Briefings in Bioinformatics, 22(2), 2021, 690–700

doi: 10.1093/bib/bbaa261
Advance Access Publication Date: 15 October 2020
Method Review

Web tools to fight pandemics: the COVID-19 experience

Daniele Mercatelli, Andrew N. Holding and Federico M. Giorgi

Corresponding author: Federico M. Giorgi, Department of Pharmacy and Biotechnology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy.
E-mail: federico.giorgi@unibo.it

Abstract

The current outbreak of COVID-19 has generated an unprecedented scientific response worldwide, with the generation of vast amounts of publicly available epidemiological, biological and clinical data. Bioinformatics scientists have quickly produced online methods to provide non-computational users with the opportunity of analyzing such data. In this review, we report the results of this effort, by cataloguing the currently most popular web tools for COVID-19 research and analysis. Our focus was driven on tools drawing data from the fields of epidemiology, genomics, interactomics and pharmacology, in order to provide a meaningful depiction of the current state of the art of COVID-19 online resources.

Key words: COVID-19; SARS-CoV-2; epidemiology; genomics; interactomics; web tools

Introduction

SARS-CoV-2 emerged in late December 2019 in Wuhan, China, as the seventh coronavirus shown to infect humans. Together with SARS-CoV, the causative agent of the severe acute respiratory syndrome outbreak in 2003, and MERS-CoV, the pathogenic coronavirus strain causing Middle East Respiratory Syndrome in 2012, SARS-CoV-2 is the third betacoronavirus of zoonotic origin to pose a serious threat to global health [1]. This novel single-stranded positive RNA [(+) ssRNA] virus causes a mild to severe respiratory disease named COVID-19 which can lead to acute respiratory distress syndrome requiring intensive care treatment, sepsis, septic shock and death in about 5.2% of diagnosed cases [2]. In less than 6 months, SARS-CoV-2 has spread in 213 countries, counting more than 10 million tested-positive cases, and more than half a million confirmed deaths [3].

SARS-CoV-2 has spread with astonishing ease across countries, and the initial lack of national or supra-national plans to face health emergencies of such magnitude has had a diverse impact over different regions [4, 5]. A global response was quickly developed in the form of collective data collection and analysis efforts, generally aimed at understanding SARS-CoV-2 biology and delivering therapeutic solutions in the form of clinical and pharmacological protocols [6].

Strategies to defeat COVID-19 have been built on previously developed methods for disease outbreak study, tracking and containment [7], most notably those for influenza [8] and SARS [9]. Specifically, classic epidemiological elements, such as case incidence data and field surveys, have been integrated with vast usage of next-generation sequencing technologies, such as those provided by Oxford Nanopore and Illumina [10]. Different methods for library preparation have been tested, based on whole metatranscriptomic approaches, PCR amplification or hybrid capture, designed to be able to extract SARS-CoV-2 sequences from both cultured media or pharyngeal swabs [11]. Sequencing methods have been paired with classic phylogenetic analyses [1] and the more debated use of network-based genetic

Daniele Mercatelli is a Postdoctoral Research Scientist at the University of Bologna. He works on the analysis of genomes and transcriptional networks in pathological contexts.
Andrew N. Holding is a Lecturer in Biomedical Science at the University of York. His focus is characterizing the role of steroid hormone receptors, by developing computational and experimental methods to answer key questions.
Federico M. Giorgi is Assistant Professor in Bioinformatics at the University of Bologna. He works on the construction of mathematical models to process epidemiological, transcriptional and genetic data.
Organization: The Department of Pharmacy and Biotechnology is a multidisciplinary structure located at the hearth of the University of Bologna, Italy, and it is dedicated to improving human health by the combination of molecular biology, pharmacology and bioinformatics.
Submitted: 8 July 2020; **Received (in revised form):** 26 August 2020

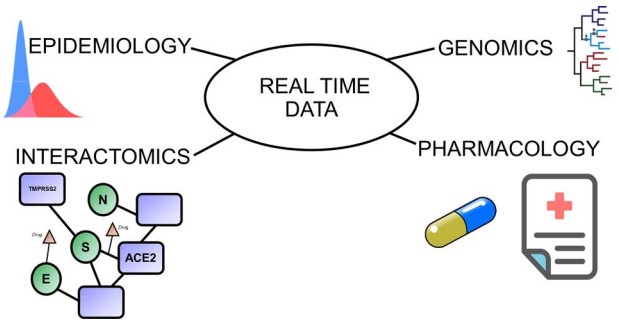

Figure 1. Diagram of COVID-19 web tools described in this review. A few example tools are indicated (for the full list, see Table 1).

divergence studies [12, 13]. Combining those distances with different clustering techniques or phylogenetic trees and network-based methods researchers can effectively track the mutational evolution of SARS-CoV-2 genomes to trace COVID-19 infection sources [14]. Genomic surveillance will also be useful to identify novel mutations conferring a drug-resistant or vaccine-resistant phenotype or altering virulence properties [15].

