Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Original article

Computational drug screening against the SARS-CoV-2 Saudi Arabia isolates through a multiple-sequence alignment approach

Pooi Ling Mok, Avin Ee-Hwan Koh, Aisha Farhana, Abdullah Alsrhani, Mohammad Khursheed Alam, S. Suresh Kumar

PII: S1319-562X(21)00052-8
DOI: https://doi.org/10.1016/j.sjbs.2021.01.051
Reference: SJBS 2196

To appear in: Saudi Journal of Biological Sciences

Received Date: 1 October 2020
Revised Date: 19 January 2021
Accepted Date: 21 January 2021

Please cite this article as: P. Ling Mok, A. Ee-Hwan Koh, A. Farhana, A. Alsrhani, M. Khursheed Alam, S. Suresh Kumar, Computational drug screening against the SARS-CoV-2 Saudi Arabia isolates through a multiple-sequence alignment approach, Saudi Journal of Biological Sciences (2021), doi: https://doi.org/10.1016/j.sjbs.2021.01.051

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.
Computational drug screening against the SARS-CoV-2 Saudi Arabia isolates through a multiple-sequence alignment approach

Pooi Ling Mok$^{1,2,3,#,*}$, Avin Ee-Hwan Koh$^{2,*}$, Aisha Farhana$^1$, Abdullah Alsrhani$^1$, Mohammad Khursheed Alam$^4$, S. Suresh Kumar$^{5,6,#}$

$^1$Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Jouf University, Sakaka, Al-Jawf Province, Saudi Arabia.
$^2$Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
$^3$Genetics and Regenerative Medicine Research Centre, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
$^4$Department of Orthodontics, College of Dentistry, Jouf University, Sakaka, Al-Jawf Province, Saudi Arabia.
$^5$Department of Medical Microbiology and Parasitology, Universiti Putra Malaysia, 43400 UPM Serdang Selangor, Malaysia.
$^6$UPM-MAKNA Cancer Research Laboratory, Institute of Bioscience, Universiti Putra Malaysia, Serdang 43400 UPM, Selangor, Malaysia.

*Equally-contributing first author

#Corresponding author: Dr. Mok Pooi Ling; Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Jouf University, Sakaka, Al-Jawf Province, Saudi Arabia; Email: Mpling@ju.edu.sa or rachelmok2005@gmail.com

#Co-corresponding author: Dr. S. Suresh Kumar; Department of Medical Microbiology and Parasitology, Universiti Putra Malaysia, 43400 UPM Serdang Selangor, Malaysia.; Email: suresh@upm.edu.my or sureshkudsc@gmail.com

Abstract

COVID-19 is a rapidly emerging infectious disease caused by the SARS-CoV-2 virus currently spreading throughout the world. To date, there are no specific drugs formulated for it, and researchers around the globe are racing against the clock to investigate potential drug candidates. The repurposing of existing drugs in the market represents an effective and economical strategy commonly utilized in such investigations. In this study, we used a multiple-sequence alignment approach for preliminary screening of commercially-available drugs on SARS-CoV sequences from the Kingdom of Saudi Arabia (KSA) isolates. The viral genomic sequences from KSA isolates were obtained from GISAID, an open access repository housing a wide variety of epidemic and pandemic virus data. A phylogenetic analysis of the present 164 sequences from the KSA provinces was carried out using the MEGA X software, which displayed high similarity (around 98%). The sequence was then analyzed using the VIGOR4 genome annotator to construct its genomic structure. Screening of existing drugs was carried out by mining data based on viral gene expressions from the ZINC database. A total of 73 hits were generated. The viral target orthologs were mapped to the SARS-CoV-2 KSA isolate sequence by multiple sequence
alignment using CLUSTAL OMEGA, and a list of 29 orthologs with purchasable drug information was generated. The results showed that the SARS CoV replicase polyprotein 1a had the highest sequence similarity at 79.91%. Through ZINC data mining, tanshinones were found to have high binding affinities to this target. These compounds could be ideal candidates for SARS-CoV-2. Other matches ranged between 27 – 52%. The results of this study would serve as a significant endeavor towards drug discovery that would increase our chances of finding an effective treatment or prevention against COVID-19.

Keywords: COVID-19; coronavirus; multiple sequence alignment; tanshinones; Saudi Arabia

1.0 Introduction

The COVID-19 disease initially began as a small outbreak that was first reported in Wuhan, Hubei, China, towards the end of the year 2019. Although the source of the first possible outbreak could not be pinpointed, it began spreading throughout the world. By January 2020, COVID-19 was announced as a global health emergency by the World Health Organization (WHO), a status that is shared by the likes of a few, namely the H1N1 swine flu (2009), Polio (2014), and Ebola (2014) outbreaks.

Countries in the middle-east have repeatedly reported infectious disease outbreaks since the last decade. In 2012, the first presumed case of MERS-CoV, a relative to the current SARS-CoV-2 virus, was reported in Saudi Arabia (Chan et al., 2012). The patient developed pneumonia followed by renal failure. The virus, known as HCoV-EMC/2012 at the time, was categorized as a betacoronavirus. Later on, this infection resulted in the second coronavirus (CoV) epidemic (Kandeel et al., 2020; van Boheemen et al., 2012). SARS-CoV-2 is the newest member under this subgenus and was found to share about 50% similarity with MERS-CoV and about 79% with SARS CoV (Pal et al., 2020). During the early phases of the COVID-19 epidemic, three initial spreading or seeding patterns were observed based on population movements into and within Saudi Arabia (Memish et al, 2020). Transmission among international pilgrims heading towards the holy sites of Mecca and Medina accounted for the first infection pattern (Ebrahim & Memish, 2020; Memish et al., 2020). The second pattern involved returning Saudi shiite pilgrims from the eastern province. Finally, the third pattern, which involves general travelers moving in and out of Saudi Arabia (Ebrahim & Memish, 2020; Memish et al., 2020).

As of September 2020, Saudi Arabia has reported more than 300,000 cases and almost 4,000 deaths (Worldometer, 2020). Proactive and intensive responses, especially during the early stages of COVID-19, as well as a previous history in handling MERS-CoV, has primarily helped Saudi Arabia to manage local outbreaks (Algaissi et al., 2020; Barry et al., 2020; Obied et al., 2020). Besides, there have been ongoing efforts in the country to curb viral transmissions, such as the study of the popular hydroxychloroquine and Ivermectin as a candidate drug treatment (Kelleni, 2020; Meo et al., 2020). At present, pharmaceutical companies are constantly devising new drugs in the laboratory before proceeding with clinical trials. However, such drug studies require costly and lengthy endeavors before new compounds can be brought over from bench to bedside. Not to mention, more than half of these investigational studies fail to pass phase-3 of clinical trials, resulting in huge losses (Fogel, 2018; Hwang et al., 2016). This process can be expedited efficiently in an economical way through computational drug discovery, which utilizes bioinformatics processes and data mining on already available huge datasets to identify new drug targets and screen existing drugs for pharmaceutical research (Bharath...
et al., 2020; Ou-Yang et al., 2012). This facilitates the repurposing of de-risked drugs and shortens development timelines immensely, which make it a highly attractive approach, especially during a sudden viral outbreak (Pushpakom et al., 2018). Hence, many paid and open-access software have been developed over the years for this purpose. A comprehensive list of databases for computer-aided drug design and screening can be found on click2drug.org, a directory under the ExPASy bioinformatics resource portal (Artimo et al., 2012).

