

Review

Not peer-reviewed version

Genomic and Bioinformatic Tools for SARS-CoV-2: Applications in COVID-19 Diagnostics, Surveillance, and Drug Discovery

[Suresh Kumar](#)^{*}

Posted Date: 15 May 2025

doi: 10.20944/preprints202505.1123.v1

Keywords: SARS-CoV-2; Bioinformatics; Spike protein; Drug repurposing; Long COVID



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

Genomic and Bioinformatic Tools for SARS-CoV-2: Applications in COVID-19 Diagnostics, Surveillance, and Drug Discovery

Suresh Kumar

Faculty of Health and Life Sciences, Management and Science University, 40100 Shah Alam, Selangor, Malaysia; sureshkumar@msu.edu.my

Abstract: The COVID-19 pandemic, caused by SARS-CoV-2, has underscored the pivotal role of bioinformatics in elucidating viral genomics, epidemiology, and therapeutics. This review provides a curated catalog of bioinformatics resources, including databases, tools, and datasets, supporting SARS-CoV-2 research and COVID-19 surveillance. We review epidemiological databases, genomic and structural repositories, analytical tools, and clinical datasets, highlighting their applications and limitations. We discuss emerging tools, such as AI-driven structural modeling, and future needs, including data integration. By synthesizing these resources, this review aims to guide virologists, bioinformaticians, and public health researchers in addressing ongoing COVID-19 challenges and preparing for future pandemics.

Keywords: SARS-CoV-2; bioinformatics; spike protein; drug repurposing; long COVID

Introduction

The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in Wuhan, China, and has profoundly impacted global health and economies [1]. The World Health Organization named this disease COVID-19, acknowledging its global impact. Coronaviruses (CoVs), belonging to the Coronaviridae family, are enveloped, positive-sense, single-stranded RNA viruses. They are categorized into three main groups: alpha-CoVs, causing gastrointestinal disorders; beta-CoVs, including SARS-CoV, Middle East respiratory syndrome (MERS-CoV), and SARS-CoV-2; and gamma-CoVs, primarily infecting avian species. These categories are based on antigenic properties and phylogenetic analyses [2]. While some CoVs cause mild illnesses like the common cold, others, such as SARS-CoV and MERS-CoV, have triggered severe epidemics. SARS-CoV caused over 8,000 cases and 774 deaths between 2002 and 2003, with a fatality rate of 9–11%, and MERS-CoV, identified in 2012, led to severe respiratory illness. SARS-CoV-2, the seventh known human coronavirus, has demonstrated efficient human-to-human transmission via respiratory droplets, surpassing previous CoV outbreaks [2]. Limited data exist to explain its genetic variation, ancestry, and zoonotic transmission mechanisms, necessitating advanced research tools [3]. Computational studies on SARS-CoV-2 variants, such as Delta and Omicron, have revealed differences in spike protein infectivity and pathogenicity, informing vaccine and therapeutic development [4,5].

Virologists study viruses affecting humans, animals, and plants, recognizing their rapid evolution in response to host immune systems and therapeutic interventions. Addressing fundamental virology questions requires powerful genome sequencing technologies and tools to manage large datasets. Understanding coronaviruses at a structural level is critical due to the persistent health risks posed by these zoonotic viruses [6]. Numerous web-based bioinformatics tools and genomic databases support this research, with some offering comprehensive but uncurated data and others providing curated but specialized resources. These tools, computer programs that analyze sequences to identify protein domains, query millions of sequences, or locate protein-coding regions,

are often publicly accessible. However, selecting appropriate tools can be challenging, as researchers must understand their applications and availability [7]. Laboratories have shared SARS-CoV-2-related data at unprecedented speeds, generating vast datasets that require bioinformaticians' expertise to address key research questions. Collaborative efforts among bioinformaticians, virologists, and molecular biologists are essential to analyze these data. These efforts drive discoveries in fundamental and applied science and inform public health initiatives [8]. A pivotal bioinformatics breakthrough was the rapid sequencing of the SARS-CoV-2 genome and the crystal structure of its protease enzyme, enabling the design of PCR primers for diagnostic protocols [9]. This article aims to: (1) provide a comprehensive list of bioinformatics resources, (2) categorize tools based on their specificity to SARS-CoV-2 or research tasks, and (3) enable researchers to select the most appropriate tools for their studies.

Epidemiology Database

Epidemiology examines disease transmission, factors influencing health conditions in specific populations, and techniques for monitoring complications. It relies on statistical methods, comprehensive design, impartial interpretation, and rigorous analysis. Epidemiological data for the COVID-19 pandemic are accessible through several databases, including Nextstrain, Johns Hopkins Coronavirus Resource Center, WHO Coronavirus Disease Dashboard, Worldometer, Healthmap, and ViralZone.

Nextstrain, a fully accessible platform, helps researchers explore genomic pathogen data, offering real-time visualizations and analytics to support epidemiological understanding and outbreak response [10]. The Johns Hopkins Coronavirus Resource Center, a publicly accessible system, enhances awareness of viral outbreaks, informs the public, and guides policymakers on response strategies, treatment improvements, and life-saving measures [11]. The WHO Coronavirus Disease Dashboard provides daily global and country-level COVID-19 case statistics, serving as a key resource for health policy monitoring [12]. Worldometer, a statistical research site, provides data on COVID-19 cases, deaths, and recoveries, supporting general surveillance [13]. Healthmap complements health policy systems by monitoring diseases using Internet-based media and other platforms, enhancing real-time surveillance [14]. ViralZone offers molecular and epidemiological information, including virion and genome diagrams, with links to UniProtKB/Swiss-Prot viral protein entries, making it valuable for molecular epidemiology [15]. These databases collectively allow researchers and policymakers to effectively track and analyze COVID-19 trends.