Epidemiology and genomics have been supported in the fight against SARS-CoV-2 by molecular biology and structural biology to define the map of functional interactions between viral and human proteins [16]. This interactomics approach collects data on protein–protein interactions and also on drug-protein networks, whereas drugs can be targeted against viral proteins or against specific host-virus interactions [17]. While SARS-CoV-2 does not mutate at a high rate, with less than 25 predicted mutations per year [18], constant mutation tracking is currently being exercised, since any mutation could in theory influence the development of specific antiviral therapies and vaccines [19]. The further elucidation of the SARS-CoV-2/human protein interaction network will greatly sustain the drug discovery process [20], and most of the hypothesis-driven clinical trials on FDA approved drugs have come from targets identified from interactome maps, like ACE2 inhibitors [21].

The urgency of the COVID-19 crisis has ignited a previously unseen global scientific response with unmatched speed and multidisciplinary spirit, witnessing entire remodulation of previously existing laboratories and research programs. This has fueled and reciprocally been fueled by the collection of a vast amount of data that have been made freely accessible and shared online [22]. In this scenario, data science and bioinformatics play a prominent role in analyzing real-time information to foster basic research and decipher viral biology, drug discovery and disease management. A number of web tools for SARS-CoV-2-related analysis have been specifically developed or readapted to enable researchers to easily access the available data and reanalyze it independently, thus sustaining the fast exchange and the advance of COVID-19-related knowledge.

Here, we will provide an overview of the current state-of-the-art of web resources specifically written or repurposed for SARS-CoV-2 research. We will cover in particular platforms designed to visualize and mine data from four major branches: epidemiology, genomics, interactomics and pharmacology/clinical studies (Figure 1 and Table 1). We specifically selected, in our list, efficient tools that are updated at least monthly, with clear user guides and/or tutorials (available at the tools web sites). In Table 2, a summary of all the features of the described tools is provided.

Epidemiologic Data Collection and Analysis Web Tools

The most established way to track a disease is to build epidemiologic maps collecting data about the number of confirmed cases, deaths and recoveries on a spatial and temporal scale. 'Dashboards' have been the most common and popular tools in this early phase, as user-friendly visualizers of geographic maps highlighting cases and deaths country- and region- wise.

One of the first dashboards to be delivered to the public was the COVID-19 Dashboard developed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [23]. The JHU dashboard offers an interactive web-based world map showing the real-time location and the number of confirmed COVID-19 cases, deaths and recoveries for all affected countries (Figure 2A). In addition to that, it offers a snapshot of local and global incidence rates, number of tests, case fatality ratios and rates of hospitalization that make it easy to compare different responses across the countries. It also allows to assess trends in specific countries and states, thus allowing us to identify areas of exponential increase in case numbers. The JHU dashboard is updated daily, and all the collected data are shared through a GitHub repository (https://github.com/CSSEGISandData/COVID-19). A similar tool, the Demographics by State COVID-19 Reporting (DSCovR) Dashboard, has been developed by Columbia University. It is an interactive Shiny web app [24] for the collection and visualization of COVID-19 cases in the USA, displaying the age, sex and race distributions for confirmed cases, number of hospitalized cases and deaths for a user-specified state. Trend comparisons across a maximum of three states can be performed through intuitive web interfaces. Alongside the JHU and DSCovR solutions, the World Health Organization (WHO) has also developed a public dashboard resource to globally map COVID-19 cases and deaths and compare them between geographic regions. The WHO Dashboard has a very quick and intuitive interface, intended to be used by a vast public beyond scientific audiences.