In this study, we employed a multiple-sequence alignment approach in identifying potential drug candidates for COVID-19 (March-Vila et al., 2017), with an emphasis on isolates obtained from KSA. The genomic sequence of SARS-CoV-2 isolates from KSA (SARS-CoV-2-SA) were obtained from the GISAID database (gisaid.org) (Shu & McCauley, 2017). At the time of this study, 164 sequences were uploaded from KSA, including the province of Jeddah, Madinah, Makkah, and Riyadh. Phylogenetic analysis was carried out on the isolates using MEGA X. The VIGOR4 annotator was then used to construct a genomic structure using the SARS-CoV-2 sequence. The ZINC database was mined to obtain a list of potential compounds against the therapeutic targets of SARS-CoV-2. It is a free resource available at zinc.docking.org to provide researchers with free access to ligand discovery, annotated compounds, purchasable drug information, and more (Irwin et al., 2012). A total of 73 hits were obtained from the database for specific viral proteins, including the ORF1ab region of SARS CoV, which codes some of the essential proteins for coronaviral replication (e.g. 3C-like protease 3CL\textsubscript{PRO}, and papain-like protease PL\textsubscript{PRO}). The targets with annotated drug compounds were then mapped to the SARS-CoV-2 KSA isolate sequence by multiple sequence alignment using CLUSTAL OMEGA. A list of 29 orthologs, including its similarity indexes, was tabulated. Through data mining in ZINC, high-affinity binding drugs to the viral targets were generated.

2.0 Methods and materials

2.1 Genomic data information and collection
The complete genomic sequences of SARS-CoV-2 isolates from Saudi Arabia (SARS-C0V-2-SA) were downloaded from the GISAID database at gisaid.org. All downloaded 164 sequences were obtained from highly affected regions in Jeddah, Madinah, Makkah, and Riyadh. These also included a list of accession IDs, submission dates, originating labs, and other relevant information (supplementary file gisaid_hcov-19).

2.2 Nucleotide sequence alignment and phylogenetic tree analysis
The sequences were first aligned using the alignment explorer feature of MEGA X (Khan, 2017; S. Kumar, Stecher, Li, Knyaz, & Tamura, 2018). The ClustalW (codons) feature was used for the alignment. The gap opening and gap extension penalty were set to 10.00 and 0.20 respectively, as recommended by Newman, Duffus, and Lee (2016) (Newman et al., 2016). The aligned sequences were exported and then analyzed to construct the phylogenetic tree. The statistical method employed was the maximum likelihood tree, and the model used was the Tamura-Nei model along with the option to delete partial gaps (S. Kumar et al., 2018; Tamura & Nei, 1993).

2.3 Construction of SARS-CoV-2-SA isolate genomic structure
The hCoV-19/Saudi Arabia/KAIMRC-Alghoribi/2020 isolate, one of the first known sequenced SARS-CoV-2 in the region from a 68 y/o local male, was used to construct the viral genome structure. The genomic sequence of the SARS-CoV-2-SA isolate was annotated using the Viral Genome ORF Reader, VIGOR4 annotator. The open-source for this tool is available at github.com/VirusBRC/VIGOR4, or online at vpirbrc.org (Wang, Sundaram, & Spiro, 2010). In brief, the sequence (in FASTA format) was loaded into the tool. The reference genome used was the SARS-CoV-2 Wuhan-Hu-1 isolate (GenBank sequence accession NC_045512.2). A total of 28 annotations in the sequence was analyzed and the results were then tabulated.

2.4 Database screening and multiple sequence alignment
A list of candidate viral drug targets was generated by screening through annotated entries from the ZINC database. The targets with purchasable drug information were then mapped to the SARS-CoV-2-SA isolate sequence by multiple sequence alignment using CLUSTAL OMEGA. The similarity index was generated, and the list of drugs for high similarity orthologs were then screened and tabulated.

2.5 Statistical method
The Maximum Likelihood method was used in the inference and construction of the phylogenetic tree with the highest log likelihood (-1396374.75) based on the Tamura-Nei model. The Neighbor-joining algorithm was applied to a matrix of pairwise distances, and the topology that passes the algorithm was selected to obtain the tree. The site proportion, where at least one unambiguous base is available in one or more sequence per descendent clade, is shown in the tree beside each internal node. The analysis involved 164 nucleotide sequences. There was a total of 30643 positions in the final dataset.

3.0 Results

3.1 High homology between all SARS-CoV-2 isolates from Saudi Arabia
The purpose of constructing a phylogenetic tree of the SARS-CoV-2 isolates from Saudi Arabia (SARS-CoV-2-SA) was to determine the homology and evolutionary relationships between these sequences. The data was obtained from GISAID.org and aligned using MEGA X. The phylogenetic tree was then generated using the whole genome sequences compiled in the software. The results suggest that all isolates from Saudi Arabia demonstrated 97-98% similarity in the genome sequences (Figure 1).

3.2 High similarity of Open Reading Frame (ORF) sequences in the SARS-CoV-2-SA isolates to the Wuhan isolate
Based on the phylogenetic analysis, the genomic sequences were found to be highly similar in all SARS-CoV-2-SA isolates. Thus, in order to determine the sequence structure and function of the viral ORFs in relation to the Wuhan isolate, the first isolate uploaded by King Abdullah International Medical Research (KAIMRC) to GISAID.org was used to build its genomic structure using VIGOR4. The data shows that ORFs from the input KAIMRC genome shows 99%-100% similarity to the Wuhan reference (Table 1). This includes the ORF1ab, ORF3a, ORF6a, ORF7a, ORF7b, ORF8, and ORF10 genes that encode for non-structural and accessory proteins essential for viral replication. In addition, there are also presence of highly similar structural ORFs for S, M, E, and N proteins.

3.3 Tanshinones show highest binding affinity to replicase 1a in SARS-CoV-2
Drug repurposing remains the most practical method for rapidly developing new treatments using existing drugs. In this study, data mining was performed by screening through the ZINC database for known anti-viral targets. A list of targeted viral genes along with the associated acting compounds were generated (Table 2). According to the database, there are currently 73 viral therapeutic targets, about half of which, contain information on known purchasable drugs against these targets. A multiple sequence alignment approach was then adopted to align these viral target sequences with that of the SARS-CoV-2-5A isolate using CLUSTAL OMEGA. The list generated 29 possible viral target groups with associated drugs that may be effective against SARS-CoV-2 (Table 3). The replicase polyprotein 1a (REP) of SARS CoV had the highest similarity (79.91%) to SARS-CoV-2. Based on chembl 2.0, tanshinones had among the highest binding affinities to REP. The E6 protein (52%) of human papilloma virus had the second highest similarity. The flavonoids, myricetin and morin, were among the listed drugs that targeted the E6 protein. The other alignments were found to be below 50%, which might still be potent against SARS-CoV-2.