Epidemiological observations indicate that SARS-CoV-2, the virus causing COVID-19, originated in Wuhan, China, and is primarily transmitted person-to-person via respiratory droplets [16]. The incubation period from infected to healthy individuals varies, depending on risk factors and symptom development [16]. Infection rates among SARS-CoV-2-infected individuals vary based on exposure duration, preventive measures, and individual health factors. Transmission includes asymptomatic and presymptomatic spread, contributing to the virus's rapid dissemination. Significant risks to disease spread include exposure types (e.g., close contact), environmental contamination, contact with animal reservoirs, immunity (humoral, cell-mediated, and vaccination status), and reinfection rates. Databases like Nextstrain and Healthmap track these factors, providing critical data for epidemiological modeling and health policy interventions.

Online Guidelines for SARS-CoV-2

Disseminating information about the COVID-19 pandemic through online platforms is essential for supporting research, surveillance, and health policy responses. Guidelines from health agencies standardize diagnostic, reporting, and prevention protocols, ensuring data consistency for epidemiological studies and bioinformatics applications. Two key platforms provide critical COVID-19 information: the Centers for Disease Control and Prevention (CDC) and World Health Organization (WHO) guidelines.

The Centers for Disease Control and Prevention (CDC), a U.S. national public health institute, protects public health by diagnosing and monitoring illness, injury, and disability in the U.S. and globally [17]. The CDC focuses on developing and implementing disease control and prevention strategies, emphasizing infectious diseases, pathogenic microorganisms, environmental health, workplace safety, health promotion, and prevention programs to enhance community health [17]. For COVID-19, the CDC provides comprehensive guidelines covering diagnosis, surveillance, and prevention measures. These guidelines standardize case definitions and testing protocols, supporting consistent data collection for epidemiological modeling and genomic research.

The World Health Organization (WHO), a global health agency, plays a crucial role in distributing COVID-19 information in collaboration with the CDC [12]. The WHO focuses on campaigning for universal health care, tracking health hazards, planning responses to health crises, and promoting population health. It provides technical assistance to nations, sets global health standards, and collects public health data. The WHO's online guidelines offer daily updates on global COVID-19 cases, vaccine development, treatments, and diagnostic tests. They also provide research and development insights and issue alerts against misinformation, ensuring reliable data for surveillance and policy-making. Together, CDC and WHO guidelines facilitate standardized data for SARS-CoV-2 research and health policy interventions.

Sequence and Structure Database Resources of COVID-19

Research on SARS-CoV-2 pathogenesis is complex and requires extensive sequence and structural database resources to understand viral mechanisms and develop vaccines or therapeutics. These databases, critical for COVID-19 research, provide genomic, proteomic, and chemical data to support studies on viral evolution, protein interactions, and drug discovery. Key resources include GISAID, NCBI SARS-CoV-2 Resources, COVID-19 Data Portal, UniProtKB COVID-19, COVID-19 Disease Map, COVID-19 Genomics UK Consortium, 2019 Novel Coronavirus Resource (2019nCoV), Protein Data Bank (PDB) COVID-19/SARS-CoV-2 Resources, PubChem, DrugBank, and ZINC Database.

Genomic Databases: The Global Initiative on Sharing All Influenza Data (GISAID) facilitates effective data exchange for influenza viruses and coronaviruses, including SARS-CoV-2 genetic sequences and clinical, epidemiological, and species-specific data for human and animal viruses. It supports research on viral spread and evolution during pandemics [18]. NCBI SARS-CoV-2 Resources, an NCBI database interface, provides comprehensive COVID-19 information, from sequence submission to literature, and links to other databases for advanced analyses [19]. The COVID-19 Data Portal integrates datasets for sharing and analysis, enabling researchers to upload and access COVID-19-related reference data within the European COVID-19 Data Platform [11]. The 2019 Novel Coronavirus Resource (2019nCoV) integrates genomic and proteomic sequences from GISAID, NCBI, NMDC, and CNCB/NGDC, providing research articles, studies, and visualization tools for genome variation analysis [12]. The COVID-19 Genomics UK Consortium delivers rapid whole-genome sequencing to NHS centers and the UK government, integrating viral genomes with clinical and epidemiological data to evaluate therapies and inform health strategies [13].

Proteomic and Structural Databases: UniProtKB COVID-19 provides pre-release data on SARS-CoV-2 protein sequences and related entries, updated independently of general UniProt releases [14]. The COVID-19 Disease Map, a community initiative, compiles molecular interaction diagrams focusing on host-pathogen interactions, sharing tools to build a knowledge repository of COVID-19 mechanisms [15]. The Protein Data Bank (PDB) COVID-19/SARS-CoV-2 Resources enables analysis of three-dimensional protein and nucleic acid structures, supporting research on SARS-CoV-2-related health and disease mechanisms [20].

