Berkala **PENELITIAN HAYATI**

ISSN: 08526834 | E-ISSN:2337-389X

Volume 26| No. 2| June | 2021

A review

Wiping out the covid-19 pandemic through bioinformatics: a review on database and web tools applications

Nur Imaniati Sumantri[1], Nagisa Eremia Anju[1], Siti Fauziyah Rahman[1,2]

[1]Biomedical Engineering Study Program, Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia, 16424

[2]Research Center for Biomedical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia, 16424

Abstract

The COVID-19 is an illness caused by SARS-CoV-2 that has ended up a widespread since March 2020 as declared by WHO. This condition gives a huge impact on miscellaneous sectors enforcing advance use of technology and information system, especially in scientific and medical community. Bioinformatics as a multidisciplinary method plays important role to overcome the COVID-19 outbreak in the early-stage through data exchange in virtual databases. There are a number of free access databases containing basic to complex information of SARS-CoV-2, such as genetic data sequence, epidemiology, evolutionary analysis, pharmacology, and so on. Bioinformatics allows us to analyze the data further to reveal new information applied in biomedical technology activities. Thus, bioinformatics helps the scientists, clinicians, and government learn the genomic characteristics of SARS-CoV-2, to trace the SARS-CoV-2 spread, and select and develop biomarker for reliable diagnostic tools, and design the drug and vaccine for SARS-CoV-2. This review aims to view insights on uses of bioinformatics methods and the databases related to SARS-CoV-2

Keywords:SARS-CoV-2, sequence analysis, vaccine design, web-tools, databases.

Received: 12 December 2020 Revised: 18 February 2021 Accepted: 23 February 2021

Introduction

The new coronavirus identified as a severe acute respiratory syndrome (SARS-CoV-2) emerged in December 2019 in Wuhan, China. It has spread globally, causing a recent worldwide pandemic disease called coronavirus disease 2019 (COVID-19). As of 18 November 2020, the SARS-CoV-2 reaches 55,064,128 positive cases with 1,328,015 deaths worldwide (https://www.who.int/ emergencies/diseases/novel-coronavirus-2019). Coronaviruses possess a positive single-stranded RNA genome that comprises the largest genomes of RNA viruses and belongs to the Coronaviridaee family (CoVs) (Pal *et al*., 2020; Martinelli *et al*., 2020). Based on their antigenic relationships, SARS-CoV-2 were divided into four genera, *Alphacoronavirus* (α), *Betacoronavirus* (β), *Gammacoronavirus* (γ), and *Deltacoronavirus* (δ) (Soremekun *et al.,* 2020). MERS-CoV, SARS-CoV, and SARS-CoV-2 are member of *Betacoronaviruses* that are responsible for deadly respiratory infections. The virus was introduced zoonotically into human population with mild infection similar with flu symptoms (Tilocca *et al*., 2020). SARS-CoV-2 has been reported as being passes on through the close contact respiratory droplets from human to human and aerial droplets on surfaces (Soremekun *et al*., 2020).

To date, appropriate diagnostic tools and effective treatment for SARS-CoV-2 are not available. There are specifications to comply with when designing diagnostic tools to prevent SARS-CoV-2 spread, they are (a) able to diagnose viral infections rapidly, effectively, and reliably at the pre-symptomatic level.; (b) efficiently target the human immune system to avoid irreversible negative impacts on human health; (c) capable of identifying immunogenic pathways and bioactive compounds (Martinelli *et al*., 2020). Besides the development of tools and treatment, immunoinformatics approaches have also been employed to discover epitopes for vaccine design. Immunoinformatic is one of the foremost promising approaches to analyze the antigenic determinants of novel and unknown molecules. It uses computational resources (e.g. program, algorithms, and statistic models) to investigate protein sequences and to foresee epitopes (Ansori *et al*., 2020; Soremekun *et al*., 2020). The application of immunoinformatics gives better perceptive on host-pathogen interaction along with immune system response and allows the researchers to do vaccine design that become crucial issue to fight against this covid-19 pandemic situation (Oli *et al*., 2020).

The databases or web tools containing updated data plays important role to support the development of several aspects, including epidemiology, pharmacology, genomic, and interactomics (Mercatelli *et al*., 2020). Several reliable and free-access databases and libraries contain comprehensive information about SARS-CoV-2, such as genomic information, pandemic spread, epitope mapping, protein function analysis, and so on. This article aims to provide literature review discussing recent progress of bioinformatics roles in the Covid-19 era, include genome organization, phylogeny analysis, and databases related to SARS-CoV2 that enhance the researchers to define the specific biomarkers towards diagnostic, therapeutic, drug and vaccine development.

