 doi:10.20944/preprints202004.0171.v2

Title: Systems Biology Approach to Characterize Potential SARS-CoV-2 Pathways Based on Protein Functional Motifs

Running title: COVID-19 protein motif and molecular pathways

Author: Haitham Sobhy

Address: Laboratory of dynamics host-pathogen interactions (DHPI), IUT Louis Pasteur, University of Strasbourg, Schiltigheim, France.

E-mail: sobhy@unistra.fr, and haithamsobhy@gmail.com

Abstract: 140 words

Text: 2429 words

Figures: 1

Supplementary file: 1

Keywords

Protein functional domains, short linear motifs, coronaviruses, COVID-19, severe acute respiratory syndrome-related coronavirus, 2019-nCoV, virus outbreak,

 doi:10.20944/preprints202004.0171.v2

Abstract

Although phylogenetic analysis shows coronaviruses (CoV) share similar genome sequences, CoVs encode different number of proteins (5 to 14), which has implication on viral pathogenicity and infection. Here, we aimed to identify (*in-silico*) the similarities between different members of coronavirus family. The analysis included 50 coronavirus proteomes, including SARS-CoV-2 (COVID-19), to find the variation of the number of protein functional motifs and domain in each coronavirus. For this role, we used the experimentally validated domain (motif) that known to be crucial for viral infection. Although human CoVs are classified within one genus, we found variations among them. SARS-CoV-1, SARS-CoV-2 and MERS-CoV encode different type of domains, which has implications on the molecular interactions triggered by each virus within human cells. Secondly, we used functional motifs to reconstruct the potential molecular pathways or interactions triggered by SARS-CoV-2 proteins within human cell.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2020 doi:10.20944/preprints202004.0171.v2

Introduction

The coronavirus outbreak (coronavirus disease-19, COVID-19) is thought to be initiated by a zoonotic virus that transmitted to human. Genome sequencing reveals that the causative virus is named severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2, SARS-2), and belongs to genus *Betacoronavirus*, family *Coronaviridae* [1-4]. The virus shares sequence homology with coronaviruses infecting bats and murine, such as bat SARS (bat-SL-CoVZC45), and classified in so-called subgenus *Sarbecovirus*. Furthermore, the first SARS-CoV (SARS-1) outbreak occurred in 2003, from zoonotic origin as well. The phylogenetic analysis showed the close relationship with bat coronavirus (e.g. BtCoV_279/2005 and BtRs-BetaCoV/YN2018). In 2012, another zoonotic virus was transmitted to human, so-called Middle East respiratory syndrome coronavirus (MERS-CoV). The virus belongs to subgenus *Merbecovirus*, and shares sequence homology with bat CoV HKU4, HKU5, and betacoronavirus Erinaceus. The third subgenus *Embevovirus*, which includes human CoV HKU1 and OC43 and share homology with murine CoV MHV-1 and betacoronavirus HKU24.

The phylogenetic analysis shows coronaviruses are relatively conserved within the same subgroup [1]. However, coronaviruses encode diverse number of proteins, ranging from 5 protein in bat CoV, to 14 in SARS CoV, table S1, figure 1A and 1B. They evolve rapidly emerging new strains that cause outbreaks and life-loss. Although the high rate of homology of CoV genomes, introduction of few nucleotides may lead to significant changes on the short protein motifs or domains. Particularly, the newly isolated viruses (despite of evolutionary relationships with other CoVs) may encode new proteins of unknown functions and utilize different molecular interactions and pathways within the host cell [5-9]. These domains could contribute in viral virulence and ability to infect wide-range of host cells.

This analysis highlights the role of functional protein domains and the potential pathways that could be triggered during CoVs infection. We used text mining approach to search for over 1500 canonical motifs, including those known to be involved in viral infection. We then compared between the functional motifs encoded by different coronaviruses, including SARS-CoV-2. Additionally, we reconstructed the potential SARS-CoV-2-human protein interactions network and pathways based on the encoded virus motifs.

