, 2021. **384**(6): p. 533-540.
81. Seow, J., et al., *Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans*. Nat Microbiol, 2020. **5**(12): p. 1598-1607.

82. Karlsson, A.C., M. Humbert, and M. Buggert, *The known unknowns of T cell immunity to COVID-19*. Sci Immunol, 2020. **5**(53).
83. Schulien, I., et al., *Characterization of pre-existing and induced SARS-CoV-2-specific CD8(+) T cells*. Nat Med, 2021. **27**(1): p. 78-85.
84. Chan, Y.A. and S.H. Zhan, *Single source of pangolin CoVs with a near identical Spike RBD to SARS-CoV-2*. bioRxiv, 2020: p. 2020.07.07.184374.

Figure legends

Figure 1. Comparative virion physiology. A) HIV virion has three main shells: Matrix (outer), capsid (intermediate) and nucleocapsid (inner shell) B) Coronavirus (CoV). (Figures reproduced with the permission of Gerard KM Goh, 2017).

Figure 2. Parent shell disorder project: Viral shapeshifter. A) Schematic diagram that illustrates the mechanism viral “shapeshifting” immune evasion in HIV (Figure reproduced with the permission of Gerard KM Goh, 2017). B) Shell disorder (maximal PIDs) of viruses that no vaccines (HIV, HSV and HCV) C) Shell disorder of the classical viruses that have effective vaccines available. The mechanism involves the movement of the surface glycoprotein arising from motions allowed within the disordered matrix (outer-shell) The outer shell PID of HIV, HSV and HCV are noticeably higher than those of the classical viruses with known vaccines. SARS-CoV-2 is placed in (B) for comparative purposes. The M PID of SARS-CoV-2 resembles the outer shell PIDs of the classical viruses than those of HSV, HCV and HIV.

Figure 3. Virulence-inner shell disorder model: “Trojan horse” immune evasion A) Correlation between filovirus/EBOV virulence (CFR) and NP PID). (Multivariate analysis: $p < 0.0001$ $r = 0.8$) B) Relationships between virulence and inner shell disorder (DENV, SARS-CoV-1/2) (DENV $r = 0.95$, $p < 0.001$).

Figure 4. SARS-CoV-2 has one of the hardest outer shell (lowest M PID) among CoVs.

Figure 5. N PID of Pangolin-CoV, bat-RaTG13 and SARS-CoV-1/2. Stepwise increment of pangolin-CoV N PIDs by the year is seen. According to the virulence-shell disorder model the 2017 pangolin-CoV (marked “XX”) is the most attenuated of the pangolin samples found.

Figure 6. Phylogenetic tree based on M and N proteins A) M protein 2) N protein. The regions shaded in green highlight important difference in the M phylogenetic tree not seen in other phylogenetic tree., Pangolins-CoVs are seen to be closer to SARS-CoV-2 than bat-RaTG13 in (A).

Figure 7. Zoonotic pathways towards human infections of SARS-CoV-1/2 and NiV. Qualitative timeline indicates that SARS-CoV-2 ancestral strains took a longer time to evolve in pangolins and humans before mutating to its current virulent form.

Figure 8. Schematic differences in the replication cycles of SARS-CoV-1 (SARS1) and SARS-CoV-2 (SARS2) as seen the Shell Disorder models and published experimental and clinical data. One and two virus copies are shown next to the cells for SARS1 and SARS2 respectively to illustrate the more efficient cell entry of SAR2, which could induce the ease of entry of more than one viral particle. The additional number of infectious particles produced by SARS1 compared to SARS2 is just a qualitative illustration that more particles are produced by SARS1. The same goes for the number of particles left after exposure to antimicrobial enzymes in body fluids.

Tables

Table 1. Categorization of coronaviruses by mainly N PID to predict levels of respiratory and fecal-oral transmission potentials ($p < 0.001$, $r = 0.9$).