* Corresponding Author:
Nur Imaniati Sumantri
Biomedical Engineering Study Program, Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia, 16424
Phone:+62217270078 Fax:+62217270077
e-mail: nur.imaniati@ui.ac.id

SARS-CoV-2 genome organizations

SARS-CoV-2 is a spherical virus-containing approximately 30 kilobases of non-segmented positive

http://dx.doi.org/10.23869/bphjbr.26.2.20212

Published by © PBI East Java. Open Access www.berkalahayati.org

sense RNA, making it the largest genome among all recognized RNA viruses. This family has a remarkable ability to easily accommodate and alter genes consisting of 6-11 open reading frames (ORFs) encoding 9,680 amino acid polyprotein (Kumar *et al*., 2020; Woo *et al*., 2020). As shown in Figure 1, the gene sequence begins with 5'-replicase ORF1ab, spike (S), envelope (E), membrane (M), nucleocapsid (N)-3', and some variable numbers of additional ORFs. Around 67 percent of the genome that encodes 16 non-structural proteins (nsps) is made up of ORF 1 (ORF1ab), while the remaining ORFs encode structural proteins and accessory proteins (Kumar *et al.,* 2020). Spike (S) proteins, envelope (E) proteins, membrane (M) proteins, and nucleocapsid (N) proteins are structural proteins, whereas nsp1, nsp2, nsp3, etc., are considered nsps (Woo *et al.,* 2020). Structural proteins in the T-cell response are found to be more immunogenic than non-structural proteins. A variety of studies have shown that the most prevalent and compliant T-cell responses are the S and N proteins (Martinelli *et al*., 2020).

ORF1ab. ORF1ab encodes polyprotein replicase in 15 to 16 nsps and occupies approximately 70 percent of their genomes. Some of these nsps, such as PLpro (nsp3), 3CLpro (nsp5), RNA-dependent RNA polymerase (Pol) (nsp12), and helicase (nsp13), have been found to encode proteins with crucial functions. Two PLpro (PL1pro and PL2pro) exist in the Alphacoronavirus and Betacoronavirus subgroup A, while PLpro exists in Betacoronavirus subgroup B, C, and D and Gammacoronavirus., The gene sequences responsible for encoding these conserved proteins are commonly used for phylogenetic analysis (Woo *et al.,* 2020).

Spike (S) proteins. The crown-like surface of the coronaviruses under electron microscopy shows up from the spike (S) proteins that critically important for receptor binding. The (S) proteins have two subunits, S1 and S2, with the S1 domains contain more divergent sequences than the S2 domains (Woo *et al.,* 2020). The receptor-binding domain (RBD) in SARS-CoV-2 (S) protein has been reported. It strongly binds to receptors called angiotensin-converting enzyme 2 (ACE2) on the surface of human cells. This binding creates a fuse of the virus with cell membrane releasing its genetic material into the cell (Wu *et al.,* 2020). In the study of human-neutralizing antibodies, evidence has shown that a vaccine targeting the S-receptor-binding-domain could effectively prevent COVID-19. The primary objective of neutralizing antibodies, improving therapeutic and vaccine design was the crown-like surface spike (S) proteins, specifically S1 domains (Salvatori *et al.*, 2020).

Haemagglutinin esterase (HE). In addition to spike (S) proteins, a distinct viral attachment called the haemagglutinin esterase (HE) protein was recently correlated for initiating infection. This property of HE is a structural glycoprotein in the bovine coronavirus (BCV) to induce neutralizing antibodies and to possess an esterase receptor-destroying activity that promotes important role for virus entry. The HE in SARS-CoV-2 functions as the classical glycan-binding lectin and receptor-degrading enzyme. O-acetylated sialic acids connect with the lectin-like spike glycoprotein of the virus to get into the host cells. Most *Betacoronavirus* notice 9-O-acetyl-SAs but switched to recognizing the 4-O-acetyl-SA form as the coronavirus did evolution (Kim, 2020*)*.