Method

In this analysis, the full protein sequences of 51 coronaviruses, including the newly isolated SARS-CoV-2 are obtained from NCBI database, March, 2020. The functional motifs in the coronavirus proteomes identified using exact text-finding (mining) implemented in Shetti

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2020 doi:10.20944/preprints202004.0171.v2

and Shetti-Motif tools, for detailed method [6], in addition to in-house scripts, which can be supported upon request, https://sites.google.com/site/haithamsobhy/software. For short, we used the datasets of functional motifs that experimentally validated to be involved in viral infection and virulence, see reviews [7-9]. Additional protein domains were downloaded from PROSITE database (https://prosite.expasy.org/). Together these construct a dataset of over 1500 experimentally validated motifs and domains. Then the number of proteins harbouring each motif calculated and normalized to number of the proteins in the proteome. We performed hierarchical clustering (MeV tool, http://mev.tm4.org/) of the matrix of motif-proteome enrichment (proteome as columns and motifs as rows), figure 1C. The aim of the matrix is to visualize the dynamics between different proteomes, and cluster each proteome based on the encoded motifs and hence the potential molecular interactions that could be triggered. Similarly, to predict the functions and potential pathways that could be triggered by each protein encoded by SARS-CoV-2, the search approach and the datasets of functional motif were used for individual SARS-CoV-2 protein.

Results

Coronaviruses encode different type of functional motifs

We counted number of proteins in each CoV proteome that harbour certain functional motif patterns. The results show that CoVs encode different type of functional motifs from one virus to another, even the closely related encode some different motifs, table S2-S8. Clustering based on the number of proteins harbouring the motifs could reveal the similarities between the viruses, and potential virus motif-host protein interactions that could be triggered by certain virus within the host cell. As expected we found both SARS-CoV-2 and bat-SL-CoVZC45 clustered in one clade close to SARS-CoV-1, whereas MERS-CoV clustered in other clade, figure 1C. This shows that the SARS-1 and 2 might trigger similar molecular pathways within host cell, compared to MERS-CoV.

Interestingly, the proteomes of the evolutionary-closely related viruses infecting the same host (e.g. SARS-CoV-1 and 2) encode some different motifs. For example, two motifs that required for cell signalling transduction and they can recognize TRAF2 protein and PDZ domain-containing proteins are deleted in SARS-CoV-2, but encoded in closely related bat-SL-CoVZC45 virus. However, SARS-CoV-2 encodes multiple domains that can recognize ubiquitination proteins, such as E3 ubiquitin ligases (E3 Ub), Elongin C (ELOC), TRAF6 and SIAH1. These motifs are either deleted or less represented by other closely related viruses. For example, the canonical PPxY motif (x means any residue) is largely encoded and utilized by

 doi:10.20944/preprints202004.0171.v2

viruses to hijack the cellular machinery, reviewed [7]. The viruses utilize the motif to recruit NEDD4 E3 ubiquitin ligases for protein degradation, and ESCRT (endosomal sorting complexes required for the transport) pathway. ESCRT pathway is crucial for HIV-1 and paramyxoviruses budding and exit from the cell. Additionally, adenoviruses utilize PPxY motif during cell entry and cellular trafficking. Here, the surface glycoprotein (S) of SARS-CoV-2 harbour PPxY motif. Two proteins of MERS-CoVs, and three proteins of *Erinaceus* CoV harbour the same motif, table S2. However, the motif is not encoded by human CoVs nor SARS-1.