Beyond dashboards, other tools have been developed to monitor COVID-19 epidemiology. One such tools, heavily focused on real-time spreadsheet-like data sharing, is Worldometers.info, a popular independent website originally intended to collect, aggregate and perform statistical analyses on demographic data from official sources. The Worldometers.info COVID-19 section contains daily updated interactive tables and graphs summarizing global and local reported cases, deaths and recoveries. It contains informative tables about symptoms and reported incubation time for SARS-CoV-2 infections, and statistics about fatality rates grouped by age, sex and existing conditions or comorbidities.

Unique functions are offered by COVID-19 Scenarios, an interactive web application that simulates COVID-19 spreading dynamics in a selected population-based upon different epidemiological assumptions as well as the effect of measures adopted to contain disease transmission [25]. COVID-19 Scenarios implements a basic age-stratified Susceptible-Exposed-Infected-Recovered (SEIR) model [26] to build projections that can be adjusted by modifying several parameters and assumptions. For example, interventions like social distancing, quarantine measures and case isolation policies can be introduced in the model to evaluate their efficacy on mitigating the severity of the outbreak on a temporal scale. Severity of reported cases, age composition of populations and other epidemiologic data can be added by users to refine their models. Since most of the lockdown measures that were taken at regional

Table 1. Web link, source and architecture of COVID-19 web tools described in this review, divided in four categories

Tool	Link	Main institution	Nation	Architecture
Epidemiology				
JHU COVID-19 Dashboard	coronavirus.jhu.edu/map.html	Johns Hopkins University	USA	Python
DSCovR	msph.shinyapps.io/dscovr_dashboard	Columbia University	USA	Shiny/R
WHO Dashboard	covid19.who.int	WHO	Worldwide	JavaScript
Worldometers	worldometers.info/coronavirus	Worldometers.info	USA	JavaScript
COVID-19 Scenarios	covid19-scenarios.org	University of Basel	Switzerland	JavaScript
Harvard COVID-19 Simulator	covid19sim.org	Harvard Medical School	USA	R
CovidSIM	covidsim.eu	ExploSYS GmbH	Germany	JavaScript
COVID-19 Trajectory viewer	apps.health-atlas.de/covid-19-grapher/	University of Leipzig	Germany	Shiny/R
COVID-19 exit strategies	scienceversuscorona.shinyapps.io/covid-exit/	Science versus Corona initiative	Worldwide	Shiny/R
Greifswald COVID-19 Simulator	kaderali.org:3838/covidsim	University of Greifswald	Germany	Shiny/R
COVID19-Tracker	ubidi.shinyapps.io/covid19	Bellvitge Biomedical Research Institute	Spain	Shiny/R
Genomics				
GISAID	gisaid.org	GISAID	Worldwide	CMS TYPO3
Nextstrain	nextstrain.org/ncov/global	University of Basel	Switzerland	Python
Covidex	cacciabue.shinyapps.io/shiny2	University of Luján	Argentina	Shiny/R
Coronapp	giorgilab.dyndns.org/coronapp	University of Bologna	Italy	Shiny/R
COVID-19 Genotyping Tool	covidgenotyper.app	University of Toronto	Canada	Shiny/R
Pangolin	pangolin.cog-uk.io	Centre for Genomic Pathogen Surveillance	UK	Python
SARS-CoV-2 Alignment Screen	macman123.shinyapps.io/ugi-scov2-alignment-screen	University College London	UK	Shiny/R
CoV-GLUE	cov-glue.cvr.gla.ac.uk	University of Glasgow	UK	JavaScript
Coronavirus3D	coronavirus3d.org	University of California Riverside	USA	JavaScript
Interactomics				
CoVex	exbio.wzw.tum.de/covex/explorer	Technical University of Munich	Germany	JavaScript
VirHostNet 2.0	virhostnet.prabi.fr	University of Lyon	France	Cytoscape web
P-HIPSTer	phipster.org	Columbia University	USA	JavaScript
Pharmacology				
COVID-19 Gene/Drug Set Library	amp.pharm.mssm.edu/covid19	Icahn School of Medicine Mount Sinai	USA	JavaScript
canSAR	corona.cansar.icr.ac.uk	CRUK Cancer Therapeutics Unit	UK	JavaScript
CORDITE	cordite.mathematik.uni-marburg.de	University of Marburg	Germany	JavaScript
COVID-19 Disease Map	covid19map.lcsb.uni.lu/minerva	University of Luxemburg	Luxemburg	JavaScript
CoV-Hipathia	hipathia.babelomics.org/covid19	Foundation for Progress and Health	Spain	Web Components
Chemical Checker	sbnb.irbbarcelona.org/covid19	Institute for Research in Biomedicine	Spain	JavaScript
Clinical Trials	clinicaltrials.gov/ct2/who_table	WHO/NIH	USA	JavaScript