Envelope (E) proteins. Envelope (E) proteins are present at different virus infection stages and consist of 75 amino acids remaining in monomeric and homo-pentameric forms. In the early phase and growth of viral contagion, envelope proteins play a significant role. Besides that, it is said that (E) proteins are extremely immunogenic. Researchers have been targeting this protein for its immunogenicity in developing antigen-based studies, including immunodiagnostic tools and prophylactic oriented studies (Tilocca *et al.*, 2020).

Membrane (M) proteins. These transmembrane proteins mediate the incorporation of the spikes into the viral envelop, and it is the most abundant envelope protein. The M protein of SARS-CoV-2 possesses a triple helix bundle and forms a single 3-trans-membrane domain viewed by in silico analysis. This protein connects with S, E, and N protein in SARS-CoV-2 (Thomas, 2020). The envelope of all coronaviruses is associated with membrane proteins.. However, their short sequences made them less targeted genes for phylogenetic studies (Woo *et al*., 2020).

Nucleocapsid (N) proteins. Nucleocapsid (N) proteins are similar to the nsps and are considered to be another prevalent target for phylogenetic analysis. Its primary role in increasing the efficiency of virus transcription and assembly also targets cloning and recombinant protein production for serological research (Joshi *et al.*, 2020; Woo *et al.*, 2020). Amino acid sequence comparisons have shown that CoV N proteins have three distinct and strongly preserved domains: two structural and separately folded structural regions, domain-1, namely the N terminal domain (NTD) and domain-3, a long C-terminal domain (CTD) that is disjoined by domain-2, a central region that is intrinsically disordered (RNA-binding domain). These three domains have been shown to bind with viral RNA in various CoVs (Joshi *et al*., 2020).

Figure 1. SARS-CoV2 gene organization consists of ORF1a, ORF1b, Spike (S), ORF3a, Envelope (E), Membrane (M), ORF6, ORF7a, ORF7b, ORF8, Nucleocapsid (N), and ORF10

Biomarker identification

According to COVID 19 R & D TRACKER, 1,026 products released related to SARS-CoV-2 consist of 464 diagnostics, 354 therapeutics, and 208 vaccines (https://www.policycuresresearch.org/ covid-19-r-d-tracker). At the initial stage of symptoms, the COVID-19 pandemic poses critical obstacles for healthcare facilities to detect viral infections as soon as possible. By using biomarkers, clinical symptoms of patients with confirmed SARS-CoV-2 infection could be interpreted more confidently. Biomarkers can help improve prognosis and outcomes during the progression of the disease, providing objective values. Consequently enabling earlier interventions and assessing moderate, severe, or critical categories of patients (Kermali *et al*., 2020).

Enlightening biomarkers from human hosts and viruses in a fast, sensitive, and affordable method play a role in assisting this development of reliable and sensitive asymptomatic diagnostic tools. Currently, the traditional method employed in acute respiratory infection is reverse-transcriptase polymerase chain reaction (RT-PCR). Real-time RT-PCR is used to determine viruses in respiratory secretions for its highly specific and sensitive detection. This available technique is unable to test a large number of people in public space rapidly and assess infections at the asymptomatic stage. A recent study, based on the analysis of molecular reactions at the transcript level, a portable instrument called bCUBE ® 2.0, was proposed to link the gaps between traditional detection of human viruses and approaches to molecular diagnostics. Furthermore, this portable instrument approach could help reinforce the disclosure of SARS-CoV-2-specific biomarkers capable of detecting SARS-CoV-2 in restricted and public environments that connect translational genomics to technologies (Martinelli *et al*., 2020).

The previous study has summarized the biomarker abnormality of SARS-CoV-2 patients with severe systemic disease and potential new biomarkers. They are hematological biomarkers (WBC count, platelet count, etc.), biochemical biomarkers (total bilirubin, cardiac troponin, creatinine, etc.), coagulation marker (prothrombin time, D-dimer), inflammatory biomarkers (C-reactive protein, serum ferritin, etc.), and potential new biomarkers (Angiotensin monocyte-lymphocyte ratio, etc.) (Ponti *et al.*, 2020).

Another study classified the biomarkers of SARS-CoV-2 detection into three types: viral RNA genome, spike protein, and glycans that can be detected through PCR, serological, and glycan biology assay, respectively. On the other hand, biomarkers, such as small-molecule compounds, monoclonal antibodies (mAbs), convalescent plasma/serum medication, allogeneic cell therapy, have a role as a treatment target as well (Zhang and Guo, 2020). Several databases and web tools has been established to facilitate SARS-CoV-2 biomarker identification, as mentioned in Table 1.