Coronaviruses encode other canonical ESCRT-interacting motifs, such as P[T/S]AP, [F/I/L/V]PxV, YxxL, and LYPxL. Although ORF1ab polyprotein of SARS-2 harbours LYPTL, LPGV and VPFV motifs, the virus does not encode P[T/S]AP motif, which is encoded only by MERS-CoV EMC/2012, table S2. P[T/S]AP motif recruits TSG101 protein, a component of ESCRT-I, whereas LYPxnL recruits ALIX, a component of ESCRT-III complex, reviewed in [7,10-13]. In fact, ESCRT pathway and viral budding are considered as potential antiviral therapeutics, such as quinolones that include FGI-104, FGI-103, FGI-106, and chloroquine.

Among all coronaviruses, only ORF1ab polyprotein of SARS-2 harbours the tubulin-beta mRNA autoregulation signal motif (PROSITE accession: PS00228). β-tubulin is able to auto-regulate its expression through the binding between the polymerized tubulin to the MREI motif of the nascent tubulin peptides, which by an unknown mechanism can terminate the translation of tubulin, causing the regulation of the expression level of tubulin. Noteworthy, all tubulins harbour MR[E/D][I/L] motif in the N-terminal end of the tubulin, which thought it is a signature of tubulin family. To test the hypothesis that this motif is widely-encoded by viruses, we searched for the tubulin motif in herpesvirus, iridovirus and poxvirus proteins. The search shows that some of these large viruses (the genome sizes are 10-20 times larger than CoV) do not encode the tubulin motif, while some of these large proteomes harbour only one version. The exact biochemical and structural mechanism by which the tubulins recognize the motif and regulate the translation is remain to be discovered. Therefore, the motif could have a regulatory function during virus replication/infection, however the exact role remains to be elucidated.

That said, SARS-1 and 2, but not MERS encode motif that recognizes host cell factor 1 (HCFC1), which is crucial to regulate the cell cycle. Additionally, two Cys-rich motifs were predicted to link between spike and envelope proteins [14]. These two motifs are encoded by SARS-1 and 2, but not MERS subgroup. Other C-rich motifs, which are needed for baculovirus virions production and nucleocapsid assembly, are encoded by multiple coronaviruses, table S2. On the other hand, coronaviruses harbour multiple integrin-binding (RGD) motif. In absence of RGD, viruses may utilize other motifs to attach to cellular receptors and enter into

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2020 doi:10.20944/preprints202004.0171.v2

host cells, such as KGE, LDV, LDI and SDI, which are encoded by coronaviruses as well, table S2.

Reconstruction of CoV-human protein interactions network based on functional motif

We used the functional motif encoded by each SARS-CoV-2 protein to predict the potential molecular interactions and pathways within the host cell, figure 1D, table S5. We observed that SARS-2 encodes multiple Ub-, and SUMO-binding motif, including the PPxY motif, which is localized on surface (S) protein. Recruiting ubiquitin proteins are essential to degrade the antiviral proteins and hijack the immune response imposed by the cells. SARS-2 encodes motifs to recognize heparan sulfate (HS), which is required for post-internalization events of viral entry, discussed in [7]. As mentioned above ORF1ab harbours MREI-like tubulin motif. Besides, it harbours the canonical motif for protein trafficking and nuclear localization signal (NLS), which is essential for trafficking through the nuclear membrane.

On the other hand, furin endoprotease belongs to group of proprotein convertases that cleave the precursor proteins to the activated form. It binds to the canonical motif RxRK/R||x, where || denotes the cleavage site. Coronaviruses characterize by the presence of R||S and RRRR||S motifs. The motif is required for interaction with furin, which leads to proteolytic activation of the spike protein [15]. As a result the virus enters into host cells. Besides, its role in the conversion of the precursor proteins, R||S motif is essential for syncytium formation [15]. ORF1ab is the largest polyprotein encoded by SARS-2, which is auto-proteolytically processed into 16 non-structural proteins (Nsp1-16), reviewed in [4]. The $Rx_{0-3}RS$ motifs can be detected in ORF1ab, which could correspond to the conversion from the polyprotein form to the non-structural protein form.