4.0 Discussion

The COVID-19 disease has resulted in an ongoing pandemic with no specific treatment to date, only supportive care that has been endorsed by the World Health Organization (WHO) (Song et al., 2020). Hence, finding a functional cure or even an effective therapy is paramount in stopping the SARS-CoV-2 global outbreaks. At present, based on the analysis of over 48,000 SARS-CoV-2 genomes worldwide from the GISAID database, there are 7 clades (G, GH, GR, L, O, S, V) (Mercatelli & Giorgi, 2020). Each possesses characteristic variants, such as the spike protein variant S-D614G (clade G) and the ORF3a variant NS3-G251 (clade V). Although sequence variation can give rise to contrasting phenotypes and hence possible differences in treatment regimes, its implications in SARS-CoV-2 is still unclear and widely debated today (Young et al., 2020). This is made even more complicated by its relationship with host factors of different countries, which have given rise to varied infection and mortality rates (Toyoshima et al., 2020). Since April 2020, the variants of SARS-CoV-2 genomes were found to be unevenly distributed and highly diversified throughout the continents, which raised the importance of adoptive measures for containing its outbreak (Jones & Manrique, 2020; Mercatelli & Giorgi, 2020). Interestingly, it has also been shown that the mutation rate of SARS-CoV-2 is low, at least in comparison to SARS CoV (Jia et al., 2020; Rausch et al., 2020). In our study of the phylogeny of SARS-CoV-2 isolates in Saudi Arabia (Figure 1), the genome sequences show very high similarity (between 97-98%). This may be beneficial because a uniformed drug treatment may be applicable to the current isolates. The genomic structure reflects that of the reference genome isolated from Wuhan (GenBank sequence accession NC_045512.2) (Table 1). Currently, more genomic sequences are being uploaded to GISAID, and this may increase the number of more diversified sequences and hence mutations that can render many available drugs ineffective (Naqvi et al., 2020; Pachetti et al., 2020). Despite this, a recent study by Vries et al (2020) has proven that clade differences do not have an impact on drugs that target highly conserved SARS-CoV-2 regions, such as 3CLPRO (Vries et al., 2020).

Developing drugs for neutralizing the infectivity of a novel pathogen in an outbreak itself is a lengthy endeavor. During a rapid outbreak that shows a high rate of infectivity and fatality, the window period of treatment becomes dangerously short. Hence, drug repurposing is an effective and economical strategy to immediately deliver the desired treatment outcomes in affected individuals with lesser
side effects. This is because the use of de-risked drugs bypasses most of the drug developmental hurdles. There are several approaches to this method. Apart from the traditional experimental approaches through in vitro and in vivo studies, an in silico approach is commonly performed due to its high throughput nature (R. Kumar et al., 2019). With continuous advancements in big data and computing technology, investigators now have an arsenal of analytical or machine learning tools to sift through large numbers of genomes, chemical structures, phenomes, and more to discover ideal drugs for novel targets (Jarada et al., 2020). In our study, we used multiple-sequence alignment, which is a rapid and simple structural-based approach in identifying potential drug candidates (March-Vila et al., 2017). Based on the data mined from ZINC, we have tabulated a list of 73 known viral therapeutic targets (Table 2). This includes those that share some similarities to the SARS-CoV-2 virus, for instance, the rhinovirus, which is a positive-sense, single-stranded RNA virus (þssRNA) much like SARS-CoV-2 (Pal et al., 2020). A large number of independent and clinical studies that are testing antiviral drugs for COVID-19 (e.g. remdesivir, favipiravir, and lopinavir) are being used commercially for their specificity to these known viral therapeutic targets (Gordon et al., 2020; Kang et al., 2020) (NCT04401579, NCT04358549, NCT04386876). Of these, we have generated a list of 29 viral orthologs that share sequence similarities to that of the SARS-CoV-2-SA isolates (Table 3). Each of these targets comes with information on purchasable drugs. Evidently, the replicase polyprotein 1a of SARS CoV, which also contains the conserved 3CL\textsuperscript{PRO} and PL\textsuperscript{PRO} sequences, had the highest similarity with SARS-CoV-2-SA (79.91%). A ZINC search revealed that tanshinones were among the highest binding affinity substances based on chembl 20. Tanshinones are a class of lipophilic phenanthrene compounds and are the main terpenoid bioactive components of Salvia miltiorrhiza, which is a dried root that is heralded for its therapeutic properties in traditional Chinese medicine (Jiang et al., 2019). It has been shown to activate the AMPK/mTOR signaling pathway, which led to apoptotic inhibition and autophagy induction in heart cells (X. Zhang et al., 2019). This involved the modulation of Bcl-2, Bax, caspase-3 and caspase-7. A study by Park et al (2012) has shown that these compounds are selective inhibitors of SARS CoV 3CL\textsuperscript{PRO} and PL\textsuperscript{PRO} cysteine proteases (Park et al., 2012), and this has been a recent subject of interest for its application in treating COVID-19 (Shahrajabian et al., 2020). These important viral proteins, along with RNA-dependent RNA polymerase (RdRp) and several others, are known to be highly conserved between the two human coronaviruses, including functional regions (C. Wu et al., 2020). Because of that, finding available drugs that target these regions have garnered significant attention (Báez-Santos et al., 2015). Indeed, several studies have already been done to identify such potential drug candidates that can bind to these viral proteins (Jo et al., 2020; Virdi et al., 2020). Tanshinones are currently being investigated in clinical trials as a potential treatment for acute myocardial infarction (NCT02524964), pulmonary hypertension (NCT01637675), and polycystic ovary syndrome (NCT01452477). Its therapeutic role in vascular diseases is partly due to its modulatory effect on vascular smooth muscle cell proliferation, which is a key contributor in arterial remodeling and hypertension (Wu et al., 2019). To date, there are no related clinical studies involving its use in coronavirus infections. However, a study by Zhang et al. (2020) has shown that cryptotanshinone, a tanshinone derivative, elicited a dose-dependent anti-viral effect on cells infected with SARS-CoV-2 in vitro. This proved that the compound does indeed hold tremendous potential.
Meanwhile, myricetin and morin are members of flavonoids, which have been shown by several studies to be chemical inhibitors of the SARS CoV 3CL\textsuperscript{PRO} and nsp3 proteins (Jo et al., 2020; Yu et al., 2012). Currently, flavonoids are being recommended as potential phytochemical-based medicines for COVID-19 treatment (Ngwa et al., 2020; Russo et al., 2020). Podofilox (podophyllotoxin), which is a human papilloma virus drug, may also be a potential candidate. However, a separate virtual screening done by Jordaan et al (2020) showed a low binding affinity of this drug to the SARS-CoV-2 protease (Jordaan et al., 2020). Podofilox is similar to the molecular scaffold of efavirenz, which is an HIV drug (Jordaan et al., 2020). Several of these anti-HIV protease drugs (e.g. atazanavir, ritonavir, darunavir, and dolutegravir) have been predicted to possess inhibitory potency against SARS-CoV-2 3CL\textsuperscript{PRO} in silico. An example of these can also be found in Table 3. Other notable mentions from this study are drugs targeting the influenza A virus (a negative-sense RNA virus) neuraminidase and matrix protein. Drugs against this virus (e.g. favipiravir, oseltamivir, etc) are currently in clinical trials and experimental treatments for COVID-19 patients (R. Wu et al., 2020) (NCT04464408, NCT04516915). Other influenza drugs like amantadine, are also being recommended as plausible treatments (Abreu et al., 2020; Araújo et al., 2020). In spite of the lower similarity indexes of other viral targets in our study, further studies should be done to investigate the potential effects of the listed drugs against these viruses. Perhaps, through other in silico methods such as molecular docking.