Drugs and vaccine development

Bioinformatics supports the researchers in conducting the simulation of bioactive compound against SARS-CoV-2 through molecular modelling (Muttaqin and Ansori, 2020). This is powerful because it devotes itself to the agile production of new drugs at drastically reduced costs. Molecular docking methods play an important role in creating drugs and vaccines, with PubChem (a database of chemical molecules) being one of the most used databases in the research (Villas-Boas *et al.*, 2020). Massive molecular simulations lead compounds to be inferred as drug candidates in the wet laboratory acitivities.

Besides thinking of the treatment, we have to prepare another scheme to fight against SARS-CoV-2, which is developing vaccines. Vaccines have been shown to reduce the rate of morbidity and mortality of various infectious diseases. The process of developing a vaccine can be allergenic, high-priced, and sluggish that requires in vitro culture of pathogens that may lead to a serious safety issue. Hence, the commitment to designing a safe and effective vaccine is essential and requires a thorough investigation to end the spread of the SARS-CoV-2 infection ultimately. Three different techniques that have been used to design the SARS-CoV-2 vaccine are concluded.

First, as described earlier, the discovery of a vaccine targets the surface-exposed spike (S) glycoprotein with the full S protein or S1-receptor-binding domain. The highly predicted antigen formulation of the SARS-CoV-2 vaccine production has recently been considered the spike glycoprotein for its fundamental success in receptor binding on the host cell and the main objective of neutralizing antibodies. The development of a vaccine using modern methods of biotechnology could effectively generate engineered recombinant proteins comprising epitopes. Therefore, an epitope-based peptide vaccine obtained immunogenic of spike glycoprotein of SARS-CoV-2 is proposed as a possible candidate for the SARS-CoV-2 vaccine (Kharisma and Ansori, 2020).

A previous study has conducted alignment analysis of 58 SARS-CoV-2 sequences and acquired the SARS-CoV-2 spike glycoprotein conserved region to further evaluated through molecular docking technique for the interaction with the receptor of SARS-CoV-2 and angiotensin converting enzyme 2 (ACE2). The study found the most promising epitope-based peptide vaccine candidate possessed some advantages, such as high level of immunogenicity, preventable autoimmune mechanisms, and the ability to form BCR/Fab molecular complexes with the smallest activation binding energy that allows a direct immune response from B-cells (Kharisma and Ansori, 2020).

Second, vaccine with a computational or in silico approach was found as a fast and worthwhile solution to predict immunogenic and cross-reactive peptides from SARS-CoV-2. This approach is used due to the fast-spreading virus and the limited access to the infected sample. In the Immune Epitope Database and Analysis Resource (IEDB) database, a comprehensive description of a common immunogenic peptide is presented between SARS-CoV and SARS-CoV-2, including those derived from SARS-CoV along with de novo prediction of SARS-CoV-2 9-mer peptides, as stated in the previous study. Additionally, the de novo search of 2019-nCoV 9-

mer peptides was carried out, and a total of 63 peptides with a high immunogenicity potential were found using NetMHCpan and iPred (Lee and Koohy, 2020).

Third, a multiepitope peptide vaccine uses both immunoinformatics and a comparative genomic method to identify the suggested peptides in the production of the T-cell epitope-based peptide vaccine using the SARS-CoV-2 envelope (E) protein as a target. Peptide-based vaccines offer promising peptides that, by selectively causing antigen-specific B- and T-cells, evoke specific pathogens. The immune response activation of the peptide vaccines is based on the chemical method of synthesizing the identified epitopes of B cells and immunodominant T cells. The four major structural proteins were analyzed. The envelope (E) protein was recognized as the foremost antigenic gene which is the most feasible to be new vaccine against SARS-CoV-2 (Abdelmageed *et al.*, 2020).