Chemical Databases: PubChem, a database of small molecules, supports biochemical research on SARS-CoV-2 structural protein interactions [21]. DrugBank, a bioinformatics and cheminformatics resource, integrates comprehensive drug and target data for drug discovery and repurposing [22]. The ZINC Database, a curated collection of chemical compounds, facilitates virtual screening for

SARS-CoV-2 therapeutic development [23]. These databases collectively advance research on viral pathogenesis, vaccine design, and drug development, with computational approaches targeting the SARS-CoV-2 spike protein showing promise for therapeutic strategies [24,25].

Other Databases Related to COVID-19 Research

Beyond epidemiology and pathogenesis, SARS-CoV-2 research requires investigation into diagnostic tools, drug discovery, and molecular pathways. Several specialized databases support these efforts, including CoV2ID, canSAR, COVID-19 WikiPathways, COVID-19 Disease Map, and text-mining tools for drug prediction. These resources facilitate studies on disease pathways, diagnostic oligonucleotides, and data processing workflows, incorporating the latest data to advance COVID-19 research.

Diagnostic Databases: CoV2ID, a database for detection and therapeutic oligonucleotides, provides a regularly updated list of SARS-CoV-2-specific oligonucleotides [26]. It analyzes diagnostic and therapeutic protocols based on viral genetic diversity, enabling researchers to access genome information, perform sequence alignments, identify optimal diagnostic targets, and study sequence conservation [26].

Drug Discovery Databases: canSAR, an integrated knowledge base, combines multidisciplinary data from biology, chemistry, pharmacology, structural biology, cellular networks, and clinical annotations. It employs machine learning to predict drug discovery opportunities, supporting the identification of SARS-CoV-2 therapeutic candidates [20]. Text-mining tools for drug prediction, such as those integrated with canSAR, extract insights from literature and chemical data to prioritize potential drugs for COVID-19 treatment [21]. Protein-protein interaction (PPI) network analysis has identified new therapeutic targets for SARS-CoV-2, enhancing drug repurposing efforts [27,28].

Pathway Analysis Databases: COVID-19 WikiPathways, a community-driven platform, hosts over 20 SARS-CoV-2-related pathway models through its COVID-19 portal. These models, available in various formats via Cytoscape apps and the NDEx platform, support pathway and network analyses for understanding viral mechanisms [22]. The COVID-19 Disease Map, a global collaboration, compiles molecular interaction diagrams of SARS-CoV-2-specific host-pathogen interactions. It provides tools and methodologies to build a knowledge repository, accelerating the development of effective treatments [15]. These databases collectively enhance SARS-CoV-2 research by providing specialized data for diagnostics, drug discovery, and molecular insights.

Bioinformatics Web Tools for COVID-19

Genome Analysis Tools

In bioinformatics, biological databases play a central role in research and studies. They enable scientists to access a wide range of biologically relevant data, including the genomic sequences of an increasingly diverse array of species. Bioinformatics also includes the development of computational instruments, techniques, and software for the collection, storage, analysis, and visualization of biological knowledge [29]. Databases alone are insufficient. The earliest analytical tools developed were genome analysis tools, such as the Virus Pathogen Database and Analysis Resource (ViPR), Galaxy-SARS-CoV-2 data analysis tools, and COVID-19 Pathway Interpretation and Analysis.

The Virus Pathogen Database and Analysis Resource (ViPR) is a data archive and analytical tool for multiple virus families [30]. ViPR contains information on human pathogenic viruses, enabling the recording of sequence records, gene and protein annotations, three-dimensional protein structures, immune epitope locations, and clinical and monitoring data derived from comparative genomics analysis [31]. It provides technical and simulation tools for metadata-driven statistical sequence analysis, multiple sequence alignment, phylogenetic tree building, BLAST results comparison, and sequence variation determination [32]. To promote the development of prophylactics, diagnostics, and therapeutics for major viruses, ViPR's instruments and data are freely available to the virology community [30].

Galaxy is an open-source, web-based platform for data-intensive biomedical research. Galaxy-SARS-CoV-2 data analysis tools provide infrastructure and workflows for SARS-CoV-2 data analyses accessible to the public [30]. These tools currently feature three major types of analyses—genomics, evolution, and cheminformatics—as well as other functions, such as proteomics, artc, and direct RNA-seq. COVID-19 Pathway Interpretation and Analysis (CoV-Hipathia) is a web tool that implements a mechanistic model of human signaling to interpret the consequences of combined changes in gene expression levels and/or genomic mutations in the context of signaling pathways known to be involved in SARS-CoV-2 infection, updated with curated versions released by the COVID-19 Disease Map curation project [15].

Modelling and Drug Design Tools

Homology modelling, also known as protein comparative modelling, creates an atomic-resolution model of a "target" protein from its amino acid sequence and an experimental three-dimensional model of a similar homologous protein [33]. This modelling becomes a valuable method for predicting protein structure when sequence knowledge is available [34]. Structural knowledge is often more critical than sequence alone for determining protein function. The accuracy of the predicted configuration depends on the degree of similarity between the model and template sequences through homology modelling [35]. If the similarity is minimal, homology modelling of the query protein does not yield relevant outcomes. Tools related to homology modelling include SWISS-MODEL, Phyre2, and MODELLER.