Coronavirus	M PID	UniProt(U) /Genbank(G) Accession Code (M Proteins) ^a	N PID	UniProt(U) Genbank(G) Accession Code (N) ^a	Group/Remarks
HCoV-229E	23	P15422	56	P15130	Group A Higher levels of respiratory transmission lower levels of fecal-oral transmission
IBV(Avian) ^c	10	P69606	56	Q8JMI6	
Bovine	7.8	P69704	53.1	Q8V432(U)	Group B Intermediate levels of respiratory and fecal-oral transmission
Rabbit	5.7	H9AA37(U)	52.2	H9AA59(U)	
PEDV (Porcine) ^c	8	P59771(U)	51.7	Q07499(U)	
Canine (Resp.)	7	A3E2F6(U)	50.5	A3E2F7(U)	
HCoV-OC43	7	Q4VID2(U)	51	P33469(U)	
SARS-CoV-1	8.6	P59596(U)	50.2	P59595(U)	
HCoV-NL63	11	Q6Q1R9(U)	49	Q6Q1R8(U)	
SARS-CoV-2	5.9	P0DTC5(U)	48.2	P0DTC9(U)	
Bats ^b	11.2 \pm 5.3	A3EXD6(U)	47.7 \pm 0.9	Q3LZX4(U)	
MHV(Murine) ^c	8	Q9JEB4(U)	46.8	P03416(U)	Group C Lower levels of respiratory transmission higher levels of fecal-oral transmission
Pangolin ^d	5.6 \pm 0.9	QIA428617(G)	46.6 \pm 1.6 ^e	QIA48630(G)	
MERS-CoV	9.1)	44.3	K0BVN3(U)	
TGEV(Porcine) ^c	14	K0BU37(U)	42.41	P04134(U)	
Canine(Ent.)	8	P09175(U)	40	Q04700(U)	
HCoV-HKU1	4.5	B8RIR2(U) Q14EA7(U)	37.4	Q0ZME3(U)	

^aUniProt(U):[<https://www.uniprot.org/>]; (G)GenBank-NCBI (). [<https://www.ncbi.nlm.nih.gov/protein/>]

^bSummary figures on bats. Further details on the bat samples can be found in Table 2. 4 out of 5 bat-CoVs are in group B. High standard deviations are seen for N/M PIDs as denoted by “ \pm ”

^cMHV(murine hepatitis virus), IBV (infectious bronchitis virus), PEDV (porcine epidemic diarrhoea virus), TGEV(Transmissible, gastroenteritis virus) .

^dDetails on the pangolin samples can be found in Table 2. Three out of 4 pangolins-CoVs are in group C. Standard deviation is denoted by “ \pm ”.

Table 2. Grouping of pangolin-CoVs and Bat-CoVs by mainly N PID with SARSCoV and SARS-CoV-2 as references.

Coronavirus	Sequence Similarity M (%)	M PID (%)	Accession: UniProt(U) GenBank(G)	Sequence Similarity N (%)	N PID (%)	Accession UniProt(U) GenBank(G)	Group
SARS-CoV-2	100	5.9	P0DTC5(U)	100	48.2	P0DTC9(U)	B
SARS-CoV-1	90.5	8.6	P59596(U)	90.5	50.2	P59595(U)	B
Civet-SARS-CoV	90.1	8.6	QZ3TE9(U)	90.01	49.1	QBZTE4(U)	B
Pangolin-CoV		5.6±0.9 ^a			46.6±1.6 ^a		
2019	98.2	6.3	QIG55948(G)	98	48.7	QIG55953(G)	B
2018	97.7	4.5	QIQ54051(G)	93.8	46.3	QIQ54056(G)	C
2017***	98.2	5.9	QIA48617(G)	94	44.9	QIA48630(G)	C
				93.32	46.5	QIA48656(G)	C
Bat-CoV		11.2±15 ^a	Q9JEB4		47.7±0.9 ^a		
RATG13	99.6	4.1	QHR63303(G)	99.1	48.5	QHR63308(G)	B
Bat 512	35.5	15.3	Q0Q463(U)	29.4	46.5	Q0Q462(U)	C
HKU3	91	7.7	Q3LZX9(U)	89.6	48	Q3LZX4(U)	B
HKU4	42.7	16.4	A3EXA0(U)	51.1	48.5	A3EXA1	B
HKU5	44.7	11.8	A3EXD6(U)	47.9	47.1	A3EXD7(U)	B

^aStandard deviation is denoted by “ \pm ”.