Table 2. Features of COVID-19 web tools described in this review

Tool	Tags	Pros	Cons
Epidemiology			
JHU COVID-19 Dashboard	Dashboard, interactive map, trend assessment, worldwide	Frequently updated, quick assessment, worldwide analysis	
DSCovR	Dashboard, interactive map, trend assessment	Comparative region analysis, demographics included	Slow to load, focused on USA
WHO Dashboard	Dashboard, interactive map, worldwide	Comparative region analysis, easy to use, frequently updated, quick assessment, worldwide analysis	
Worldometers	Spreadsheet, worldwide	Easy to use, frequently updated, quick assessment, worldwide analysis	
COVID-19 Scenarios	Interactive simulator, worldwide	Demographics Included, high number of parameters	Non-trivial to tailor the simulation for specific regions
Harvard COVID-19 Simulator	Interactive simulator	Frequently updated	Focused on USA
CovidSIM	Interactive simulator	High number of parameters	Non-trivial to tailor the simulation for specific regions
COVID-19 Trajectory viewer	Interactive simulator	Comparative region analysis	
COVID-19 Exit Strategies	Interactive simulator	Comparison of several exit strategies	Tunable parameters are few
Greifswald COVID-19 Simulator	Interactive simulator	Predict effect of social contact reduction	Focused on specific countries and German regions
COVID19-Tracker	Case number visualizer and predictor	Frequently updated	Focused on Spain
Genomics			
GISAID	Data repository, worldwide	Database fully downloadable, frequently updated, precomputed multiple sequence alignment	
Nextstrain	Dashboard, nucleotide mutation analysis, phylogenesis, worldwide	Frequently updated, simulation of mutation spread over time, worldwide	Difficult to zoom into specific regions of the interactive phylogenetic tree
Covidex	Phylogenetic categorization	Allows user-provided data, intuitive tutorial	Works exclusively with user-provided Data
Coronapp	Amino acid mutation analysis, nucleotide mutation analysis, frequency of mutations over time	Allows user-provided data, nucleotide and protein mutations, worldwide	Slow to load
COVID-19 Genotyping Tool	Phylogenetic categorization via 2D clustering	Allows user-provided data	Analysis is very slow, maximum number of sequences is only 10
Pangolin	Phylogenetic categorization, lineage assigner	Allows user-provided data, intuitive assignment of lineage	Analysis is slow
SARS-CoV-2 Alignment Screen	Nucleotide mutation analysis	Mutation analysis can be focused on specific genomic regions or genes	Not frequently updated
CoV-GLUE	Amino acid mutation analysis, nucleotide mutation analysis, spreadsheet	Mutation analysis can be focused on specific genomic regions or genes, mutations categorized as replacements/insertions/deletions	
Coronavirus3D	Amino acid mutation analysis, 3D structure	Allows to project mutations on viral protein structures from PDB, frequently updated	

Continued

Table 2. Continued

Tool	Tags	Pros	Cons
Interactomics			
CoVex	Interactome visualizer	Allows to identify known drugs for selected target proteins	
VirHostNet 2.0	Interactome visualizer	Prediction of novel interactions on user-provided protein sequences	Analysis is slow
P-HIPSTer	Interaction list	Prediction of novel interactions using sequence- and structure-based machine learning	Not focused on SARS-CoV-2
Pharmacology			
COVID-19 Gene/Drug Set Library	Curated lists of genes and drugs	Lists can be searched, new sets can be proposed	No link with external databases
canSAR	Database of clinical trials, drugs and druggable targets	Intuitive visualization of druggable interactome, drug prediction	
CORDITE	Database of clinical trials, drugs and druggable targets	Quick search	Not frequently updated
COVID-19 Disease Map	Database of drugs and pathways	Search for relevant interactions between viral proteins and human pathways	Interactome Labels are hard to read, Not frequently updated, no examples provided, not focused on SARS-CoV-2
CoV-Hipathia	Analysis of druggable pathways affected by gene expression changes	Allows user-provided data	Analysis is slow
Chemical Checker	Database of drugs	Drugs ranked by evidence quality and quantity, frequently updated	
Clinical Trials	Database of clinical trials	Frequently updated, fully comprehensive	Not categorized by drugs