Table 1. List of databases containing information of COVID-19

No	Databases	Link	Description	Reference
1	Clinicaltrials.gov	https://clinicaltrials.gov/	Gives easy access to data on publicly and privately funded clinical trials on a wide variety of diseases and disorders for patients, their family members, health care providers, researchers, and the public	Zarin, D, et al, 2011
2	COVID-evidence	https://covid-evidence.org/	Provides information about worldwide planned, ongoing, and completed trials on any intervention to treat or prevent SARS-CoV-2-infections	Janiaud, P, et al, 2020
3	Europe PMC	https://europepmc.org/	Enables access to a worldwide collection of life science publications and preprints from trusted sources around the globe	The Europe PMC Consortium
4	NCBI	https://www.ncbi.nlm.nih.gov/sars-cov-2/	Provides literature and sequences of SARS-CoV-2	Bethesda (MD), 2004
5	ViPR (The Virus Pathogen Database and	https://www.viprbrc.org/	An integrative and comprehensive publicly available database and analysis resource to search, analyze, visualize, save and share data for viral pathogens	Pickett B, et al, 2008
6	INSDC (International Nucleotide	http://www.insdc.org/	A long-standing foundational program that works between DDBJ, EMBL-EBI and NCBI. INSDC encompasses the variety of raw data readings enriched with contextual knowledge relating to samples and	Arita M, et al, 2020
7	NEXSTRAIN	https://nextstrain.org/sars-cov-2/	This page lists all of the SARS-CoV-2 analysis that we are aware of using the next strain. Researcher could use our latest Nextclade tool to compare sequences with the SARS-CoV-2 reference series, assign	Hadfield J, et al, 2018
8	hCoV-19 database	https://db.cngb.org/datamart/disease/DATAdis19/	By integrating the published coronavirus sequence information from several open source data platforms, the hCoV-19 New Coronavirus Sequence database is developed by the China National GeneBank	Chenz FZ, et al, 2020
9	GISAID	https://www.gisaid.org/	The GISAID Initiative facilitates the rapid exchange of data from all COVID-19-causing influenza viruses and coronaviruses. In order to help researchers understand how viruses grow and spread during	Eibe S, et al, 2017
10	CoV-RDB	https://covdb.stanford.edu/	The Coronavirus Antiviral Research Database is built to speed up the production of antiviral therapy for SARS-CoV-2 by helping scientists, clinical researchers, public health officials and funding agencies to	Tzou PL, et al, 2020
11	GOBIOM	https://www.excelra.com/covid-19-biomarker-	The COVID-19 Biomarker Database is a compilation of manually curated biomarkers from published clinical trials, evaluating potential drugs or biologics for the treatment of SARS-CoV-2. The database	Jagarlapudi S, Kishan K, 2009
12	DRUGBANK	https://go.drugbank.com/	DrugBank Online is a detailed, free-to-access, online database that contains drug and drug target information. We integrate extensive drug (i.e. chemical, pharmacological and pharmaceutical) data with	Wishart D, et al, 2017
13	DockCoV2	https://covirus.cc/drugs/	DockCoV2 focuses on predicting the binding affinity of FDA-approved and Taiwan National Health Insurance (NHI) drugs with the seven proteins mentioned above. This database contains a total of	Chen T, et al, 2020
14	IUPHAR/BP Guide to PHARMACOL	https://www.guidetopharmacology.org/GRAC/Coronavi	The International Union of Basic and Clinical Pharmacology (IUPHAR) / British Pharmacological Society (BPS) Guide to PHARMACOLOGY is an expertly compiled ligand-activity-target	Faccenda E, et al, 2020
15	Therapeutic Target Database	http://bidd.nus.edu.sg/group/cjttd/	Therapeutic Target Database (TTD) is a database to provide information about the known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information and the	Wang Y, et al, 2019
16	CoV3D	https://cov3d.ibbr.umd.edu/	CoV3D is a database of experimentally determined coronavirus protein structures. This resource can aid in efforts for rational vaccine design, targeting by immunotherapies, biologics, and small molecules,	Gowthaman E, et al, 2020
17	GESS (Global Evaluation of SARS-CoV-	https://wan-bioinfo.shinyapps.io/GESS/	In order to uncover the viral mutations, GESS was established, based on the analysis of single nucleotide variants (SNVs) from high-coverage and high-quality hCoV-19 viral genomes downloaded from	Eibe S, et al, 2017

Evolutionary study

Over the years, genome sequencing appeared to be an effective way for drug development. Genome sequencing has benefits that help explain the molecular basis of pathogenic diseases, the mechanism of action of drugs, and help determine individual chemotherapeutic targets. The molecular basis and corresponding pathways of the unique SARS-CoV-2 sequence, which will lead to greater production of biomarkers and therapeutic targets

in COVID-19, are therefore important to understand (Soremekun *et al*., 2020).