SWISS-MODEL, a popular web-based modelling tool, aims to make protein simulation accessible to all life science researchers worldwide [33]. Researchers have modelled the full SARS-CoV-2 proteome based on the NCBI reference sequence and annotations from UniProt. Phyre2 is a web-based tool for simulating and analyzing protein structure, behavior, and variants [34]. Phyre2 provides an easy and reliable framework for specifying protein bioinformatics tools for all researchers. MODELLER is used for homology or comparative modelling of three-dimensional protein structures [35]. Users provide an alignment with known related structures of a sequence to be modelled, and MODELLER automatically calculates a model containing all non-hydrogen atoms. By meeting spatial restraints, MODELLER performs comparative protein structure modelling and can perform several additional tasks. Computational approaches using these tools have been instrumental in designing drugs and vaccines against the SARS-CoV-2 spike protein [24,25].

Docking Tools

Docking, a molecular modelling technique, predicts the preferred orientation of one molecule to another when bound to form a stable complex [36]. Molecular docking, due to its ability to predict the binding conformation of small molecule ligands to the appropriate target binding site, is a commonly used method in structure-based drug design [37]. There are two types of docking servers: receptor-based docking servers and ligand-based docking servers. Receptor-based docking servers include COVID-19 Docking Server, SwissDock, DockThor, MTiOpenScreen, EasyVS, CaverWeb, DOCK Blaster, e-LEA3D, ezCADD, iScreen, and Pharmit. Ligand-based docking servers include SwissTargetPrediction, TargetNet, SuperPred, SEA, Anglerfish, ChemProt, HitPickV2, MolTarPred, and SuperPred.

Vaccine Design Tools

Computational vaccinology includes epitope mapping, antigen collection, and immunogen design using computational methods. Tools that promote in silico prediction of immune responses to emerging infectious diseases and cancers can facilitate the production and distribution of novel vaccines to the clinic [29]. This integrated collection of immunoinformatic tools includes scoring and triage algorithms for candidate antigens, the selection of immunogenic and retained T cell epitopes, the re-engineering or removal of regulatory T cell epitopes, and the redesign of immunogenicity-

inducing and disease-protective antigens for humans and livestock [38]. Vaccine design tools include SYFPEITHI, MHCBN, EPIMHC, Propred, MHCpred, NetMHC 4.0 Server, and Epitope Prediction and Analysis Tools.

SYFPEITHI is a database comprising over 7,000 peptide sequences known to bind class I and class II MHC molecules. It provides information on peptide sequences, anchor positions, MHC specificity, source proteins, source organisms, epitope prediction, and publication references [38]. MHCBN is a curated database containing detailed information about Major Histocompatibility Complex (MHC) binding, non-binding peptides, and T-cell epitopes [39]. The newest version 4.0 of the database provides information about peptides interacting with TAP and MHC-linked autoimmune diseases. EPIMHC is a relational database of MHC-binding peptides and T-cell epitopes observed in real proteins, accessible through a web server designed to facilitate computational vaccinology research [39]. Propred predicts MHC Class-II binding regions in an antigen sequence using quantitative matrices derived from published literature, assisting in locating promiscuous binding regions useful for selecting vaccine candidates [39].

MHCpred uses the additive method to predict the binding affinity of major histocompatibility complex (MHC) class I and II molecules and the Transporter associated with Processing (TAP). Allele-specific Quantitative Structure Activity Relationship (QSAR) models were generated using partial least squares (PLS) [40]. The NetMHCIIpan-4.0 server predicts peptide binding to any MHC II molecule of known sequence using Artificial Neural Networks (ANNs) [40]. The Immune Epitope Database Analysis Resource provides a collection of tools for the prediction and analysis of immune epitopes [41]. It serves as a companion site to the Immune Epitope Database (IEDB), a manually curated database of experimentally characterized immune epitopes. The tools include T Cell Epitope Prediction Tools (MHC class I & II binding predictions, peptide processing predictions, and immunogenicity predictions) and B Cell Epitope Prediction Tools (predicting regions of proteins likely to be recognized as epitopes in the context of a B cell response). Epitope analysis tools enable detailed analysis of known epitope sequences or groups of sequences.

COVIDep provides an up-to-date set of B-cell and T-cell epitopes that can serve as potential vaccine targets for SARS-CoV-2, the virus causing the COVID-19 pandemic. The identified epitopes are experimentally derived from SARS-CoV (the virus that caused the 2003 SARS outbreak) and have a close genetic match with available SARS-CoV-2 sequences [41]. Computational studies on mutations like D614G, N501Y, and S477N have further informed vaccine design by assessing their impact on SARS-CoV-2 infectivity [42].

COVID-19 Dataset

A COVID-19 dataset is a collection of data related to the COVID-19 pandemic. It contains the latest available public data on COVID-19, including daily situation updates, the epidemiological curve, and global geographical distribution [43]. COVID-19 dataset sources include the COVID-19 Open Patent Dataset (hosted by Lens.org), NCATS OpenData | COVID-19, GEO Datasets, Sequence Read Archive (SRA), and CORD-19 COVID-19 Open Research Dataset.

The Lens has compiled free, open datasets of patent documents, scholarly research metadata, and biological sequences from patents, depositing them in a machine-readable, explorable format [43]. The Lens is building an interactive tool for understanding the landscape of patent and research works in any domain, including human coronaviruses and COVID-19. NCATS generates a collection of datasets by screening a panel of SARS-CoV-2-related assays against all approved drugs. These datasets and assay protocols are made immediately available to the scientific community [43]. GEO DataSets is a database of gene expression curated profiles maintained by NCBI and included in the Gene Expression Omnibus [19]. The Sequence Read Archive (SRA) stores raw sequence data from next-generation sequencing technologies, including Illumina, 454, IonTorrent, Complete Genomics, PacBio, and Oxford Nanopores. SRA also stores alignment information in the form of read placements on a reference sequence [19]. The CORD-19 COVID-19 Open Research Dataset provides researchers with free and open tools and datasets to find new insights about SARS-CoV-2 [43].