5.0 Conclusion

COVID-19 outbreak will not disappear in a short time and will add more to the rate of mortality and morbidity if appropriate intervention is not correctly applied. Due to the rapid human-to-human transmission of infection worldwide, finding an effective treatment or a functional cure is paramount to ensure our survival. Here, we employed multiple sequence alignment, which is a fast and simple approach in drug repurposing, to identify candidate drugs for COVID-19, with an emphasis on the isolates obtained from Saudi Arabia. Our study showed that these isolates show high sequence similarity (around 98%), and its genomic structure reflects that of the Wuhan reference isolate genome. By using multiple sequence alignment, we showed that among the list of viral target orthologs, the SARS CoV replicase polyprotein 1a sequence had the highest similarity (79.91%). Based on the ZINC database, tanshinones were among the highest compounds with binding affinity to the target, and it has been reported to selectively inhibit SARS CoV 3CL\textsuperscript{PRO} and PL\textsuperscript{PRO} cysteine proteases in the literature. These proteins have garnered significant attention in the discovery of compounds against SARS-CoV-2. However, tanshinones have yet to be investigated in current clinical trials for COVID-19, and thus, these compounds make potential drug candidates.

Acknowledgement

The authors extend their appreciation to the Deanship of Scientific Research at Jouf University for funding this work through research grant no. CV-15-41.

Author Contributions: M.P.L. and S.K. were responsible for conceiving the experimental study design and analyzing the data; A.E.H.K. performed the experiments and produced the figure and tables; M.P.L. and A.E.H.K. composed the manuscript; A.F. and A.A. analyzed and edited the manuscript. All authors were involved in reviewing the manuscript.

Conflicts of Interest

The authors declare that they have no conflict of interest.
References

Abreu, G. E. A., Aguilar, M. E. H., Covarrubias, D. H., & Durán, F. R. (2020). Amantadine as a drug to mitigate the effects of COVID-19. *Medical Hypotheses, 140*, 109755. https://doi.org/10.1016/j.mehy.2020.109755

Algaissi, A. A., Alharbi, N. K., Hassanain, M., & Hashem, A. M. (2020, June 1). Preparedness and response to COVID-19 in Saudi Arabia: Building on MERS experience. *Journal of Infection and Public Health*. Elsevier Ltd. https://doi.org/10.1016/j.jiph.2020.04.016

Araújo, R., Aranda-Martínez, J. D., & Aranda-Abreu, G. E. (2020). Amantadine Treatment for People with COVID-19. *Archives of Medical Research*. https://doi.org/10.1016/j.arcmed.2020.06.009

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., … Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research, 40*(W1). https://doi.org/10.1093/nar/gks400

Báez-Santos, Y. M., St. John, S. E., & Mesecar, A. D. (2015, March 1). The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antiviral Research*. Elsevier B.V. https://doi.org/10.1016/j.antiviral.2014.12.015

Barry, M., Al Amri, M., & Memish, Z. A. (2020). Covid-19 in the shadows of MERS-CoV in the Kingdom of Saudi Arabia. *Journal of Epidemiology and Global Health, 10*(1), 1–3. https://doi.org/10.2991/jegh.k.200218.003

Bharath, B. R., Damle, H., Ganju, S., & Damle, L. (2020). In silico screening of known small molecules to bind ACE2 specific RBD on Spike glycoprotein of SARS-CoV-2 for repurposing against COVID-19. *F1000Research, 9*. https://doi.org/10.12688/f1000research.24143.1

Chan, J. F. W., Li, K. S. M., To, K. K. W., Cheng, V. C. C., Chen, H., & Yuen, K. Y. (2012, December). Is the discovery of the novel human betacoronavirus 2c EMC/2012 (HCoV-EMC) the beginning of another SARS-like pandemic? *Journal of Infection*. J Infect. https://doi.org/10.1016/j.jinf.2012.10.002

De Vries, M., Mohamed, A. S., Prescott, R. A., Valero-Jimenez, A. M., Desvignes, L., O’connor, R., … Dittmann, M. (2020). Comparative study of a 3CL pro inhibitor and remdesivir against both major SARS-1 CoV-2 clades in human airway models 2 3. *BioRxiv*, 2020.08.28.272880. https://doi.org/10.1101/2020.08.28.272880

Ebrahim, S. H., & Memish, Z. A. (2020, March 14). COVID-19: preparing for superspreader potential among Umrah pilgrims to Saudi Arabia. *The Lancet*. Lancet Publishing Group. https://doi.org/10.1016/S0140-6736(20)30466-9

Fogel, D. B. (2018, September 1). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communications*. Elsevier Inc. https://doi.org/10.1016/j.conctc.2018.08.001

Gordon, C. J., Tchesnokov, E. P., Woolner, E., Perry, J. K., Feng, J. Y., Porter, D. P., & Götte, M. (2020, May 15). Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc. https://doi.org/10.1074/jbc.RA120.013679

Hwang, T. J., Carpenter, D., Lauffenburger, J. C., Wang, B., Franklin, J. M., & Kesselheim, A. S. (2016). Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Internal Medicine, 176*(12), 1826–1833. https://doi.org/10.1001/jamainternmed.2016.6008

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012, July 23). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*. American Chemical Society. https://doi.org/10.1021/ci3001277

Jarada, T. N., Rokne, J. G., & Alhajj, R. (2020, July 22). A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. *Journal of Cheminformatics*. BioMed Central. https://doi.org/10.1186/s13321-020-00450-7

Jia, Y., Shen, G., Zhang, Y., Huang, K.-S., Ho, H.-Y., Hor, W.-S., … Wang, W.-L. (2020). Analysis of the
mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. BioRxiv, 2020.04.09.034942. https://doi.org/10.1101/2020.04.09.034942

Jiang, Z., Gao, W., & Huang, L. (2019, March 14). Tanshinones, critical pharmacological components in salvia miltiorrhiza. Frontiers in Pharmacology. Frontiers Media S.A. https://doi.org/10.3389/fphar.2019.00202

Jo, S., Kim, S., Kim, D. Y., Kim, M. S., & Shin, D. H. (2020). Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro. Journal of Enzyme Inhibition and Medicinal Chemistry, 35(1), 1539–1544. https://doi.org/10.1080/14756366.2020.1801672

Jo, S., Kim, S., Shin, D. H., & Kim, M. S. (2020). Inhibition of SARS-CoV 3CL protease by flavonoids. Journal of Enzyme Inhibition and Medicinal Chemistry, 35(1), 145–151. https://doi.org/10.1080/14756366.2019.1690480

Jones, L. R., & Manrique, J. M. (2020). Quantitative phylogenomic evidence reveals a spatially structured SARS-CoV-2 diversity. Virology, 550, 70–77. https://doi.org/10.1016/j.virol.2020.08.010

Jordaan, M. A., Ebenezer, O., Damoyi, N., & Shapi, M. (2020). Virtual screening, molecular docking studies and DFT calculations of FDA approved compounds similar to the non-nucleoside reverse transcriptase inhibitor (NNRTI) efavirenz. Heliyon, 6(8), e04642. https://doi.org/10.1016/j.heliyon.2020.e04642

Kandeel, M., Ibrahim, A., Fayez, M., & Al-Nazawi, M. (2020). From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes. Journal of Medical Virology, 92(6), 660–666. https://doi.org/10.1002/jmv.25754

Kang, C. K., Seong, M. W., Choi, S. J., Kim, T. S., Choe, P. G., Song, S. H., ... Oh, M. D. (2020). In vitro activity of lopinavir/ritonavir and hydroxychloroquine against severe acute respiratory syndrome coronavirus 2 at concentrations achievable by usual doses. The Korean Journal of Internal Medicine, 35(4), 782–787. https://doi.org/10.3904/kijm.2020.157

Kelleni, M. T. (2020, July 1). Nitazoxanide/azithromycin combination for COVID-19: A suggested new protocol for early management. Pharmacological Research. Academic Press. https://doi.org/10.1016/j.phrs.2020.104874

Khan, N. T. (2017). MEGA - Core of Phylogenetic Analysis in Molecular Evolutionary Genetics. https://doi.org/10.4172/2329-9002.1000183