One of the genome sequencing methods is a phylogenetic analysis. Phylogenetics helps to identify regions of similarity and differences among genomes by comparing at different levels. From this analysis, the results are usually depicted as branching, treelike diagram, or known as the phylogenetic tree (Choudhuri, 2014). The first alignment published by Zhou et al. of an earlier coronavirus stated that pangolin coronavirus sequences are poorly defined, while the sequence similarity of the bat coronavirus to the human SARS-CoV-2 virus was 96.2 percent (Zhou *et al.*, 2020). CoVs that are most closely linked to SARS-CoV-2are CoVs for horseshoe bats. In addition, CoVs found in Malaysian pangolins are also closest to SARS-CoV-2. However, even there is still poor information about the exact origin of SARS-CoV-2, the CoVs belonging to Sarbecovirus in horseshoe bats showed a high probability to be the origin of SARS-CoV-2 (Nakagawa and Miyazawa, 2020).

Another study conducted by Fahmi *et al.* (2020), also showed similar result about the ancestor of SARS-CoV-2. They analyzed the phylogenetic profiling cluster of homolog proteins from Sarbecovirus and Hibecovirus, and found that were two nonstructural proteins, NS7b and NS8, were exclusively conserved among 2019-nCoV, BetaCoV_RaTG, and BatSARS-like Cov. Those NSPs were known to give impact on immune response signaling in experimental model infected by SARS-CoV-2 (Fahmi *et al.*, 2020). Besides, bioinformatics helps researchers analyze the spread of SARS-CoV2 to identify its evolutionary relationship in each country. Phylogenetic analysis research with genomes isolated from SARS-CoV-2 reveals that the emergence of the virus in one country results from many separate foreign introductions and inland transportation (Adebali *et al.*, 2020).

Databases for SARS-CoV-2 research

As the SARS-CoV-2 has become a pandemic and made complicated matters worldwide, the governments, researchers, and other stakeholders from various institutions make a big deal to do collaboration to solve this crisis. The advancement of technology is useful to gather and analyze the data, either clinical, epidemiological, or molecular data related to SARS-CoV-2, assist the researchers to draw a conclusion that encourage the government to make an exigent policy in ceasing the viral spread. Such fundamental scientific communication should be facilitated by adequate information technology architecture.

In recent years, there has been huge development of bioinformatic tools, including the database and prediction tools. Several databases and tools are popping up one by one, rapidly, during this pandemic situation. They become so fruitful that allow the researchers to conduct a number of in silico study which is a short cut to get into experimental study dealing with SARS-CoV-2 issue. There are a number of free-access databases (web tools) to support information exchange and advance bioinformatics analysis on SARS-CoV-2, as seen in Table 1

Conlclusion

The pandemic situation shows the urgency of information disclosure relating to SARS-CoV-2 worldwide. Molecular datasets pass over and over again in the databases giving essential information to assess and overcome the COVID-19 pandemic. The use of bioinformatics assists lots of parties to take a part in early stages of SARS-CoV-2 mutations analysis, biomarker development, drug and vaccine design, and so on. This method grants us to produce significant result of SARS-CoV-2 related research in undoubted efficiency criteria. However, clinical studies and wet laboratory operations remain the primary source of clinical decisions and research facts. Thus, bioinformatics, machine learning, and biomedicine should be combined towards comprehensive analysis to improve this COVID-19 pandemic management.

Acknowledgement

We gratefully acknowledge the funding from Universitas Indonesia through Publikasi Terindeks Internasional (PUTI) Q1 2020 No. NKB-1422/UN2.RST/HKP.05.00/2020.

References

Abdelmageed MI., Abdelmoneim AH., Mustafa MI., Elfadol NM., Murshed NS., Shantier SW., and Makhawi AM. 2020. Design of a multiepitope-based peptide vaccine against the E protein of human COVID-19: an immunoinformatics approach. *Biomed Research International,* 1-12.

Adebali O., Bircan A., Circi D., Islek B., Kilinc Z., Selcuk B., and Turhan B. 2020. Phylogenetic analysis of SARS-CoV-2 genomes in Turkey. *Turkish Journal of Biology, 44*(3), 146-156.

Ansori ANM., Kusala MKJ., Normalina I., Indrasari S., Alamudi MY., Nidom RV., Santoso KP., Rachmawati K., and Nidom CA. 2020. Immunoinformatic investigation of three structural protein genes in Indonesian SARS-CoV-2 isolates. *Systematic Reviews in Pharmacy, 11*(7), 422-434.

Arita M., Karsch-Mizrachi I., and Cochrane G. 2020. The international nucleotide sequence database collaboration. *Nucleic Acids Research, 49*(D1), D121-D124.