Tracker Tools

ClinicalTrials is a database of privately and publicly funded clinical studies conducted worldwide. The ClinicalTrials Protocol Registration and Results System (PRS) is used to register a clinical study or submit results information for a registered study [44]. COVID-evidence is a continuously updated database of worldwide available evidence on interventions for COVID-19 [45]. It combines automatic search strategies with expert manual extraction of key trial characteristics performed in duplicate, providing information about planned, ongoing, and completed trials on any intervention to treat or prevent SARS-CoV-2 infections [45]. Covid-19 TrialsTracker lists all COVID-19 clinical trials and observational studies registered on a WHO primary clinical trial registry or ClinicalTrials.gov. Data fields include trial ID number, date of registration, sponsor, recruitment status, completion date, and countries covered [44]. The code and dataset can be downloaded. The Cochrane COVID-19 Study Register is a freely available, continually updated, annotated reference collection of human studies on COVID-19, including interventional, observational, diagnostic, prognostic, epidemiological, qualitative, and economic designs [43]. Cochrane's COVID-19 Study Register is study-based, linking references to the same study (e.g., press releases, trial registry records, preprints, journal articles, retraction notices, and expressions of concern) to a single study record.

COVID-19 Literature Mining Tools

The iSearch COVID-19 portfolio is NIH's comprehensive, expert-curated source for publications and preprints related to COVID-19 or SARS-CoV-2 [46]. Its COVID-19 Portfolio tool leverages the cutting-edge analytical capabilities of the iSearch platform, offering powerful search functionality and faceting, and includes articles from PubMed and preprints from arXiv, bioRxiv, ChemRxiv, medRxiv, Research Square, and SSRN [43]. The portfolio is updated daily with the latest available data. LitCovid is a curated literature hub for tracking up-to-date scientific information about SARS-CoV-2, providing central access to 71,048 (and growing) relevant articles in PubMed [47]. PubMed Central (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM). Publishers voluntarily make their COVID-19 and coronavirus-related publications and supporting data immediately accessible in PMC and other public repositories to support ongoing public health emergency response efforts [48]. Computational analyses of SARS-CoV-2 gene expression have also revealed associations with long COVID symptoms, providing insights into chronic disease mechanisms [49].

Other Tools

BioExcel-CV19 is a platform designed to provide web access to atomistic-MD trajectories for macromolecules involved in the COVID-19 disease. The project is part of open-access initiatives promoted by the worldwide scientific community to share information about COVID-19 research [46]. The BioExcel-CV19 web server interface presents the resulting trajectories with a set of quality control analyses and system information. CoV-AbDab, the Coronavirus Antibody Database, consolidates antibodies known to bind SARS-CoV-2 and other betacoronaviruses, such as SARS-CoV-1 and MERS-CoV. It includes relevant metadata, including evidence of cross-neutralization, antibody/nanobody origin, full variable domain sequence (where available), germline assignments, epitope region, links to relevant PDB entries, homology models, and source literature [50]. PhylomeDB is a public database for complete catalogs of gene phylogenies (phylomes). It allows users to interactively explore the evolutionary history of genes through the visualization of phylogenetic trees and multiple sequence alignments [48]. Additionally, PhylomeDB provides genome-wide orthology and paralogy predictions based on phylogenetic tree analysis. 3DBIONOTES-WS is a web application designed to automatically annotate biochemical and biomedical information onto structural models, including post-translational modifications, genomic

variations associated with diseases, short linear motifs, immune epitope sites, disordered regions, and domain families [51].

Table 1. Bioinformatics Databases and Tools for SARS-CoV-2 Research.

Website	Key Features	URL References
Nextstrain	Genomic epidemiology	https://nextstrain.org/ncov/global
COVID-19 Map - Johns Hopkins Coronavirus Resource Center	Epidemiology, updated data visualization, statistically represented data	https://coronavirus.jhu.edu/map.html
WHO Coronavirus Disease (COVID-19) Dashboard	Overview of pandemic in graphics, data-represented tables	https://covid19.who.int
Worldometer	Overview of pandemic in graphics, data-represented tables	https://www.worldometers.info/coronavirus
Healthmap	Overview of pandemic in graphics, data-represented tables	https://www.healthmap.org/covid-19
CDC Guidelines	Freely accessible platform for updated COVID-19 information	https://www.cdc.gov/coronavirus/2019-ncov/index.html
WHO Guidelines	Freely accessible platform for updated COVID-19 information	https://www.who.int/emergencies/diseases/novel-coronavirus-2019
GISAID	CoVsurver mutations App, FluSurver mutations App, influenza genomic epidemiology, phylodynamics, Submitting Data to EpiFlu, hCoV-19 Reference Sequence	https://www.gisaid.org
NCBI SARS-CoV-2 Resources	Sequence Submission, Literature, Sequence-Related Resources, Clinical Resources, Extra Linked Resources	https://www.ncbi.nlm.nih.gov/sars-cov-2
COVID-19 Data Portal	Viral Sequencing, Host Sequencing, Expression, Proteins, Biochemistry, Literature	https://www.covid19dataportal.org
COVID-19 UniProtKB	Protein and biological sequences archives and analysis tools	https://covid-19.uniprot.org/uniprotkb?query=*
COVID-19 Disease Map	Provides COVID-19 disease map and related data and literature resources	https://covid.pages.uni.lu
COVID-19 Genomics UK Consortium	Modelling Phylogenetics Display, Sample logistics, Metadata/Patient linkage, bioinformatics, Sequencing, Clinical/Virology, public health interpretation	https://www.cogconsortium.uk
2019 Novel Coronavirus Resource (2019nCoV)	Genomic sequencing, variations, online tools, literature	https://bigd.big.ac.cn/ncov/?lang=en
COVID-19/SARS-CoV-2 Resources	Data deposition, visualization, analysis	https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true
PDB structure-COVID-19/SARS-CoV-2 Resources	Data deposition, visualization, analysis	https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true
PubChem	Chemicals, identifiers, Bioactivities, Literature	https://pubchem.ncbi.nlm.nih.gov