Kumar, R., Harilal, S., Gupta, S. V., Jose, J., Thomas, D. G., Uddin, M. S., ... Mathew, B. (2019, November 15). Exploring the new horizons of drug repurposing: A vital tool for turning hard work into smart work. European Journal of Medicinal Chemistry. Elsevier Masson SAS. https://doi.org/10.1016/j.ejmech.2019.111602

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution, 35(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

March-Vila, E., Pinzi, L., Sturm, N., Tinivella, A., Engkvist, O., Chen, H., & Rastelli, G. (2017). On the Integration of In Silico Drug Design Methods for Drug Repurposing. Frontiers in Pharmacology, 8(MAY), 298. https://doi.org/10.3389/fphar.2017.00298

Memish, Z. A., Aljerian, N., & Ebrahim, S. H. (2020). Tale of three seeding patterns of SARS-CoV-2 in Saudi Arabia. The Lancet Infectious Diseases. Lancet Publishing Group. https://doi.org/10.1016/S1473-3099(20)30425-4

Meo, S. A., Klonoff, D. C., & Akram, J. (2020). Efficacy of chloroquine and hydroxychloroquine in the treatment of COVID-19. European Review for Medical and Pharmacological Sciences. Verduci Editore s.r.l. https://doi.org/10.26355/eurrev_202004_21038

Mercatelli, D., & Giorgi, F. M. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Frontiers in Microbiology, 11, 1800. https://doi.org/10.3389/fmicb.2020.01800

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., ... Hassan, M. I. (2020, October 1). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies:
Structural genomics approach. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier B.V. https://doi.org/10.1016/j.bbadis.2020.165878

Newman, L., Duffus, A. L. J., & Lee, C. (2016). Using the free program MEGA to build phylogenetic trees from molecular data. *American Biology Teacher, 78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

Ngwa, W., Kumar, R., Thompson, D., Lyerly, W., Moore, R., Reid, T. E., ... Toyang, N. (2020). Potential of Flavonoid-Inspired Phytomedicines against COVID-19. *Molecules (Basel, Switzerland), 25*(11). https://doi.org/10.3390/molecules25112707

Obied, D. A., Alhamlan, F. S., Al-Qahtani, A. A., & Al-Ahdal, M. N. (2020). Containment of COVID-19: the unprecedented response of Saudi Arabia. *Journal of Infection in Developing Countries, 14*(7), 699–706. https://doi.org/10.3855/jidc.13203

Ou-Yang, S. S., Lu, J. Y., Kong, X. Q., Liang, Z. J., Luo, C., & Jiang, H. (2012, September 27). Computational drug discovery. *Acta Pharmacologica Sinica*. Nature Publishing Group. https://doi.org/10.1038/aps.2012.109

Pal, M., Berhanu, G., Desalegn, C., & Kandi, V. (2020). Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. *Cureus, 12*(3). https://doi.org/10.7759/cureus.7423

Park, J. Y., Kim, J. H., Kim, Y. M., Jeong, H. J., Kim, D. W., Park, K. H., ... Ryu, Y. B. (2012). Tanshinones as selective and slow-binding inhibitors for SARS-CoV cysteine proteases. *Bioorganic and Medicinal Chemistry, 20*(19), 5928–5935. https://doi.org/10.1016/j.bmc.2012.07.038

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., ... Pirmohamed, M. (2018, December 28). Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery*. Nature Publishing Group. https://doi.org/10.1038/nrd.2018.168

Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C., & Kearney, M. F. (2020). Low genetic diversity may be an Achilles heel of SARS-CoV-2. *Proceedings of the National Academy of Sciences, 202017726*. https://doi.org/10.1073/pnas.202017726

Russo, M., Moccia, S., Spagnuolo, C., Tedesco, I., & Russo, G. L. (2020, September 1). Roles of flavonoids against coronavirus infection. *Chemico-Biological Interactions*. Elsevier Ireland Ltd. https://doi.org/10.1016/j.cbi.2020.109211

Shahrajabian, M. H., Sun, W., Shen, H., & Cheng, Q. (2020). Chinese herbal medicine for SARS and SARS-CoV-2 treatment and prevention, encouraging using herbal medicine for COVID-19 outbreak. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 70*(5), 437–443. https://doi.org/10.1080/09064710.2020.1763448

Shu, Y., & McCauley, J. (2017, March 30). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance. European Centre for Disease Prevention and Control (ECDC).* https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Song, Y., Zhang, M., Yin, L., Wang, K., Zhou, Y., Zhou, M., & Lu, Y. (2020, August 1). COVID-19 treatment: close to a cure? A rapid review of pharmacotherapies for the novel coronavirus (SARS-CoV-2). *International Journal of Antimicrobial Agents*. Elsevier B.V. https://doi.org/10.1016/j.ijantimicag.2020.106080

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution, 10*(3), 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., & Kiyotani, K. (2020). SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics, 1–8*. https://doi.org/10.1038/s10038-020-0808-9

van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T. M., Raj, V. S., Zaki, A. M., ... Fouchier, R. A. M. (2012). Genomic characterization of a newly discovered coronavirus associated with acute
respiratory distress syndrome in humans. *MBio*, 3(6). https://doi.org/10.1128/mBio.00473-12
Virdi, R. S., Bavisotto, R. V., Hopper, N. C., & Frick, D. N. (2020). Discovery of Drug-like Ligands for the Mac1 Domain of SARS-CoV-2 Nsp3. *BioRxiv: The Preprint Server for Biology*. https://doi.org/10.1101/2020.07.06.190413
Wang, S., Sundaram, J. P., & Spiro, D. (2010). VIGOR, an annotation program for small viral genomes. *BMC Bioinformatics*, 11, 451. https://doi.org/10.1186/1471-2105-11-451
Worldometer. (n.d.). Saudi Arabia Coronavirus: 291,468 Cases and 3,233 Deaths - Worldometer. Retrieved September 10, 2020, from https://www.worldometers.info/coronavirus/country/saudi-arabia/
Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., … Li, H. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B*, 10(5), 766–788. https://doi.org/10.1016/j.apsb.2020.02.008
Wu, R., Wang, L., Kuo, H. C. D., Shannar, A., Peter, R., Chou, P. J., … Kong, A. N. (2020, June 1). An Update on Current Therapeutic Drugs Treating COVID-19. *Current Pharmacology Reports*. Springer. https://doi.org/10.1007/s40495-020-00216-7
Wu, Y. ting, Bi, Y. M., Tan, Z. Bin, Xie, L. P., Xu, H. L., Fan, H. J., … Zhou, Y. C. (2019). Tanshinone I inhibits vascular smooth muscle cell proliferation by targeting insulin-like growth factor-1 receptor/phosphatidylinositol-3-kinase signaling pathway. *European Journal of Pharmacology*, 853, 93–102. https://doi.org/10.1016/j.ejphar.2019.03.021
Young, B. E., Fong, S.-W., Chan, Y.-H., Mak, T.-M., Ang, L. W., Anderson, D. E., … Ng, L. F. P. (2020). Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet*, 396(10251), 603. https://doi.org/10.1016/s0140-6736(20)31757-8
Yu, M. S., Lee, J., Lee, J. M., Kim, Y., Chin, Y. W., Jee, J. G., … Jeong, Y. J. (2012). Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsP13. *Bioorganic and Medicinal Chemistry Letters*, 22(12), 4049–4054. https://doi.org/10.1016/j.bmcl.2012.04.081
Zhang, X., Wang, Q., Wang, X., Chen, X., Shao, M., Zhang, Q., … Wang, Y. (2019). Tanshinone IIA protects against heart failure post-myocardial infarction via AMPKs/mTOR-dependent autophagy pathway. *Biomedicine & Pharmacotherapy*, 112, 108599. https://doi.org/10.1016/j.biopha.2019.108599
Zhang, Z. R., Zhang, Y. N., Li, X. D., Zhang, H. Q., Xiao, S. Q., Deng, F., … Zhang, B. (2020, December 1). A cell-based large-scale screening of natural compounds for inhibitors of SARS-CoV-2. *Signal Transduction and Targeted Therapy*. Springer Nature. https://doi.org/10.1038/s41392-020-00343-z
Figure and Figure Legend

Figure 1. The phylogenetic tree of SARS-CoV-2 isolates from Saudi Arabia. The tree shows the evolutionary relationship of all 164 SARS-CoV-2-SA isolates obtained from the GISAID database. The relationship shows a high similarity between the tested samples (97-98%). The figure was generated using the Maximum Likelihood method and Tamura-Nei model in MEGA X.