SARS-2 harbours clathrin-binding motifs and clathrin adaptor protein (AP)-binding motifs, which are required for endocytosis. Coronaviruses enter the cells by fusion or endocytosis, which may require clathrin for some strains, reviewed in [16]. The integrins (ITGs) and heparan sulfates could have role in the entry into host cell. Noteworthy, HIV-1 utilizes AP-binding motifs to direct anti-tetherin (BST2) to the lysozyme and antagonizes the antiviral immune response, reviewed in [7]. Regarding the cellular signalling, ORF1b polyprotein harbours motifs involved in multiple cellular signalling, including JAK, MAPK, TRADD, TRAF6 and caspases-binding motifs. Caspases and TRAFs were linked with inflammation and apoptosis [17,18]. Multiple viruses (e.g. herpesviruses and influenza) hijack the caspase pathways for fruitful infection and to regulate the programmed cell death.

Among the interesting motifs, ORF1ab harbours the canonical motif for binding with palmitoyl acyltransferase, in addition to multiple thiol disulphide and Cys-rich motifs, which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2020 doi:10.20944/preprints202004.0171.v2

are needed for protein palmitoylation [19]. Noteworthy, the envelope (E) protein of some strains of coronaviruses were shown to be palmitoylated [20-22]. Additional lipid post-translational modification event of proteins is myristoylation [23]. ORF1ab harbours the canonical MGxxxS motif for binding with N-myristoyltransferase (NMT1), which adds a myristoyl group to the N-terminal glycine residue of the proteins [23]. Of note, myristoylation is crucial for egress of some viruses, including coronaviruses [24]. The myristoylation is crucial for activation and inhibition of the immune response by phosphorylation of tyrosine residues in so-called, immunoreceptor tyrosine-based activation motif (ITAM) and the immunoreceptor tyrosine-based inhibition motif (ITIM) [23]. SARS-2 proteins encode both ITAM and ITIM motifs.

SARS-2 proteins harbour multiple domain-interacting motifs, such as motifs recognizing PDZ, SH2, SH3 domains. These motifs are essential for cellular signalling and protein trafficking within the host cells. For example, the PDZ-binding and DLLV motifs in SARS-1 E protein can influence the subcellular localization of PALS1 (MPP5), which may disrupt the tight junction and apicobasal polarity of the cell [25]. The L-rich and PDZ-binding motifs are used by multiple viruses to recruits mTORC1 and kinase signaling, and translation initiation [7]. In SARS-2, ORF7b harbours two L-rich motifs, including one confers to the LxxLL pattern, which is crucial for HIV retrotransposition, and papillomavirus-induced oncogenesis and cell transformation. S, N, M and ORF1ab proteins harbour additional L-rich motifs.

An interesting phenomenon in viruses is the ability of a virus to disturb the host transcription for the sake of the viral gene expression. An example is the adenovirus E1A oncoprotein that regulates host transcription by binding to transcription regulators, histone acetyltransferase/CREB-binding protein (p300/CBP), through the Fx[DE]xxxL motif [26,27]. Similarly, the adenoviral E1A and the Epstein-Barr virus EBNA2 oncoproteins interact with the C-terminal Mynd domain of ZMYND11 (BS69) transcription regulator, which is facilitated by PxLxP motif in the viral proteins [28]. The transcription regulation motifs can be observed on both ORF1ab (not ORF1a) and N protein sequences, suggesting their potential roles in virus replication. Furthermore, oncoproteins with conserved LxCxE motif can phosphorylate and inactivate retinoblastoma protein (RB1), leading to initiation of the gene expression and virus replication, reviewed in [7]. Although the motif is a five residues long, ORF1ab is the only protein that harbours this motif. Additional contributor on viral replication is ORF3a protein, which harbours the HCFC1-binding motif. HCFC1 regulates the cell cycle by recruiting the regulator p300 and histone deacetylase (HDACs).

 doi:10.20944/preprints202004.0171.v2

Discussion

This analysis highlights the fact that coronaviruses harbour different type of functional motifs. The newly isolated virus could acquire new motifs and adapt new molecular functions or pathways. The fact that SARS-1, MERS and SARS-2 are different and they trigger different molecular pathways should be considered in the future studies.