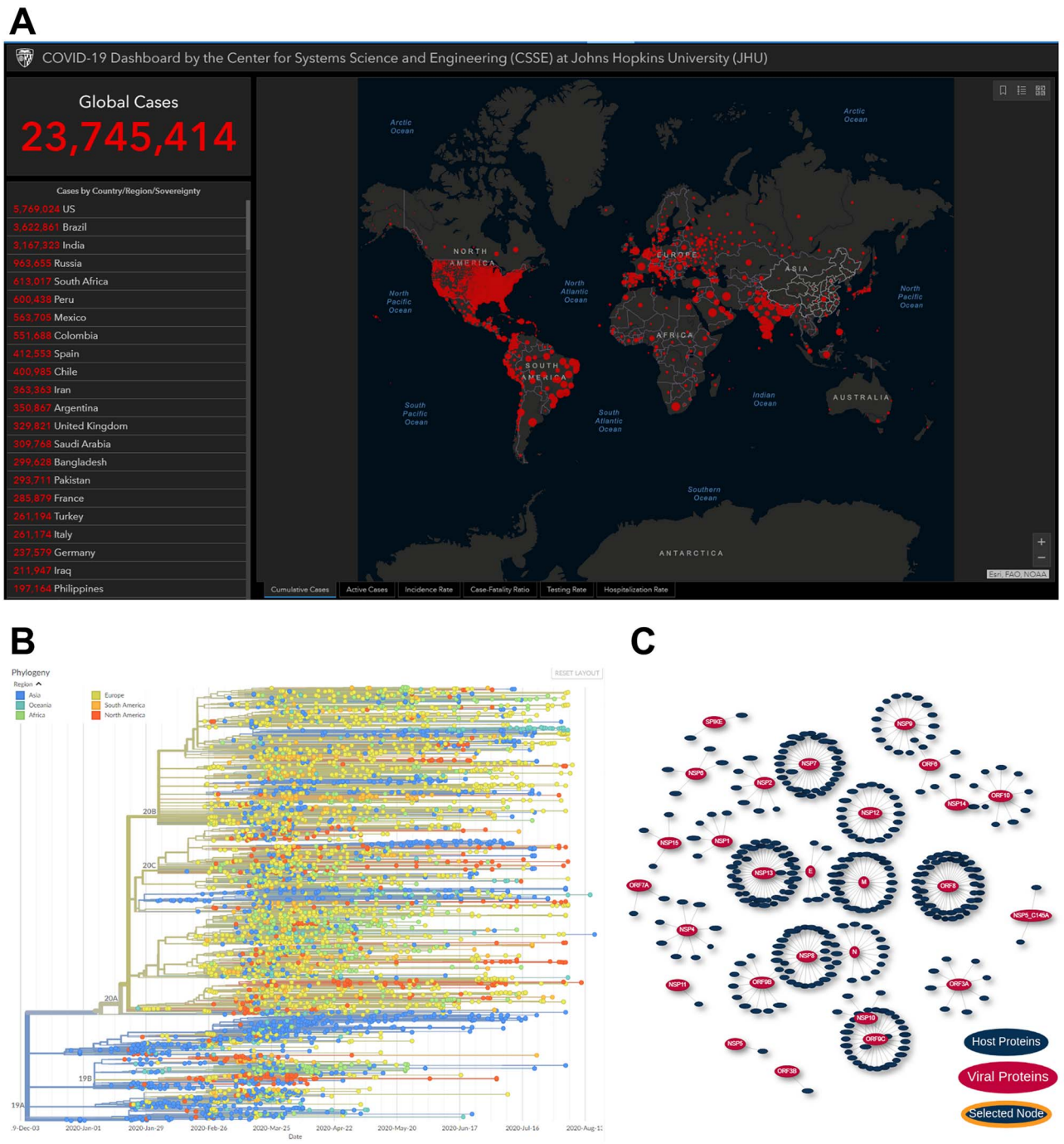

Figure 2. Screenshots of selected COVID-19 webtools described in this manuscript: the epidemiological dashboard from Johns Hopkins University (A), the time-wise phylogenetic tree representation from NextStrain (B) and the human/SARS-CoV-2 interactome visualized by Covex (C).