*** Possible vaccine strain for SARS-CoV-2 detected

Figure 1

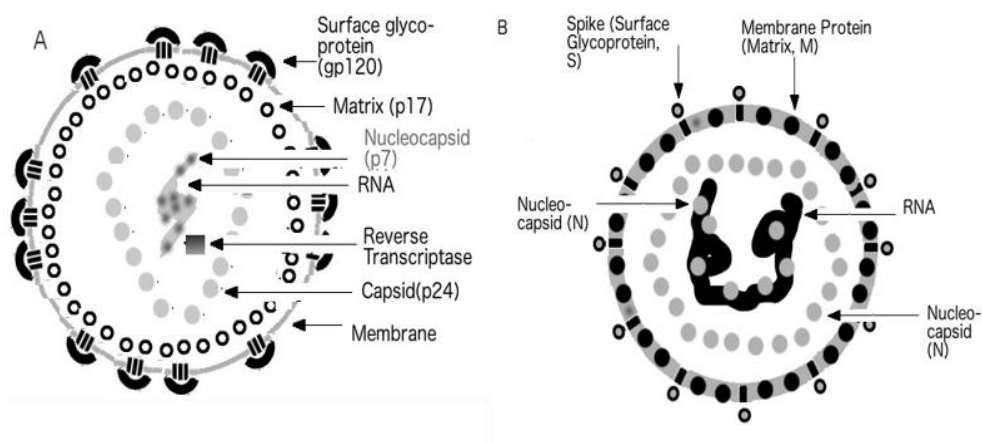
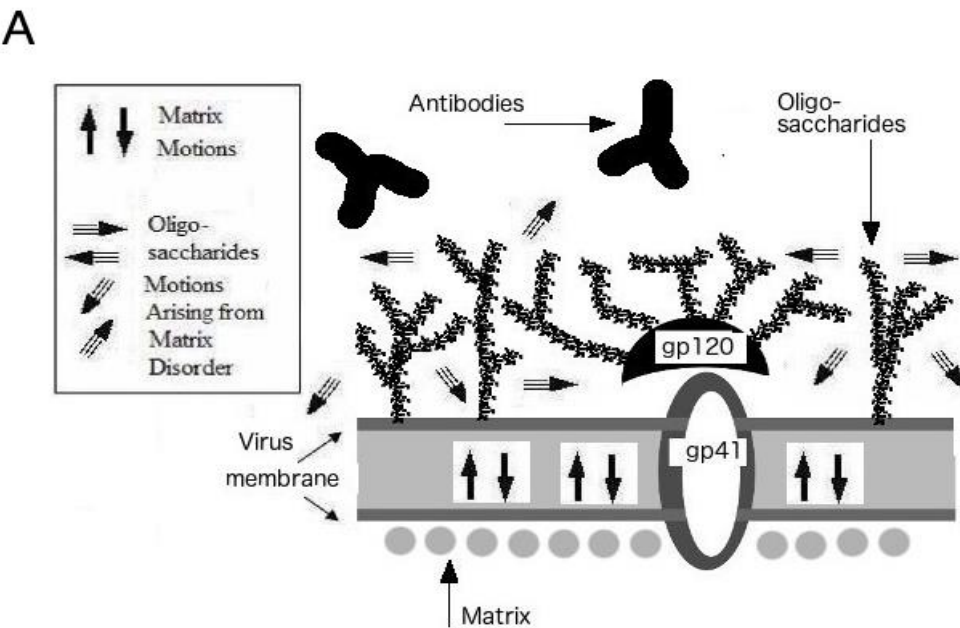