Chen FZ., You LJ., Yang F., Wang LN., Guo XQ., Gao F., Hua C., Tan C., Fang L., Shan RQ., Zeng WJ., Wang B., Wang R., Xu X., and Wei XF. 2020. CNGBdb: China National GeneBank DataBase. *Hereditas, 42*(8), 799-809.

Chen FT., Chang YC., Hsiao Y., Lee KH., Hsiao YC., Lin HY., Tu YCE., Huang HC., Chen CY., and Juan HF.. 2020. DockCoV2: a drug database against SARS-CoV-2. *Nucleic Acids Research, 49*(D1), D1152-D1159.

Choudhuri S., 2014. *Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools*, *1st ed.* Maryland, Elsevier, p 209-218.

Elbe S., and Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges, 1*(1), 33-46.

Faccenda E., Armstrong JF., Davenport AP., Harding SD., Pawson AJ., Southan C., and Davies JA. 2020. (Retrieved on December, 1 2020) Coronavirus information. IUPHAR/BPS guide to pharmacology. Retrieved from https://www.guidetopharmacology.org/coronavirus.jsp.

Fahmi M., Kubota Y., and Ito M. 2020. Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. *Infections, Genetics, and Evolution,* 81.

Forster P., Forster L., Renfrew C., and Forster M. 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of The National Academy of Sciences, 117*(17), 9241-9243.

Gowthaman R., Guest JD., Yin R., Adolf-Bryfogle J., Schief WR., and Pierce BG. 2020. CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Research, 49*, 282-287.

Hadfield J., Megill C., Bell SM., Huddleston J., Potter B., Callender C., Sagulenko P., Bedford T., and Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics 34*(23): 4121-4123.

Jagarlapudi SARP., and Kishan KVR. 2009. Database systems for knowledge-based discovery. *Methods in Molecular Biology, 575*, 159-172.

Janiaud P., Axfors C., Saccilotto R., Hemkens L., and Schmitt A. 2020. COVID-evidence: a living database of trials on interventions for COVID-19. *OSF*.

Joshi A., Joshi BC., Mannan MA., and Kaushik V. 2020. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Informatics in Medicine Unlocked* 19.

Kermali M., Khalsa RK., Pillai K., Ismail Z., and Harky A. 2020. The role of biomarkers in diagnosis of COVID-19 – a systematic review. *Life Sciences* 254.

Kharisma VD., and Ansori ANM. 2020. Construction of epitope-based peptide vaccine against SARS-CoV-2: immunoinformatics study. *Journal of Pure and Applied Microbiology, 14,* 999-1005.

Kim CH. 2020. SARS-CoV-2 evolutionary adaptation toward host entry and recognition of receptor o-acetyl sialylation in virus–host interaction. *International Journal of Molecular Sciences*. 21(12) : 4549.

Kumar S., Nyodu R., Maurya VK., and Saxena SK. 2020. Morphology, genome organization, replication, and pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Medical Virology: From Pathogenesis to Disease Control,* 23-31.

Lee CH., and Koohy H. 2020. In silico identification of vaccine targets for 2019-nCoV. *F1000research, 9,* 145.

Martinelli F., Perrone A., Noce ID., Colombo L., Priore SL., and Romano S. 2020. Application of a portable instrument for rapid and reliable detection of SARS-CoV-2 infection in any environment. *Immunological Reviews* 295: 4-10.

Mercatelli D., Holding AN., and Giorgi FM. 2020. Web tools to fight pandemics: the COVID-19 experience. *Briefings in Bioinformatics,* 1-11.

Muttaqin SS., and Ansori ANM. 2020. Candidate inhibitors of SARS-Cov-2 main protease with 3D structures similar to N3. *Research Journal of Biotechnology* 15(11): 60-64.

Nakagawa S., and Miyazawa T. 2020. Genome evolution of SARS-CoV-2 and its virological characteristics. *Inflammation And Regeneration, 40*(1), 17.

Oli AN., Obialor WO., Ifeanyichukwu MO., Odimegu DC., Okoyeh JN., Emechebe GO., Adejumo SA., and Ibeanu GC. 2020. Immunoinformatics and vaccine development: an overview. *ImmunoTargets and Therapy, 9,* 13–30.

Pal M., Berhanu G., Desalegn C., and Kandi V. 2020. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): an update. *Cureus, 12*(3), 7423.