level were due mainly to avoid overloading of local Intensive Care Units (ICUs), COVID-19 Scenarios is a powerful tool to make projections on SARS-CoV-2 transmission and severity at the country level, also testing how different measures may mitigate risks. Demographic data can be monitored over timespans, predicting the periodicity of infections due to seasonal changes across countries. This is a very useful function since preventive social distancing and lockdown measures have a significant impact on the quality of life and economic stability of a country. All projections made through COVID-19 Scenarios can be downloaded in .json or .tsv format. Online graphs can be easily printed in .pdf through the web interface and shared through weblinks, while the app source code has been made available by developers on GitHub. A similar concept can be found in the Harvard COVID-19 Simulator, an interactive Shiny web tool to inform COVID-19 intervention policy decisions in the USA, and CovidSIM, which applies a SEIR simulation model to estimate the potential health impact of the COVID-19 pandemic on a hypothetical European country [27]. COVID-19 trajectory viewer is another similar tool to monitor cases over time and predict

trajectories in the future [28]. The common meaning of trajectory tools is to evaluate how different countries are responding to the disease spreading in order to 'flatten the curve' i.e. to keep infection rates manageable to help local public health services. Predicting the effects of different strategies to respond to COVID-19 is the focus of the COVID-19 Exit Strategies, which shows the result of several overlapping approaches to lift lockdown in terms of the number of cases, such as radical reopening, contact tracing and intermittent reopening/lockdown.

Regional but high-quality tools are also available. For example, a sophisticated web tool in German and English language, called Greifswald COVID-19 simulator, can depict epidemic curves across a selection of countries and German regions, and predict the effects of social contact reduction policies on case numbers and hospital capacity. Another regional tool is COVID19-Tracker, a Shiny web app specifically designed to collect and visualize epidemiologic data from Spain. Data on COVID-19 diagnosed cases, ICU admissions and mortality are automatically collected daily, and data trends can be visualized for each of the different implemented analysis, including case fatality rates, infection time (i.e. the time required from exposure to COVID-19 diagnosis), the effects of intervention measures in containing the outbreak, 3-days trend projections over newly diagnosed cases and deaths and reproduction number (R0) estimation. Analyses can be sorted by age and geographical region (community) [29]. COVID19-Tracker Graphical User Interface supports three languages: Spanish-Castilian, Catalan and English.

Genomic Analysis Web Tools

The 29 903 nucleotides genomic sequence of SARS-CoV-2 was shared on 5 January 2020 through GenBank (MN908947.3/-NC_045512.2). It was annotated based on sequence similarity to other coronaviruses [30]. This led researchers all around the world performing sequence analysis to compare the originally isolated virus with local cases, in search for mutations that may trace back cases, both spatially and temporally, to further explain potential differences in transmission and overall case fatality rates reported across countries [18]. Like many other RNA viruses, SARS-CoV-2 naturally accumulates mutations. While most changes have limited to no effects, some mutations may affect viral biology; for example, the Spike protein D614G mutation has been associated with a higher case/fatality rate [31]. Detecting and tracking the emergence of those mutations defining viral clades or subtypes is essential to ameliorate our knowledge on the viral biology, epidemiology, evolution and pathogenesis. The emergence of subtypes, in fact, may alter the interactions between viral particles and host, thus affecting the development of effective antiviral drugs and therapies. Genomic analyses have also had a critical role in the design of *ad hoc* molecular diagnostics to ameliorate the reliability and ease of use of kit and tests for viral detection, supporting decision making in disease management, especially in the case of host genetic heterogeneity [10].

SARS-CoV-2 sequence data have been publicly available through different sources. Most of the analyses that have been made worldwide rely on sequences shared through the GISAID Initiative database (Global Initiative on Sharing All Influenza Data) [8]. The GISAID platform was launched in 2008 with the aim of helping researchers focused on virology/epidemiology by the sharing of data collected between the WHO Collaborating Centers and National Influenza Centers. Available data on the GISAID portal include genetic sequence and related clinical and epidemiological data, as well as geographical information linked to the submitting laboratories. This platform greatly supported researchers in understanding SARS-CoV-2 origin, evolution and spread since the first evidence that a novel virus was circulating in China. Access to data is granted upon registration, and it includes all submitted sequences in FASTA format, as well as precomputed multiple sequence alignments and comprehensive sample description tables (including date and region of collection and clinical parameters of the host).