Furthermore, we used functional motifs and domain to predict the potential interactions that can be triggered by each of coronavirus SARS-CoV-2 proteins, figure 1D. In consistent with our analysis, a recent interactomics study shows that SARS-CoV-2 proteins interacts with multiple ubiquitin ligases, kinases and lipid modifications of proteins [29]. Moreover, there are evidences that SARS-CoVs uses PDZ-binding motifs (PBM) during infection. To study the implication of PBM on the viral infection, PBM located in SARS-1 E protein was mutated and deleted [30]. Astonishingly, the virus restored the motif after several passages *in vitro* or in mice, however mutated PBM in nsp1 protein leads to virus attenuation. Of note, about 5-6 proteins of SARS-CoV-2 harbour PDZ-binding motifs; the same applies on SH2 and SH3 domains, and Ub and SUMO-biding motifs. Noteworthy, ORF1ab is the largest protein and it contains almost a copy of all the binding motifs. These findings can be read as that multiple copies of the same motifs (particularly those encoded by ORF1ab) are kept by virus for restoration of the motifs, in case of drastic loss of the crucial motifs, however this possibility remains to be validated.

Finally, the advantage of our *in silico* analysis is the usage of dataset of the motifs (domains) that have been experimentally validated by multiple methods for other viruses and sometimes by multiple viruses, which increases the robustness, discussed in [5-7]. The resulted datasets can make the future functional studies easy and can enrich the attempts to understand coronaviruses (e.g. COVID-19) infection and the attempts to find antiviral drug. Most of the domains used in the analysis are confer to a structure that increases their importance during binding with host cell protein domains. Interestingly, viruses could acquire the motifs from an evolutionary distant virus or organism. It is of interest to study the molecular mechanisms govern the transfer of the protein motifs and domains, which helps to predict the future and emerging pathogens.

References

1. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.*, et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **2020**, *395*, 565-574.