Figure 2



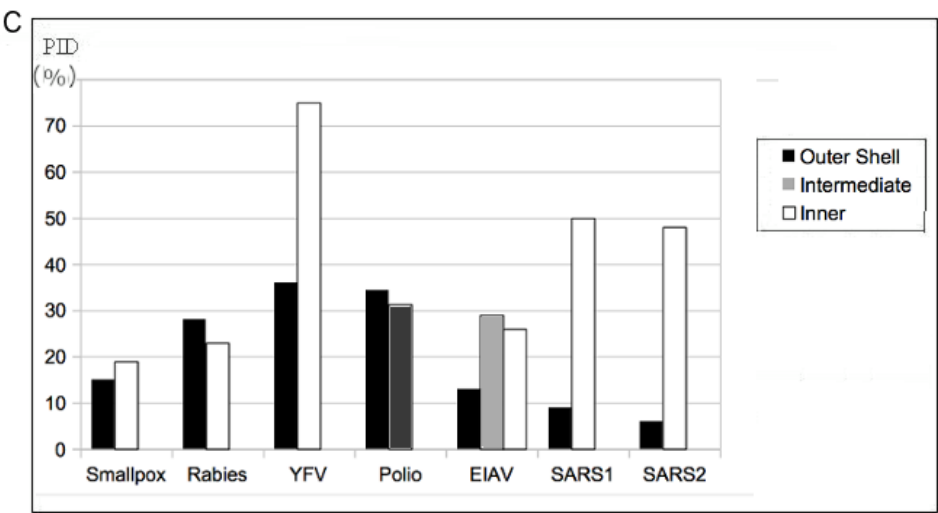
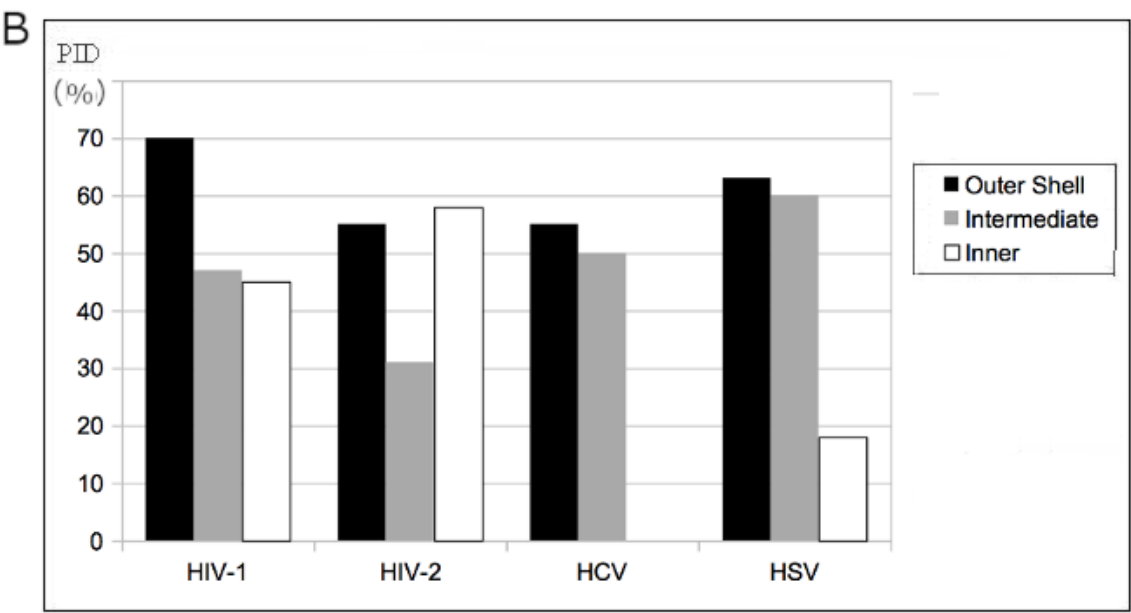
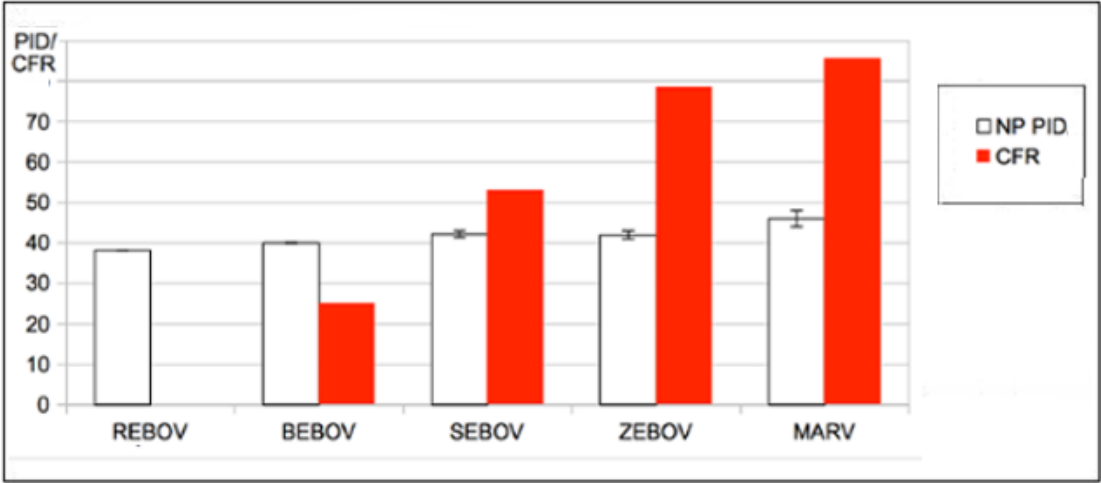