The exceptional sharing of genomic data during the COVID-19 pandemic, with more than 50 000 SARS-CoV-2 genomes publicly available on the GISAID repository, has fueled the development of close to real-time data visualization and analysis tools. One of the first and most known platforms specifically developed for real-time tracking of pathogen evolution is the Nextstrain project [32]. Nextstrain combines a daily updated database of viral genomes, a bioinformatics pipeline for phylodynamics analysis, and an interactive visualization platform into a single open web application (Figure 2B), to assist virologists and epidemiologists to share and analyze data. The architecture of Nextstrain is well designed and responds to the need for a continual surveillance to prevent uncontrolled outbreaks. The existence of Nextstrain allowed researchers from across the world respond promptly to the emergence of COVID-19 and to track the population movements associated with outbreaks via a combination of time, geography and mutation data [33]. Many other web tools rely on Nextstrain architecture to visualize data. Nextstrain source code (mainly JavaScript and Python) is shared on GitHub. It is worthwhile to highlight that the GISAID and Nextstrain teams are closely collaborating: upon connection via the GISAID website, users can browse phylogenetic trees and graphs produced by the Nextstrain Next-hCoV-19 interactive web interface, which provides updated phylogenetic and mutational analysis of GISAID-shared data.

Viral sequence subtype classification can be computationally demanding, especially for longer sequences. Most of the web tools for phylogenomic analyses offer an easy, fast and accurate classification of viral genomes taking advantage of different methods to maximize analysis performances [34]. Almost all of the web apps share a similar architecture and support phylodynamic analysis on preload sequence data or let users to submit their own sequences in FASTA format (multi-FASTA is frequently supported), and also offer visualization functions for the results. Covidex is a web application for phylogenetic analysis of viral genomes based on machine learning, with freely available code at https://sourceforge.net/projects/covidex/. Users can train their own models by uploading data on the Covidex model generator which implements a random forest classifier over a k-mer database, or run their analyses using Covidex-preloaded models [35]. Coronapp is an interactive Shiny based web-application for fast detection and monitoring of SARS-CoV-2 mutations worldwide [36]. The coronapp engine generates within seconds a mutational map using GISAID data or user-submitted sequences, then visualizes the results producing bar plot graphs summarizing most frequent mutations detected and showing the abundance and frequency of each individual mutation in an interactive dot plot. Each analysis can be visualized on a global scale or country-wise, and the results are made available for download in tabular format. The source code of coronapp is available on GitHub. Similar functions are also offered by the COVID-19 Genotyping Tool (CGT), where sequence similarity is visualized in interactive maps by Uniform Manifold Approximation and Projection [37] or Minimum Spanning Trees [38] of sequence

networks [39]. Results can be visualized by filtering for sample-associated metadata, which includes information on where and when the samples were taken, and, if available, also the travel history of the individual whose sample was collected from. CGT allows users to upload up to 10 complete COVID-19 sequences in FASTA format to be processed with preloaded GISAID data, but processing may take several minutes before visualization.

Pangolin (Phylogenetic Assignment of Named Global Outbreak LINeages) is currently the leading web resource to monitor the current status of worldwide SARS-CoV-2 mutations, allowing also the annotation and phylogenetic localization of user-provided sequences. One of the two currently most popular SARS-CoV-2 lineage nomenclatures was defined by the Pangolin analysis [40], while another one has been proposed by GISAID. The main viral clades identified by both systems are the same, and there exists a 1:1 correspondence between the two nomenclatures [18]. Visualization of the results is made possible by the connection with Microreact, a web application to display datasets through trees, geographical, temporal and associated metadata views [41]. A snapshot of the current distribution of genomic diversity can also be obtained through SARS-CoV-2 Alignment Screen, an R Shiny interactive web resource focused on visualizing specific viral genes [42]. Pangolin's whole source code is available at https://github.com/hCoV-2019/pangolin.

Amino acid sequence variations detection is one of the main features of CoV-GLUE, another web application offering the possibility for users to browse GISAID genomic data comparative analysis by filtering the results using country, collection date and phylogenetic lineage as criteria. Changes within the virus genome are organized in three separate tabs (replacements, insertions and deletions), that can be further filtered by user-selected criteria including viral protein, codon position, sequence distance, the number of sequences in which the replacement is found and the type of occurring amino acid substitutions. CoV-GLUE provides phylogenetic classification and mutation analysis, with an interactive phylogenetic tree showing the placement of the submitted sequences within the global reference evolutionary model. The final analysis also includes a diagnostic primer/probe design report to detect polymorphisms which may reduce the effectiveness of detection assays that can be very useful when locally detected sequences are analyzed and monitored [43].