DrugBank	Provides clinical information on drugs	https://go.drugbank.com
ZINC Database	Chemical compound virtual screening	http://zinc15.docking.org
CoV2ID	Reference genome, Alignments, Oligos, Genome variation	http://covid.portugene.com/cgi-bin/COVid_home.cgi
canSAR	Druggable Interactome, Clinical Trials, Drugs, Chemical Probes, 3D Structures	https://corona.cansar.icr.ac.uk
COVID-19 WikiPathways	Pathways construction	https://www.wikipathways.org/index.php/Portal:COVID-19
Disease Maps and Text Mining for Drug Prediction	Pathways, networks, ontology	https://covid19map.lcsb.uni.lu/minerva/index.xhtml?id=hackathon_covid19_map_v3
The Virus Pathogen Database and Analysis Resource (ViPR)	Data searching and archive, data analysis, and provides workbench	https://www.viprbrc.org/brc/home.spg?decorator=vipr
Galaxy-SARS-CoV-2 Data Analysis Tools	Genomics, Cheminformatics, Evolution, Direct RNAseq, Proteomics, Artic	https://covid19.galaxyproject.org
COVID-19 Pathway Interpretation and Analysis	Differential signalling, Prediction, Perturbation effect, Variant interpreter	http://hipathia.babelomics.org/covid19
SWISS-MODEL	Protein structure homology-modelling	https://swissmodel.expasy.org
Phyre2	Fold recognition server for predicting the structure and/or function of protein sequence	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
MODELLER	Comparative protein structure modeling	https://salilab.org/modeller
SYFPEITHI	Peptide sequences, anchor positions, MHC specificity, source proteins, source organisms, publication references	http://www.syfpeithi.de
MHCBN	Query searching, TAP search, Peptide mapping, Dataset Creation, Similarity search	http://crdd.osdd.net/raghava/mhcbn
EPIMHC	Alignment Analysis, Databases, Computational Immunology, Modelling & 3D-structure Analysis, Sequence Manipulation & Analysis, Similarity Searches	http://imed.med.ucm.es/epimhc
Propred	Sequence submission	http://crdd.osdd.net/raghava/propred
MHCPred	Heteroclitic peptide calculation	http://www.ddg-pharmfac.net/mhcpred/MHCPred
NetMHC 4.0 Server	Prediction of peptide-MHC class I binding using artificial neural networks (ANNs)	http://www.cbs.dtu.dk/services/NetMHC
Epitope Prediction and Analysis Tools	T Cell Tools, B Cell Tools, Analysis Tools, Tools-API, Datasets, Contribute Tools, References	http://tools.immuneepitope.org/main
COVID-19 Open Patent Dataset	Open datasets on COVID-19-related data	https://about.lens.org/covid-19

(hosted by Lens.org)		
NCATS OpenData	COVID-19	Open data browser, assays, animal models, omics efforts
GEO Datasets	Alignment Analysis, Databases, Sequence Manipulation & Analysis, Similarity Searches, Sequences alignment	https://www.ncbi.nlm.nih.gov/gds/?term=%28covid-19%20OR%20SARS-COV-2%29%20AND%20gse%5Bentry%20type%5D
Refined Structure Data	Provides COVID-19 viral structures and literature	https://covid19.bioreproducibility.org
Sequence Read Archive (SRA)	Data browsing, searches, Submission and online tools	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=search_obj&m=&s=&term=txid2697049%5Borgn%5D&go=Search
CORD-19 COVID-19 Open Research Dataset	Literature	https://www.semanticscholar.org/cord19
ClinicalTrials	Sequences annotations and archives, online tools analysis	https://clinicaltrials.gov/ct2/results?cond=COVID-19
COVID-evidence	Databases and literature	https://covid-evidence.org
Covid-19 TrialsTracker	Sequencing, Visualization, data archives and interactive	http://covid19.trialstracker.net
Cochrane COVID-19 Study Register	Databases and literature	https://covid-19.cochrane.org
COVID-19: A Living Systematic Map of the Evidence	Genome mapping	http://epi.ioe.ac.uk/cms/Projects/DepartmentofHealthandSocialCare/Publishedreviews/COVID-19Livingssystematicmapoftheevidence/tabid/3765/Default.aspx
Vaccine Tracker	Archives, Databases, Literature	https://vaclshmt.shinyapps.io/ncov_vaccine_landscape
iSearch COVID- 19 Portfolio	Query search, literature	https://icite.od.nih.gov/covid19/search
LitCovid	General mechanism, transmission, diagnosis, treatment, prevention, case Report, forecasting	https://www.ncbi.nlm.nih.gov/research/coronavirus
PubMed Central (PMC) COVID-19 Initiative	Literature, online tools, Archives, sequencing	https://www.ncbi.nlm.nih.gov/pmc/about/covid-19
BioExcel-CV19	Modelling and structural modelling	https://bioexcel-cv19.bsc.es/#/
CoV-AbDab: The Coronavirus Antibody Database	Databases, data archives	http://opig.stats.ox.ac.uk/webapps/covabdab
Coronavirus Phylomes	Phylomes, search query, sequencing	http://beta.phylomedb.org/covid19
3DBIONOTES- WS	Literature, proteomics, genomics	https://3dbionotes.cnb.csic.es/ws/covid19