2. Narayanan, K.; Huang, C.; Makino, S. Sars coronavirus accessory proteins. *Virus Res* **2008**, *133*, 113-121.
3. Schoeman, D.; Fielding, B.C. Coronavirus envelope protein: Current knowledge. *Virol J* **2019**, *16*, 69.
4. Song, Z.; Xu, Y.; Bao, L.; Zhang, L.; Yu, P.; Qu, Y.; Zhu, H.; Zhao, W.; Han, Y.; Qin, C. From sars to mers, thrusting coronaviruses into the spotlight. *Viruses* **2019**, *11*.
5. Sobhy, H. Virophages and their interactions with giant viruses and host cells. *Proteomes* **2018**, *6*.
6. Sobhy, H. A bioinformatics pipeline to search functional motifs within whole-proteome data: A case study of poxviruses. *Virus Genes* **2017**, *53*, 173-178.
7. Sobhy, H. A review of functional motifs utilized by viruses. *Proteomes* **2016**, *4*.
8. Davey, N.E.; Trave, G.; Gibson, T.J. How viruses hijack cell regulation. *Trends Biochem Sci* **2011**, *36*, 159-169.
9. Hraber, P.; O'Maille, P.E.; Silberfarb, A.; Davis-Anderson, K.; Generous, N.; McMahon, B.H.; Fair, J.M. Resources to discover and use short linear motifs in viral proteins. *Trends Biotechnol* **2020**, *38*, 113-127.
10. Williams, R.L.; Urbe, S. The emerging shape of the escrt machinery. *Nat Rev Mol Cell Biol* **2007**, *8*, 355-368.
11. Sette, P.; Jadwin, J.A.; Dussupt, V.; Bello, N.F.; Bouamr, F. The escrt-associated protein alix recruits the ubiquitin ligase nedd4-1 to facilitate hiv-1 release through the lypxnl l domain motif. *J Virol* **2010**, *84*, 8181-8192.
12. Han, Z.; Madara, J.J.; Liu, Y.; Liu, W.; Ruthel, G.; Freedman, B.D.; Harty, R.N. Alix rescues budding of a double ptap/ppey l-domain deletion mutant of ebola vp40: A role for alix in ebola virus egress. *J Infect Dis* **2015**, *212 Suppl 2*, S138-145.
13. Wolff, S.; Ebihara, H.; Groseth, A. Arenavirus budding: A common pathway with mechanistic differences. *Viruses* **2013**, *5*, 528-549.
14. Wu, Q.; Zhang, Y.; Lu, H.; Wang, J.; He, X.; Liu, Y.; Ye, C.; Lin, W.; Hu, J.; Ji, J.*, et al.* The e protein is a multifunctional membrane protein of sars-cov. *Genomics Proteomics Bioinformatics* **2003**, *1*, 131-144.
15. Yamada, Y.; Liu, D.X. Proteolytic activation of the spike protein at a novel rrrr/s motif is implicated in furin-dependent entry, syncytium formation, and infectivity of coronavirus infectious bronchitis virus in cultured cells. *J Virol* **2009**, *83*, 8744-8758.
16. Belouzard, S.; Millet, J.K.; Licitra, B.N.; Whittaker, G.R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **2012**, *4*, 1011-1033.
17. Man, S.M.; Karki, R.; Kanneganti, T.D. Molecular mechanisms and functions of pyroptosis, inflammatory caspases and inflammasomes in infectious diseases. *Immunol Rev* **2017**, *277*, 61-75.
18. Nainu, F.; Shiratsuchi, A.; Nakanishi, Y. Induction of apoptosis and subsequent phagocytosis of virus-infected cells as an antiviral mechanism. *Front Immunol* **2017**, *8*, 1220.
19. Sobocinska, J.; Roszczenko-Jasinska, P.; Ciesielska, A.; Kwiatkowska, K. Protein palmitoylation and its role in bacterial and viral infections. *Front Immunol* **2017**, *8*, 2003.
20. Corse, E.; Machamer, C.E. The cytoplasmic tail of infectious bronchitis virus e protein directs golgi targeting. *J Virol* **2002**, *76*, 1273-1284.
21. Liao, Y.; Yuan, Q.; Torres, J.; Tam, J.P.; Liu, D.X. Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology* **2006**, *349*, 264-275.
22. Lopez, L.A.; Riffle, A.J.; Pike, S.L.; Gardner, D.; Hogue, B.G. Importance of conserved cysteine residues in the coronavirus envelope protein. *J Virol* **2008**, *82*, 3000-3010.

23. Udenwobele, D.I.; Su, R.C.; Good, S.V.; Ball, T.B.; Varma Shrivastav, S.; Shrivastav, A. Myristoylation: An important protein modification in the immune response. *Front Immunol* **2017**, *8*, 751.
24. Du, Y.; Zuckermann, F.A.; Yoo, D. Myristoylation of the small envelope protein of porcine reproductive and respiratory syndrome virus is non-essential for virus infectivity but promotes its growth. *Virus Res* **2010**, *147*, 294-299.
25. Teoh, K.T.; Siu, Y.L.; Chan, W.L.; Schluter, M.A.; Liu, C.J.; Peiris, J.S.; Bruzzone, R.; Margolis, B.; Nal, B. The sars coronavirus e protein interacts with pals1 and alters tight junction formation and epithelial morphogenesis. *Mol Biol Cell* **2010**, *21*, 3838-3852.
26. Pelka, P.; Ablack, J.N.; Fonseca, G.J.; Yousef, A.F.; Mymryk, J.S. Intrinsic structural disorder in adenovirus e1a: A viral molecular hub linking multiple diverse processes. *J Virol* **2008**, *82*, 7252-7263.
27. Ferreon, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Structural basis for subversion of cellular control mechanisms by the adenoviral e1a oncoprotein. *Proc Natl Acad Sci U S A* **2009**, *106*, 13260-13265.
28. Ansieau, S.; Leutz, A. The conserved mynd domain of bs69 binds cellular and oncoviral proteins through a common pxlxp motif. *J Biol Chem* **2002**, *277*, 4906-4910.
29. 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.*, et al.* A sars-cov-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**.
30. Jimenez-Guardeno, J.M.; Regla-Nava, J.A.; Nieto-Torres, J.L.; DeDiego, M.L.; Castano-Rodriguez, C.; Fernandez-Delgado, R.; Perlman, S.; Enjuanes, L. Identification of the mechanisms causing reversion to virulence in an attenuated sars-cov for the design of a genetically stable vaccine. *PLoS Pathog* **2015**, *11*, e1005215.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2020 doi:10.20944/preprints202004.0171.v2