Table 1. SARS-CoV-2-SA genomic construction using the Viral Genome ORF Reader, VIGOR4 annotator. The isolate sequence was previously submitted to GISAID database by King Abdullah International Medical Research Center (KAIMRC). The genome consists of ORF1a, ORF1b, Spike, Envelope, Membrane, ORF6, ORF7a, ORF7b, ORF8, Nucleocapsid, and ORF10, respectively. The ORF regions code for the accessory proteins and non-structural proteins, which are involved in virus replication and assembly.

Table 2. List of targeted viral genes and the total number of its acting compounds. Using ZINC database. A total of 73 viral therapeutic targets were found. Each target includes its protein function, subclass, orthologs, as well as other information. ‘Observations’ refers to the number of individual reports on compounds that were tested on their respective target. ‘Substances’ refers to the number of compounds for the target. ‘Purchasable’ refers to commercially available compounds, and finally ‘Predicted’ shows the similarity ensemble approach (SEA) predictions-based candidate compounds based on chembl 20.

Table 3. Percent identity matrix of therapeutic targets mapped to the SARS-CoV-2-SA isolate sequence. Of the 73 genes, the orthologs with known purchasable drugs were mapped to the SARS-CoV-2 KSA isolate sequence by multiple sequence alignment using CLUSTAL OMEGA. A list of 29 orthologs, their respective similarity index, and purchasable drug information was then generated. Compounds binding to targets with higher similarity to SARS-CoV-2 could be potential drug candidates for further study.

| Start..Stop | Gene    | Gene Product Name            | Reference   | Peptide Length | Ref. Length | % Identity | % Similarity | % Coverage |
|-------------|---------|------------------------------|-------------|----------------|-------------|------------|-------------|------------|
| 272..13489  | ORF1a   | ORF1a polyprotein            | YP_009725295.1 | 4405           | 4405        | 99.95      | 99.95       | 100.00     |
| 13448..13486| ORF1a   | nsp11                        | YP_009725312.1 | 13             | 13          | 100.00     | 100.00      | 100.00     |
| 272..13474, 13474..21561 | ORF1ab | ORF1ab polyprotein          | YP_009724389.1 | 7096           | 7096        | 99.96      | 99.96       | 100.00     |
| 272..811    | ORF1ab  | leader protein               | YP_009725297.1 | 180            | 180         | 100.00     | 100.00      | 100.00     |
| ORF1ab  | nsp2  | YP_009725298.1 | 638 | 638 | 100.00 | 100.00 | 100.00 |
|---------|-------|---------------|-----|-----|--------|--------|--------|
| 812..2725 | ORF1ab nsp2 | YP_009725298.1 | 638 | 638 | 100.00 | 100.00 | 100.00 |
| 2726..8560 | ORF1ab nsp3 | YP_009725299.1 | 1945 | 1945 | 99.95 | 99.95 | 100.00 |
| 8561..10060 | ORF1ab nsp4 | YP_009725300.1 | 500 | 500 | 100.00 | 100.00 | 100.00 |
| 10061..10978 | 3C-like proteinase | YP_009725301.1 | 306 | 306 | 100.00 | 100.00 | 100.00 |
| 10979..11848 | ORF1ab nsp6 | YP_009725302.1 | 290 | 290 | 99.66 | 100.00 | 100.00 |
| 11849..12097 | ORF1ab nsp7 | YP_009725303.1 | 83 | 83 | 100.00 | 100.00 | 100.00 |
| 12098..12691 | ORF1ab nsp8 | YP_009725304.1 | 198 | 198 | 100.00 | 100.00 | 100.00 |
| 12692..13030 | ORF1ab nsp9 | YP_009725305.1 | 113 | 113 | 100.00 | 100.00 | 100.00 |
| 13031..13447 | ORF1ab nsp10 | YP_009725306.1 | 139 | 139 | 100.00 | 100.00 | 100.00 |
| 13448..16243 | ORF1ab RNA-dependent RNA polymerase | YP_009725307.1 | 932 | 932 | 99.89 | 99.89 | 100.00 |
| 16244..18046 | ORF1ab helicase | YP_009725308.1 | 601 | 601 | 100.00 | 100.00 | 100.00 |
| 18047..19627 | ORF1ab 3'-to-5' exonuclease | YP_009725309.1 | 527 | 527 | 100.00 | 100.00 | 100.00 |
| 19628..20665 | ORF1ab endoRNAse | YP_009725310.1 | 346 | 346 | 100.00 | 100.00 | 100.00 |
| 20666..21559 | ORF1ab 2'-O-ribose methyltransferase | YP_009725311.1 | 298 | 298 | 100.00 | 100.00 | 100.00 |
| 21569..25390 | S surface glycoprotein | YP_009724390.1 | 1273 | 1273 | 100.00 | 100.00 | 100.00 |
| 25399..26226 | ORF3a ORF3a protein | YP_009724391.1 | 275 | 275 | 99.64 | 100.00 | 100.00 |
| 26251..26478 | E envelope protein | YP_009724392.1 | 75 | 75 | 100.00 | 100.00 | 100.00 |
| 26529..27197 | M membrane glycoprotein | YP_009724393.1 | 222 | 222 | 100.00 | 100.00 | 100.00 |
| 27208..27393 | ORF6 ORF6 protein | YP_009724394.1 | 61 | 61 | 100.00 | 100.00 | 100.00 |
| 27400..27765 | ORF7a ORF7a protein | YP_009724395.1 | 121 | 121 | 100.00 | 100.00 | 100.00 |
| 27762..27893 | ORF7b ORF7b protein | YP_009725296.1 | 43 | 43 | 100.00 | 100.00 | 100.00 |
| 27900..28265 | ORF8 ORF8 protein | YP_009724396.1 | 121 | 121 | 100.00 | 100.00 | 100.00 |
| 28280..29539 | N nucleocapsid phosphoprotein | YP_009724397.2 | 419 | 419 | 99.76 | 99.76 | 100.00 |
| 29564..29680 | ORF10 ORF10 protein | YP_009725255.1 | 38 | 38 | 100.00 | 100.00 | 100.00 |
Table 2. List of targeted viral genes and the total number of its acting compounds. Using ZINC database, a total of 73 viral therapeutic targets were found. Each target includes its protein function, subclass, orthologs, as well as other information. ‘Observations’ refers to the number of individual reports on compounds that were tested on their respective target. ‘Substances’ refers to the number of compounds for the target. ‘Purchasable’ refers to commercially available compounds, and finally ‘Predicted’ shows the similarity ensemble approach (SEA) predictions-based candidate compounds based on chembl 20.