Pickett BE., Sadat EL., Zhang Y., Noronha JM., Squires RB., Hunt V., Liu M., Kumar S., Zaremba S., Gu Z., Zhou L., Larson CN., Dietrich J., Klem EB., and Scheuermann RH. 2011. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research, 40,* 593-598.

Ponti G., Maccaferri M., Ruini C., Tomasi A., and Ozben T. 2020. Biomarkers associated with COVID-19 disease progression. *Critical Reviews in Clinical Laboratory Sciences, 57*(6), 389-399.

Salvatori G., Luberto L., Maffei M., Aurisicchio L., Roscilli G., Palombo F., and Marra E. 2020. SARS-CoV-2 spike protein: an optimal immunological target for vaccines. *Journal of Translational Medicine, 18*(1).

SARS-CoV-2 [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2020 11 18]. Available from: https://www.ncbi.nlm.nih.gov/sars-cov-2/.

Soremekun SO., Omolabi FK., and Soliman MES. 2020. Identification and classification of differentially expressed genes reveal potential molecular signature associated with SARS-CoV-2 infection in lung adenocarcinomal cells. *Informatics in Medicine Unlocked* 20.

The Europe PMC Consortium. 2015. Europe PMC: a full-text literature database for the life sciences and platform for innovation. *Nucleic Acids Research, 43,* 1042-1048.

Thomas S. 2020. The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter SemiSWEET. *Pathogens and Immunity, 5*(1), 342–363.

Tilocca B., Soggiu A., Sanguinetti M., Babini G., Maio FD., Britti D., Zecconi A., Bonizzi L., Urbani A., and Roncada P. 2020. Immunoinformatic analysis of the SARS-CoV-2 envelope protein as a strategy to assess cross-protection against COVID-19. *Microbes and Infection, 22*(4-5), 182-187.

Tzou PL., Tao K., Nouhin J., Rhee SY., Hu BD., Pai S., Parkin N., and Shafer RW. 2020. Coronavirus Antiviral Research Database (CoV-RDB): an online database designed to facilitate comparisons between candidate anti-coronavirus compounds. *Viruses, 12*(9).

Villas-Boas GR., Rescia VC., Paes MM., Lavorato SN., Magalhaes-Filho MF., Cunha MS., … and Oesterreich SA. 2020. The new coronavirus (SARS-CoV-2): a comprehensive review on immunity and the application of bioinformatics and molecular modeling to the discovery of potential anti-SARS-CoV-2 agents. *Molecules, 25*(18), 4086.

Wang Y, Zhang S, Li F., Zhou Y., Zhang Y., Wang Z., Zhang R., Zhu J., Ren Y., Tan Y., Qin C., Li Y., Li X., Chen Y., and Zhu F. 2019. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Research, 48*(D1), D1031-D1041.

Wishart DS., Feunang YD., Guo AC., Lo EJ., Marcu A., Grant JR., Sajed T., Johnson D., Li C., Sayeeda Z., Assempour N., Iynkkaran I., Liu Y., Maciejewski A., Gale N., Wilson A., Chin L., Cummings R., Le D., Pon A., Knox C., and Wilson M. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research, 46*(D1), D11074-D11082.

Woo PCY., Huang Yi., Huang Yi., Lau SKP., and Yuen KY. 2010. Coronavirus genomics and bioinformatics analysis. *Viruses, 2*(8), 1804-1820.

Wu J., Deng W., Li S., and Yang X. 2020. Advances in research on ACE2 as a receptor for 2019-nCoV. *Cellular and Molecular Life Sciences, 78*(2), 531-544.

Zarin DA., Tse T., Williams RJ., Califf MR., and Ide NC. 2011. The ClinicalTrials.gov results database — update and key issues. *The New England Journal of Medicine* 364(9): 852-860.

Zhang L., and Guo H. 2020. Biomarkers of COVID-19 and technologies to combat SARS-CoV-2. *Advances in Biomarker Sciences and Technology, 2,* 1-23.

Zhou P., Yang XL., Wang XG., Hu B., Zhang., Zhang W., Si HR., Zhu Y., Huang CL., Chen HD., Chen J., Luo Y., Guo H., Jiang RD., Liu MQ, Chen Y., Shen XR., Wang X., Zheng XS., Zhao K., Chen QJ., Deng F., Liu LL, Yan B., Zhan FX., Wang YY., Xiao GF., and Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature, 579,* 270-273.