Figure 3

A



B

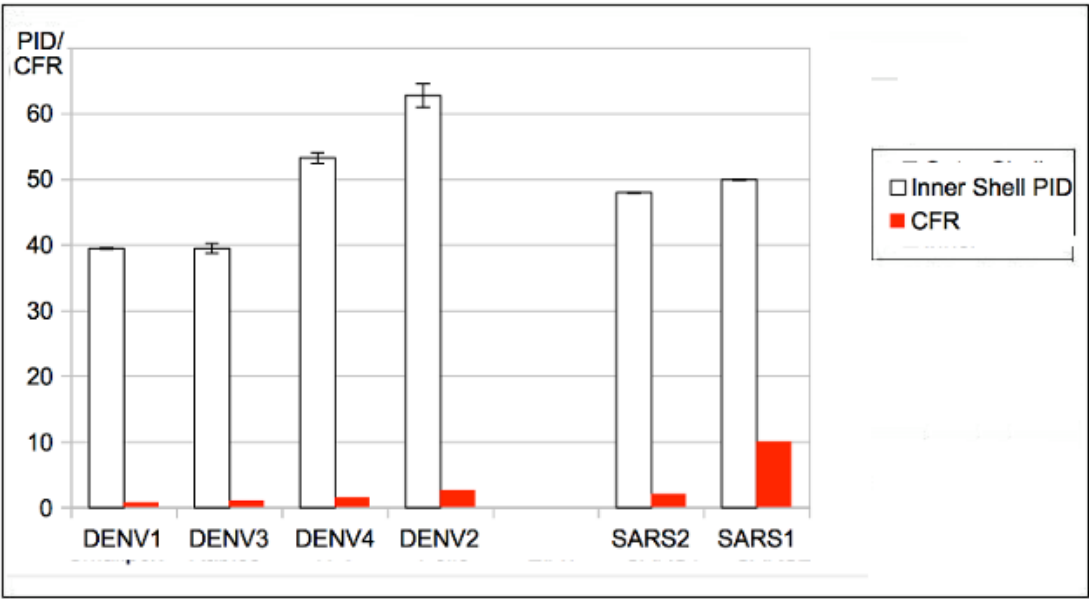


Figure 4

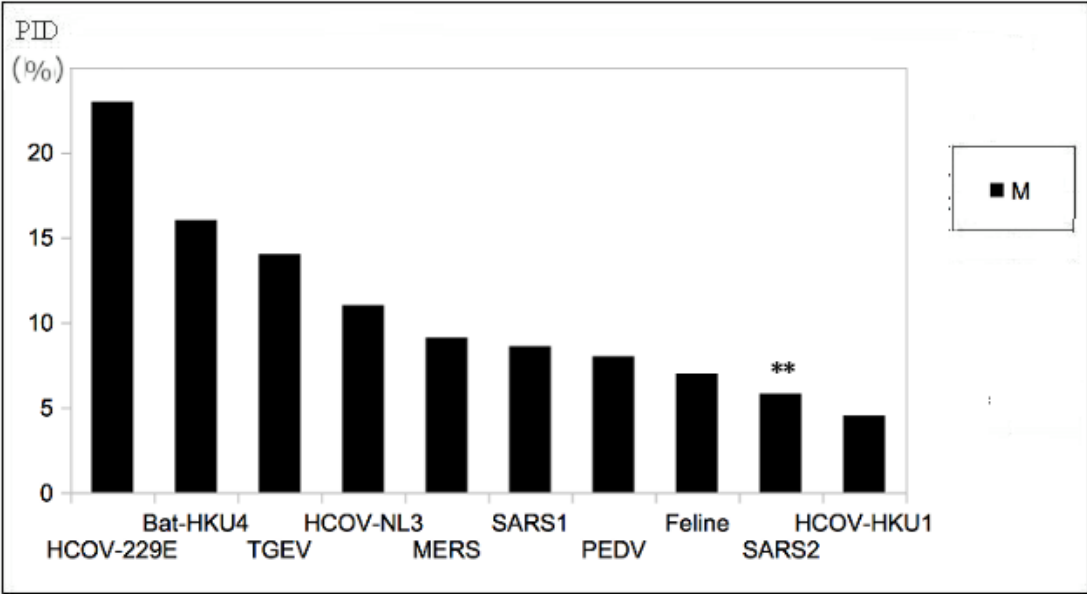