Conclusions

Bioinformaticians and computer scientists worldwide are working diligently to combat SARS-CoV-2 following the COVID-19 outbreak. Operating behind the scenes, they have provided vital knowledge to help medical agencies develop medicines, diagnostic tools, and vaccines to defeat COVID-19. Detailed sequencing of the SARS-CoV-2 biological sequence was the first bioinformatic breakthrough in this outbreak. Research institutions have sequenced and elucidated the genome of SARS-CoV-2 and the crystal structure of its protease enzyme. The protease structure is linked to the

N3 inhibitor, which is critical knowledge for drug development efforts. This sequencing marks the initial stage in developing medications, diagnostic tools, and a vaccine for the hazardous virus. These efforts benefit from identifying genes that code for viral replication and the protein responsible for host cellular uptake. Recent computational analyses have further explored long COVID mechanisms, highlighting associations with chronic diseases like Alzheimer’s, enhancing our understanding of post-acute sequelae [49,52].

References

1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3.
2. Satapathy P, Kumar P, Sood M, et al. SARS-CoV-2 evolution: Decoding the genomic dynamics of a global pandemic. *Front Microbiol*. 2024;15:1334152. doi:10.3389/fmicb.2024.1334152.
3. Chiara M, Horner DS, Gissi C, Pesole G. Comparative genomics reveals early emergence and biased spatiotemporal distribution of SARS-CoV-2. *Mol Biol Evol*. 2021;38(6):2547-2565. doi:10.1093/molbev/msab049.
4. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. *J Med Virol*. 2022;94(4):1641-1649. doi:10.1002/jmv.27526.
5. Kumar S, Karuppanan K, Subramaniam G. Omicron (BA.1) and sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: A comparative sequence and structural-based computational assessment. *J Med Virol*. 2022;94(10):4780-4791. doi:10.1002/jmv.27927.
6. Jin Y, Lei C, Hu D, et al. Structural basis for the zoonotic transmission of SARS-CoV-2. *Cell*. 2023;186(8):1632-1645. doi:10.1016/j.cell.2023.02.015.
7. Liu Z, Chen H, Wang X, et al. An amalgamation of bioinformatics and artificial intelligence for COVID-19 management: From discovery to clinic. *Comput Struct Biotechnol J*. 2024;22:100123. doi:10.1016/j.csbj.2024.100123.
8. Robson B. Bioinformatics in the age of pandemics: Challenges and opportunities. *Comput Biol Med*. 2023;155:106678. doi:10.1016/j.combiomed.2023.106678.
9. Harcourt J, Tamin A, Lu X, et al. Severe acute respiratory syndrome coronavirus 2 from patient with coronavirus disease, United States. *Emerg Infect Dis*. 2020;26(6):1266-1273. doi:10.3201/eid2606.200516.
10. Hadfield J, Megill C, Bell SM, et al. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-4123. doi:10.1093/bioinformatics/bty407.
11. Johns Hopkins Coronavirus Resource Center. Home. Available from: <https://coronavirus.jhu.edu/>. Accessed May 14, 2025.
12. World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. Available from: <https://covid19.who.int/>. Accessed May 14, 2025.
13. Worldometer. About us. Available from: <https://www.worldometers.info/about/>. Accessed May 14, 2025.
14. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc*. 2008;15(2):150-157. doi:10.1197/jamia.M2554.
15. ViralZone. Molecular and epidemiological information. Available from: <https://viralzone.expasy.org/>. Accessed May 14, 2025.
16. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun*. 2020;109:102433. doi:10.1016/j.jaut.2020.102433.



17. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19). Available from: <https://www.cdc.gov/coronavirus/2019-ncov/index.html>. Accessed May 14, 2025.
18. GISAID. Global Initiative on Sharing All Influenza Data. Available from: <https://www.gisaid.org/>. Accessed May 14, 2025.
19. NCBI. SARS-CoV-2 Resources. Available from: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>. Accessed May 14, 2025.
20. Protein Data Bank. COVID-19/SARS-CoV-2 Resources. Available from: <https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true>. Accessed May 14, 2025.
21. PubChem. National Library of Medicine. Available from: <https://pubchem.ncbi.nlm.nih.gov/>. Accessed May 14, 2025.
22. DrugBank. Drug and target data. Available from: <https://go.drugbank.com/>. Accessed May 14, 2025.
23. ZINC Database. Chemical compound virtual screening. Available from: <http://zinc15.docking.org/>. Accessed May 14, 2025.
24. Kumar S. Drug and vaccine design against novel coronavirus (2019-nCoV) spike protein through computational approach. Preprints.org. 2020. doi:10.20944/preprints202004.0292.v1.
25. Kumar S. Online resource and tools for the development of drugs against novel coronavirus. In: Roy K, ed. In Silico Modeling of Drugs Against Coronaviruses: Computational Tools and Protocols. New York, NY: Springer US; 2021:735-759.
26. CoV2ID. Detection and therapeutic oligonucleotides. Available from: http://covid.portugene.com/cgi-bin/COVid_home.cgi. Accessed May 14, 2025.
27. Kumar S. Protein-protein interaction network for the identification of new targets against novel coronavirus. In: Roy K, ed. In Silico Modeling of Drugs Against Coronaviruses: Computational Tools and Protocols. New York, NY: Springer US; 2021:213-230.
28. Kumar S. COVID-19: A drug repurposing and biomarker identification by using comprehensive gene-disease associations through protein-protein interaction network analysis. Preprints.org. 2020. doi:10.20944/preprints202003.0360.v1.
29. Cleemput S, Dumon W, Fonseca V, et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. Bioinformatics. 2020;36(11):3552-3555. doi:10.1093/bioinformatics/btaa145.
30. Pickett BE, Greer DS, Zhang Y, et al. Virus Pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics database and analysis resource for the coronavirus research community. Viruses. 2012;4(11):3209-3226. doi:10.3390/v4113209.
31. Gómez-Carballa A, Bello X, Pardo-Seco J, et al. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res. 2020;30(10):1434-1448. doi:10.1101/gr.266221.120.
32. Sun J, Zhuang Z, Zheng J, et al. Generation of a broadly useful model for COVID-19 pathogenesis, vaccination, and treatment. Cell. 2020;182(3):734-743.e5. doi:10.1016/j.cell.2020.06.010.
33. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303. doi:10.1093/nar/gky427.
34. Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845-858. doi:10.1038/nprot.2015.053.
35. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics. 2016;54:5.6.1-5.6.37. doi:10.1002/cpbi.3.

36. McBryde E, Meehan M, O'Neill A. Role of modelling in COVID-19 policy development. *Paediatr Respir Rev.* 2020;35:57-60. doi:10.1016/j.prrv.2020.06.013.
37. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Appl Intell.* 2020;50(11):3913-3925. doi:10.1007/s10489-020-01770-9.
38. Rammensee HG, Bachmann J, Emmerich NP, et al. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics.* 1999;50(3-4):213-219. doi:10.1007/s002510050595.
39. Bhasin M, Singh H, Raghava GP. MHCBN: A comprehensive database of MHC binding and non-binding peptides. *Bioinformatics.* 2003;19(5):665-666. doi:10.1093/bioinformatics/btg055.
40. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* 2007;8:4. doi:10.1186/1471-2105-8-4.
41. Vita R, Mahajan S, Overton JA, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2019;47(D1):D339-D343. doi:10.1093/nar/gky1006.
42. Mathavan S. Evaluation of the effect of D614G, N501Y and S477N mutation in SARS-CoV-2 through computational approach. *Preprints.org.* 2020. doi:10.20944/preprints202012.0457.v1.
43. Lens.org. COVID-19 Open Patent Dataset. Available from: <https://about.lens.org/covid-19/>. Accessed May 14, 2025.
44. ClinicalTrials.gov. COVID-19 clinical trials. Available from: <https://clinicaltrials.gov/ct2/results?cond=COVID-19>. Accessed May 14, 2025.
45. COVID-evidence. Worldwide evidence on COVID-19 interventions. Available from: <https://covid-evidence.org/>. Accessed May 14, 2025.
46. iSearch COVID-19 Portfolio. NIH-curated publications and preprints. Available from: <https://icite.od.nih.gov/covid19/search/>. Accessed May 14, 2025.
47. LitCovid. Curated literature hub for SARS-CoV-2. Available from: <https://www.ncbi.nlm.nih.gov/research/coronavirus/>. Accessed May 14, 2025.
48. PubMed Central. COVID-19 Initiative. Available from: <https://www.ncbi.nlm.nih.gov/pmc/about/covid-19/>. Accessed May 14, 2025.
49. Das S, Kumar S. Exploring the mechanisms of long COVID: Insights from computational analysis of SARS-CoV-2 gene expression and symptom associations. *J Med Virol.* 2023;95(9):e29077. doi:10.1002/jmv.29077.
50. CoV-AbDab. Coronavirus Antibody Database. Available from: <http://opig.stats.ox.ac.uk/webapps/covabdab/>. Accessed May 14, 2025.
51. 3DBIONOTES-WS. Biochemical and biomedical annotation. Available from: <https://3dbionotes.cnb.csic.es/ws/covid19>. Accessed May 14, 2025.
52. Shajahan SR, Kumar S, Ramli MDC. Unravelling the connection between COVID-19 and Alzheimer's disease: A comprehensive review. *Front Aging Neurosci.* 2023;15:1274452. doi:10.3389/fnagi.2023.1274452.

53. Xie Y, Chen G, Wu W, et al. A bioinformatics approach combined with experimental validation analyzes the efficacy of azithromycin in treating SARS-CoV-2 infection in patients with IPF and COPD. *Sci Rep.* 2025;15:10009. doi:10.1038/s41598-025-94801-9.
54. World Health Organization. Types of data requested to inform May 2025 COVID-19 vaccine antigen composition deliberations. Available from: <https://www.who.int/news/item/25-03-2025-types-of-data-requested-to-inform-may-2025-covid-19-vaccine-antigen-composition-deliberations>. Accessed May 14, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.