| Name             | Description                              | Sub Class    | Orthologs | Observations | Substances | Purchasable | Predicted |
|------------------|------------------------------------------|--------------|-----------|--------------|------------|-------------|-----------|
| NA               | Neuraminidase                             | hydrolyase   | 14        | 977          | 480        | 20          | 23041     |
| M                | Matrix protein 2                          | IC-other     | 1         | 32           | 32         | 12          | 1025      |
| DPOL_HHV11       | DNA polymerase catalytic subunit          | transferase  | 1         | 6            | 6          | 4           | 31646     |
| TK               | Thymidine kinase                          | enzyme-other | 5         | 302          | 147        | 35          | 20453     |
| TAT_HV112        | Protein Tat                               | TF-other     | 1         | 58           | 49         | 3           | 3374      |
| NS4A             | Non-structural protein 4A                 | protease     | 1         | 1            | 1          | 0           | 4879      |
| UL80             | Capsid scaffolding protein                | protease     | 1         | 208          | 182        | 4           | 176053    |
| RIR1_HHV11       | Ribonucleoside-diphosphate reductase large subunit | enzyme-other | 1         | 30           | 26         | 0           | 88152     |
| E2               | Regulatory protein E2                     | TF-other     | 3         | 1            | 1          | 1           | 2429      |
| UL54             | DNA polymerase catalytic subunit          | enzyme-other | 1         | 87           | 80         | 4           | 148906    |
| KITH_VZVD        | Thymidine kinase                          | enzyme-other | 1         | 9            | 9          | 2           | 697       |
| NS5B             | NS5B protein                              | transferase  | 2         | 1186         | 925        | 56          | 213055    |
| NS3              | Genome polyprotein                        | protease     | 1         | 980          | 927        | 47          | 64711     |
| UL26             | Capsid scaffolding protein                | protease     | 2         | 55           | 55         | 1           | 82085     |
| POLG_HCV1        | Genome polyprotein                        | protease     | 1         | 96           | 95         | 2           | 75703     |
| PROTEASE         | Protease                                 | protease     | 2         | 953          | 818        | 51          | 158148    |
| TAT              | Protein Tat                               | TF-other     | 2         | 12           | 12         | 0           | 7728      |
| REV              | Protein Rev                              | cytosolic-other | 1     | 19           | 15         | 1           | 3001      |
| ENV              | Envelope glycoprotein gp160              | surface-antigen | 5     | 55           | 51         | 21          | 1590740   |
| ORF_36           | Thymidine kinase                          | enzyme-other | 1         | 0            | 0          | 0           | 0         |
| U38              | DNA polymerase catalytic subunit          | enzyme-other | 1         | 12           | 12         | 2           | 243208    |
| U53              | Capsid scaffolding protein                | protease     | 1         | 33           | 33         | 0           | 28153     |
| SCAF_EBV9        | Capsid scaffolding protein                | protease     | 1         | 28           | 28         | 0           | 72426     |
| POL              | Pol polyprotein                           | enzyme-other | 6         | 12880        | 9001       | 876         | 799012    |
| Gene Name                  | Protein Description                                      | Function     | Other 1 | Other 2 | Other 3 | Other 4 | Protein Length |
|---------------------------|----------------------------------------------------------|--------------|---------|---------|---------|---------|----------------|
| GAG-PRO-POL               | Gag-Pro-Pol polyprotein                                   | enzyme-other | 1       | 3       | 3       | 2       | 494            |
| POLG_POL1M                | Genome polyprotein                                        | protease     | 1       | 1       | 1       | 0       | 3037           |
| R1A_CVHS                 | Replicase polyprotein 1a                                  | enzyme-other | 1       | 40      | 35      | 13      | 11327          |
| GAG-PRO                   | Gag-Pro polyprotein                                       | protease     | 1       | 33      | 33      | 0       | 62841          |
| HA                       | Hemagglutinin                                             | surface-antigen | 3       | 8       | 6       | 0       | 1051           |
| ABL                      | Tyrosine-protein kinase transforming protein Abl         | kinase       | 1       | 1       | 1       | 0       | 10324          |
| E                        | Endolysin                                                 | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| E1                       | Replication protein E1                                    | enzyme-other | 1       | 32      | 31      | 27      | 233548         |
| PRIM_HHV11                | DNA primase                                               | nuclear-other | 1       | 7       | 7       | 3       | 1929           |
| VP16_HHV11                | Tegument protein VP16                                     | nuclear-other | 1       | 2       | 2       | 1       | 55             |
| US28                      | G-protein coupled receptor homolog US28                   | GPCR-A       | 2       | 46      | 14      | 3       | 4752           |
| DPOL_VZVD                 | DNA polymerase catalytic subunit                          | enzyme-other | 1       | 8       | 7       | 5       | 64879          |
| MC087R                    | MC087R                                                   | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| POLG_BVDVC                | Genome polyprotein                                        | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| Q86831_AVIMB              | Polypeptide II                                            | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| POLG_GBVB                 | Genome polyprotein                                        | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| POLG_DEN26                | Genome polyprotein                                        | enzyme-other | 1       | 11      | 10      | 0       | 13876          |
| PSET                      | Polynucleotide kinase                                     | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| POLG_WNV                  | Genome polyprotein                                        | enzyme-other | 1       | 71      | 52      | 5       | 191900         |
| POLG_HCVCO                | Genome polyprotein                                        | enzyme-other | 1       | 34      | 34      | 2       | 707            |
| REP                      | Replicase polyprotein 1ab                                 | enzyme-other | 1       | 9       | 9       | 3       | 9189           |
| POLG_HRV16                | Genome polyprotein                                        | enzyme-other | 1       | 5       | 5       | 1       | 4352           |
| Q82323_9DELA              | Protease                                                  | unclassified | 1       | 20      | 20      | 0       | 76070          |
| POLG_HCVBK                | Genome polyprotein                                        | enzyme-other | 1       | 7       | 7       | 0       | 2591           |
| V-FPS                     | Tyrosine-protein kinase transforming protein Fps          | kinase       | 1       | 1       | 1       | 1       | 0              |
| 30                        | DNA ligase                                                | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| GAG                       | Gag polyprotein                                           | unclassified | 2       | 15      | 15      | 3       | 1050           |
| DPOL_HHV1K                | DNA polymerase catalytic subunit                          | enzyme-other | 1       | 4       | 4       | 2       | 869            |
| 43                        | DNA polymerase                                            | enzyme-other | 1       | 0       | 0       | 0       | 0              |
| G                        | Glycoprotein G                                            | unclassified | 1       | 4       | 4       | 0       | 17091          |
| Gene/Protein ID | Gene/Protein Name                          | Biochemical Category | GeneID | MW      | pI      | GeneID | MW      |
|----------------|-------------------------------------------|----------------------|--------|---------|---------|--------|---------|
| UL23           | Thymidine kinase                          | unclassified         | 1      | 8       | 8       | 6      | 4926    |
| RPOL_BPT7      | T7 RNA polymerase                         | enzyme-other         | 1      | 0       | 0       | 0      | 0       |
| HBCAG          | External core antigen                     | unclassified         | 1      | 0       | 0       | 0      | 0       |
| THYX_PBCV1     | Probable thymidylate synthase             | enzyme-other         | 1      | 10      | 10      | 0      | 4316    |
| N1L            | Virokine/NFkB inhibitor                    | unclassified         | 1      | 22      | 12      | 8      | 31528   |
| PA             | Polymerase acidic protein                 | unclassified         | 2      | 94      | 51      | 9      | 36557   |
| Q3ZDS5_9HEPC   | N55B                                      | unclassified         | 1      | 6       | 2       | 0      | 2097    |
| UL97           | Phosphotransferase pUL97                  | unclassified         | 2      | 294     | 159     | 23     | 24831   |
| HIVRT          | Reverse transcriptase                     | unclassified         | 1      | 1       | 1       | 0      | 557     |
| M2             | Matrix protein 2                          | enzyme-other         | 1      | 13      | 13      | 3      | 533     |
| UNG            | Uracil-DNA glycosylase                    | enzyme-other         | 2      | 4       | 4       | 1      | 1259    |
| E6             | Protein E6                                | unclassified         | 1      | 2       | 2       | 2      | 0       |
| Q76353_9HIV1   | Integrase                                 | unclassified         | 1      | 153     | 146     | 2      | 63112   |
| POLG_CXB3N     | Genome polyprotein                        | enzyme-other         | 1      | 1       | 1       | 0      | 171     |
| R1A_CVHNL      | Replicase polyprotein 1a                  | enzyme-other         | 1      | 0       | 0       | 0      | 0       |
| PB2            | Polymerase basic protein 2                | unclassified         | 1      | 31      | 31      | 0      | 1789    |
| A0A0K1CY61_9HEPC | Nonstructural protein NS3-4A        | unclassified         | 1      | 59      | 47      | 6      | 996     |
| HIV1_ENV       | GP41                                      | unclassified         | 1      | 1       | 1       | 1      | 35089   |
| Q91H74_9FLAV   | Genome polyprotein                        | unclassified         | 1      | 30      | 25      | 5      | 24580   |
Table 3. Percent identity matrix of therapeutic targets mapped to the SARS-CoV-2-SA isolate sequence. Of the 73 genes, the orthologs with known purchasable drugs were mapped to the SARS-CoV-2-SA isolate sequence by multiple alignment using CLUSTAL OMEGA. A list of 29 orthologs, their respective similarity index, and purchasable drugs information was then generated. Compounds binding to targets with higher similarity to SARS-CoV-2 could be potential drug candidates for further study.