Figure 5

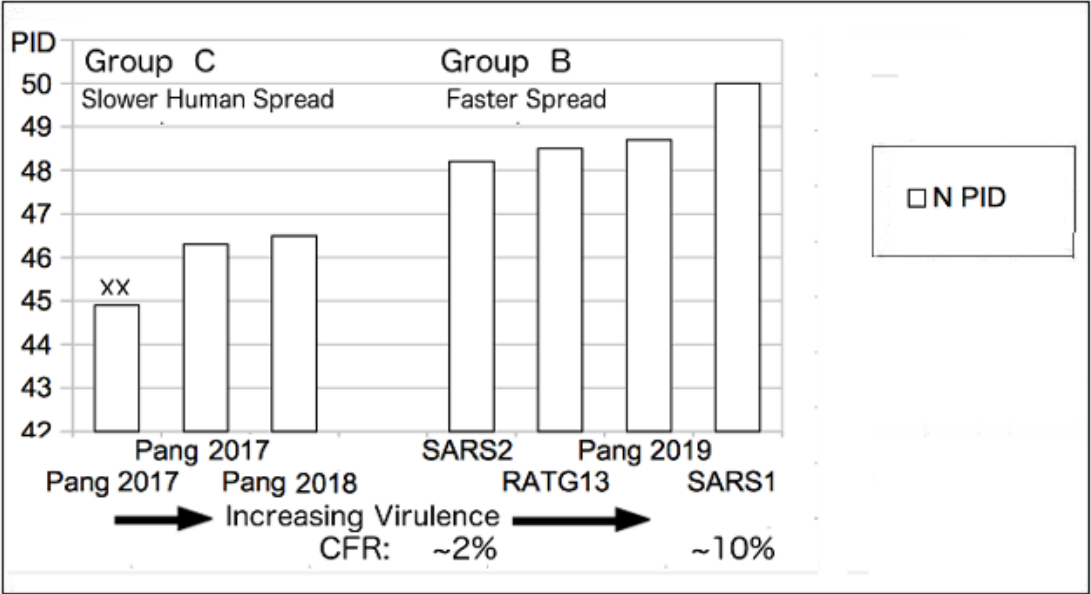
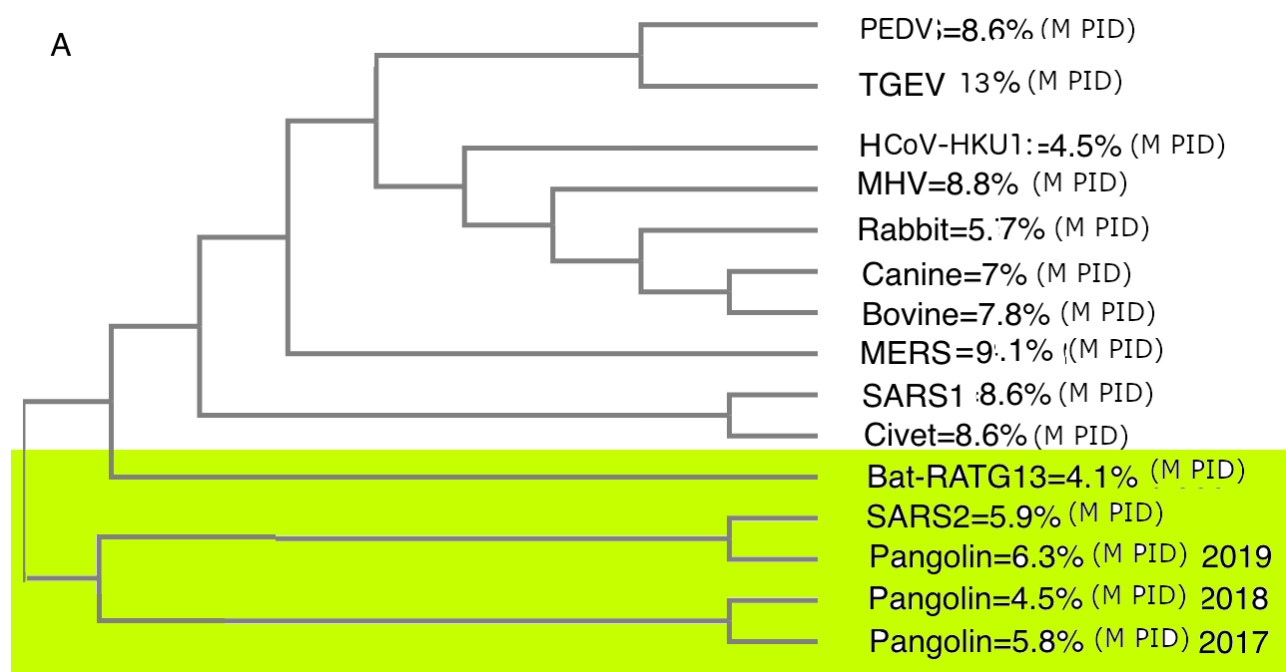