Figure Legend

Figure 1. The genetic architecture of (A) MERS-CoV and (B) SARS-CoV-1, modified from [4]. (C) The heatmap visualizes the present of proteins harbour certain motif (rows), whereas the columns represent the coronavirus proteomes. The proteomes were clustered (hierarchical clustering) based on the present of enriched proteins, data constructed from table S3. (D) The reconstruction of potential protein-protein interactions network or pathways that could be triggered by each SARS-CoV-2 protein. **Abbreviations**: **ALIX** = Programmed cell death 6-interacting protein; **Anti-tetherin** = Motif antagonizes the bone marrow stromal antigen 2 (BST2); **AP-1 / 2** = Adaptor protein complex AP-1 or AP-2; **APOBEC3G** = Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G; **ATP-binding** = Walker's motifs (ATP-binding motif); **C2HC Zn** = Motif recognizes the CysCysHisCys (C2HC) type zinc finger domain; **CASPs** = Caspases; **COP1** = E3 ubiquitin-protein ligase COP1; **E3 Ub** = E3 ubiquitin ligases; **EIF2A** = Eukaryotic translation initiation factor 2A; **ELOC** = Elongin C; **ESCRT** = Endosomal sorting complexes required for the transport; **FZR1 (Cdh1)** = Fizzy-related protein homolog; **HCFC1** = Host cell factor 1; **HS** = Heparan sulfate; **ITGs** = Integrins; **ITAM** = Immunoreceptor tyrosine-based activation motif; **ITIM** = Immunoreceptor tyrosine-based inhibition motif; **JAK** = Tyrosine-protein kinase; MAPK = Mitogen-activated protein kinase; **NECAP1** = Adaptin ear-binding coat-associated protein 1; **NEDD4** = E3 ubiquitin-protein ligase NEDD4; **NES** = Nuclear export signal; **NLS** = Nuclear localization signal; **NMT1** = Glycylpeptide N-tetradecanoyltransferase 1; **P in black box** = Phosphorylation; **p300/CBP** = Histone acetyltransferase/CREB-binding protein; **PACS1** = Phosphofurin acidic cluster sorting protein 1; **PDZ** = Motif recognizes the PDZ-domain, post-synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-I protein (zo-1); **PP-1A** = Serine/threonine-protein phosphatase PP1-alpha catalytic subunit; **RB1** = Retinoblastoma-associated protein; **SH2 and SH3** = Motif recognizes the SRC Homology 2 and 3 domain; **SUMO** = Small ubiquitin-related modifier binding motif; **TR** = Thyroid hormone (TH) receptors (TRs); **TRADD** = Tumor necrosis factor receptor type 1-associated DEATH domain protein; **TRAF6** = TNF receptor-associated factor 6; **WASL** = WAS/WASL-interacting protein family member 1; **ZMYND11** = Zinc finger MYND domain-containing protein 11.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2020 doi:10.20944/preprints202004.0171.v2