| Name        | Target description | Orthologs | Ortholog ID          | Virus                | %       | Purchasable | Examples                                      |
|-------------|--------------------|-----------|----------------------|----------------------|---------|-------------|-----------------------------------------------|
| REP         | Replicase polyprotein 1a | 1         | R1A_CVHSA            | SARS-CoV            | 79.91   | 13          | Tanshinone                                    |
| E6          | Protein E6         | 1         | VE6_HPV16            | Human papilloma virus | 51.99   | 2           | Myricetin, Morin, Oseltamivir, Zanamivir, Rapivab |
| NA          | Neuraminidase      | 14        | NRAM_I34A1           | Influenza A virus    | 47.76   | 20          |                                              |
| E2          | Regulatory protein E2 | 3         | VE2_HPV16            | Human papilloma virus | 47.61   | 1           | Podoflox                                      |
| M           | Matrix protein 2   | 1         | M2_I72A2             | Influenza A virus    | 45.09   | 12          | Amantadine, Ramantadine, Ciuprevir, Victrelis |
| NS3         | Genome polyprotein | 1         | A3EZI9_9HEPC         | Hepatitis C virus    | 40.55   | 47          |                                              |
| POLG_WNV    | Genome polyprotein | 1         | POLG_WNV             | West Nile virus      | 40.09   | 5           | ZINC3249673                                  |
| Q76353_9HIV1| Integrase          | 1         | Q76353_9HIV1         | Human immunodeficiency virus 1 | 39.19 | 2 | Raltegravir |
| UL26        | DNA polymerase catalytic subunit | 2         | SCAF_HHV11          | Human herpesvirus 1  | 38.26   | 1           | ZINC3625576                                  |
| TAT_HV112   | Tat protein        | 1         | TAT_HV112            | Human immunodeficiency virus 1 | 37.9  | 1           | ZINC5155                                     |
| ENV         | Envelope polyprotein GP160 | 5         | ENV_HV1H2            | Human immunodeficiency virus 1 | 37.7  | 21          | ZINC1780082                                  |
| UL54        | DNA polymerase     | 1         | DPOL_HCMVA           | Human cytomegalovirus | 35.79 | 4           | Foscarinet                                   |
| UL80        | Capsid scaffolding protein | 1         | SCAF_HCMVA           | Human cytomegalovirus | 35.69 | 4           | ZINC901466                                   |
| POL         | Pol polyprotein    | 6         | P88142_9HIV2         | Human immunodeficiency virus 2 | 35.67 | 876 | Elvicitabine, Trovirdine, Efavirenz, Intelence |
| HIVRT       | Reverse transcriptase | 2         | Q06347_9HIV2        | Human immunodeficiency virus 2 | 35.48 | 23          | Pleconaril, Pirodavir                        |
| PROTEASE    | Protease           | 2         | Q4U254_9ENTO         | Human rhinovirus     | 35.28   | 51          |                                              |
| Gene ID | Protein Name                          | Gene ID | Protein Name                          | Organism            | IC50 (nM) | ZINC Code                  |
|---------|---------------------------------------|---------|---------------------------------------|---------------------|-----------|-----------------------------|
| PRIM_HHV11 | DNA primase                          | 1       | PRIM_HHV11                            | Human herpesvirus 1  | 35.23     | ZINC1675992                 |
| UL26    | Capsid scaffolding protein            | 2       | SCAF_HHV11                            | Human herpesvirus 1  | 35.05     | ZINC3625576                 |
| DPOL_VZVD | DNA polymerase catalytic subunit      | 1       | DPOL_VZVD                             | Varicella-zoster virus | 34.45     | Aphidicolin, Foscarnet      |
| E1      | Replication protein E1                | 1       | VE1_HPV11                             | Human papilloma virus | 34.39     | ZINC3600349                 |
| DPOL_HHV11 | DNA polymerase catalytic subunit     | 1       | DPOL_HHV11                            | Human herpesvirus 1  | 34.33     | Aphidicolin                 |
| GAG     | Gag polyprotein                       | 2       | GAG_AVIER                             | Avian erythroblastosis virus | 34.33     | ZINC6584476                 |
| PA      | Polymerase acidic protein             | 2       | PA_I34A1                              | Influenza A virus    | 34.02     | ZINC3626195                 |
| POLG_HRV16 | Genome polyprotein                   | 1       | POLG_HRV16                            | Human rhinovirus     | 33.87     | ZINC40975895                |
| UNG     | Uracil-DNA                            | 2       | UNG_VACCW                             | Vaccinia virus       | 33.82     | ZINC359756                  |
| US28    | G-protein coupled receptor homolog US28 | 2   | US28_HCMVA                            | Human cytomegalovirus | 30.89     | Metitepine                  |
| VP16_HHV11 | Tegument protein VP16               | 1       | VP16_HHV11                            | Human herpesvirus 1  | 30.84     | ZINC3831128                 |
| TK      | Thymidine kinase                      | 5       | KITH_HHV1S                            | Human herpesvirus 1  | 28.11     | Brivudine, Sorivudine       |
| N1L     | Virokine/NFkB inhibitor               | 1       | Q49PX0_9POXV                          | Vaccinia virus       | 27.78     | ZINC1557545                 |
Conflicts of Interest
The authors declare that they have no conflict of interest.