Figure 6



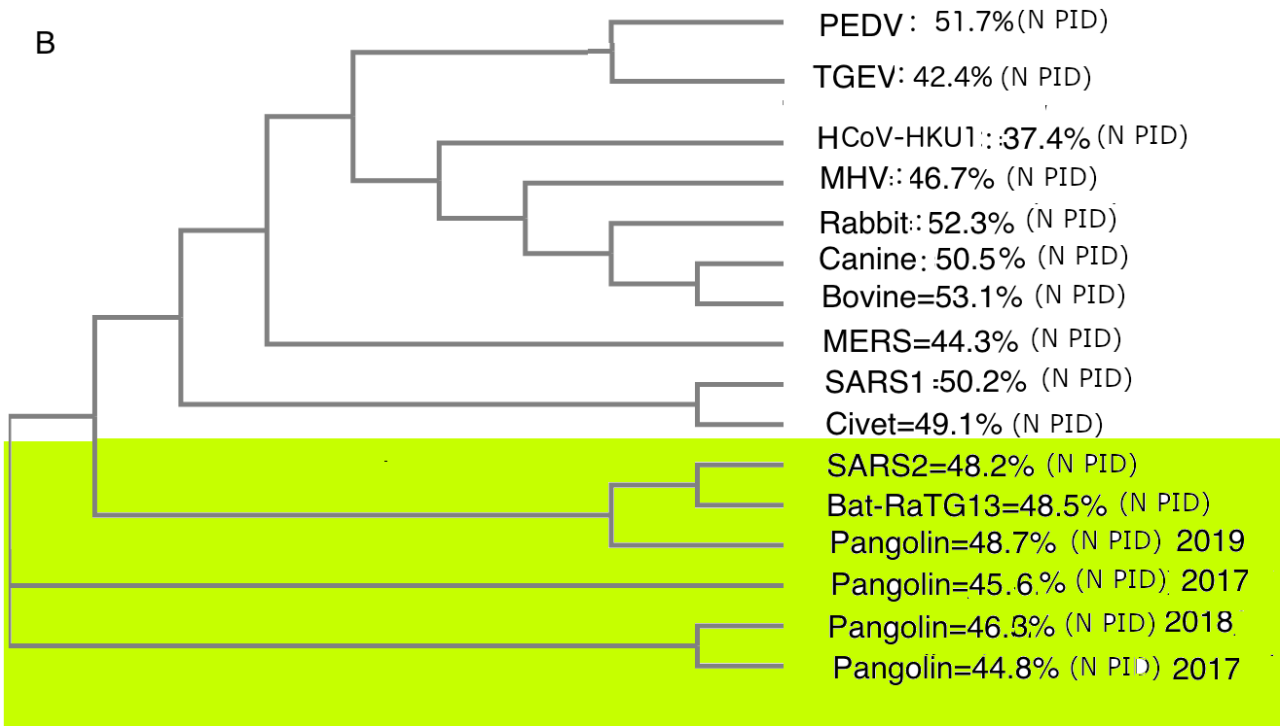


Figure 7

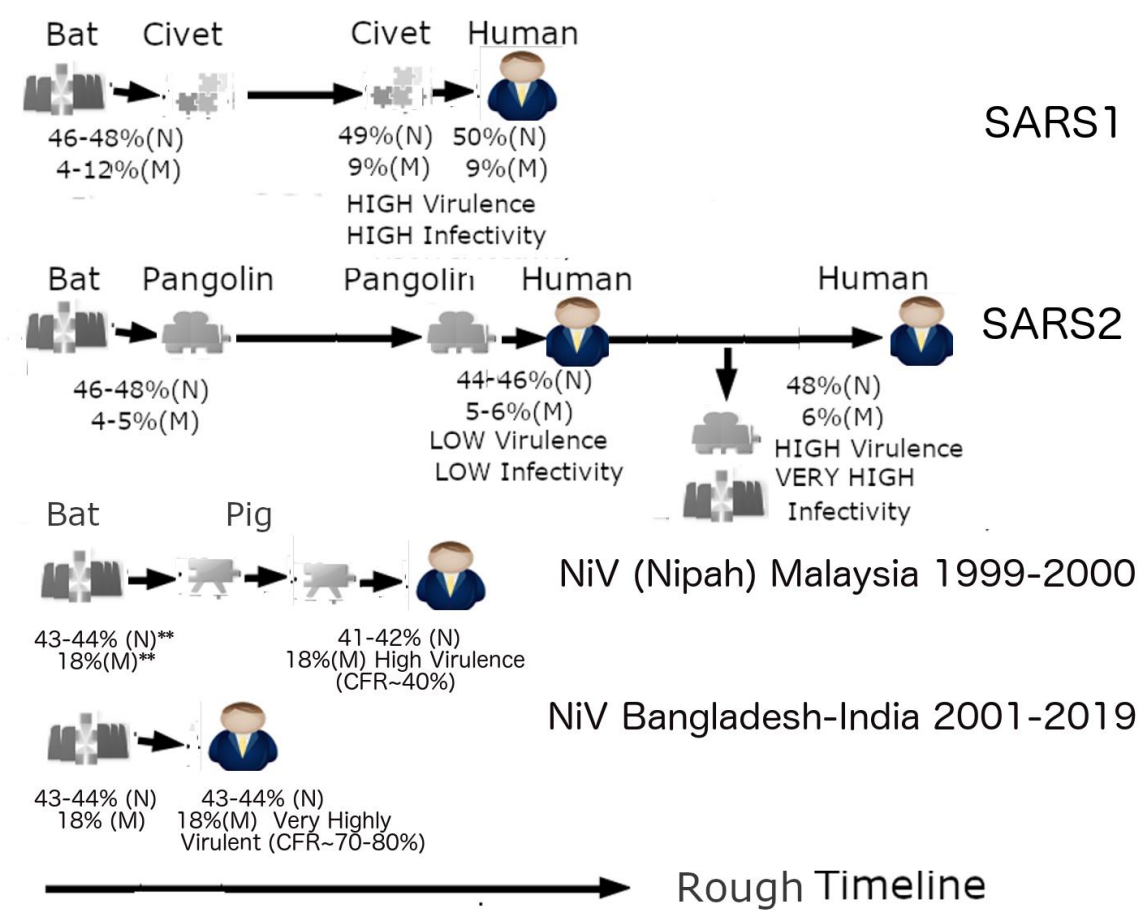


Figure 8