Changes in SARS-CoV-2 genomic sequence may affect viral proteins 3D structures and functions. Unique features are offered by Coronavirus3D server [44], a web-based viewer for 3D visualization and analysis of the SARS-CoV-2 protein structures from the Protein Data Bank (PDB) integrating information on SARS-CoV-2 mutational patterns retrieved from China National Center for Bioinformation. Since changes in SARS-CoV-2 genomic sequence may affect viral proteins 3D structures and functions, this tool may be particularly interesting to identify and model mutations causing peculiar phenotypes, like more transmissible or aggressive strains, or to test the hypothesis for insurgence of viral subtypes showing altered recognition patterns or altered protein functions. Interactive visualization of the selected structure or models with coloring options according to the mutation frequency is provided.

Interactomics Web Tools

One of the most urgent challenges of the COVID-19 emergency is the identification of effective antiviral drugs and therapies to limit transmission and manage the most severe cases. Understanding the molecular mechanisms driving viral-host interactions is crucial to identify potential target and drug candidates, or to foster vaccine and drug development. The identification of Protein–Protein Interactions (PPIs), PPI networks or protein-drug interaction networks greatly facilitates the discovery of potential targets and drug candidates, particularly following *in silico* drug repurposing approaches [45], sustained by characterization of SARS-CoV-2 protein structures and complexes, and by antiviral drugs or inhibitors, measured experimentally or inferred by molecular docking analysis [46].

PPI network analysis approaches have been adopted at the early onset of COVID-19 pandemic [20] and now can be easily performed by using several different web tools. For example, CoVex (CoronaVirus Explorer) is a systems medicine-oriented web platform integrating known coronavirus PPIs and drug-protein interactions into an interactive large-scale interactome (Figure 2C). Users may search the human interactome for druggable targets or repurposable drug candidates (or a combination of the two) by querying the interactive map and obtaining tables describing potential drugs. Each drug receives a score rating the relevance of the selected drug to the target, together with information about FDA approval, ongoing or closed clinical trials and relevant literature information. Alternatively, users can upload a list of targets as starting seeds to guide the analysis [47]. CoVex source code is available upon request. A manual curation of Coronaviridae/host PPI networks is at the base of the COVID-19 updated version of VirHostNet 2.0, a Cytoscape-based web library providing access to a complete resource of virus–virus and virus–host PPIs and network visualization [48]. The sources of VirHostNet 2.0 interaction database are manually curated and derive primarily from the Krogan lab human/SARS-CoV-2 experimental interactome [17]. Three main query modes are available in VirHostNet: a simple query for protein/domain/pathway/taxonomy/literature search, browsing the database by viral family or species of interest, or by prediction and visualization of virus-virus and virus-host PPIs from raw primary protein sequences, submitted in FASTA format, which is particularly of interest when investigating possible intra-host drug-resistant viral quasispecies. The P-HIPSTer (Pathogen-Host Interactome Prediction using STructurE similaRity) algorithm gathers information from several viruses and drugs to infer potential PPIs for the user-selected virus (SARS-CoV-2 is one of the options). It exploits both sequence- and structure-based information to infer interactions between pathogen and human proteins. The P-HIPSTer database can be accessed through the web interface by selecting a virus or a human protein of interest to obtain a table containing known virus-host PPIs, also indicating which of the results have PDB structures available [49].

Pharmacology Web Tools

Since the first cases of COVID-19 have been described, an increasing number of papers and preprints have become available reporting *in silico* and *in vitro* testing of potential drug candidates, as well as clinical experiences of drug repurposing, leading to a potentially harmful confusing landscape considering how media have overexposed several potential remedies. An attempt to organize this rapidly evolving knowledge, to make it more suitable for researchers, is the main purpose of the COVID-19 Gene and Drug Set Library [50]. This site gathers NGS and transcriptomic experiments to offer a collection of differentially expressed genes in coronavirus-infected human cells together with gene and drug sets related to COVID-19

research delivered through a web-based interface. Currently, users can browse among 420 different gene sets (and a total of 16 765 unique genes) and 120 drug sets (1397 unique drugs), download the original datasets of interest and contribute to expand the COVID-19 Gene and Drug Set Library by submitting their own datasets. While supported by a large database, the web interface is not providing links to external databases, which could provide extra information on drugs (e.g. DrugBank [51]) or genes (e.g. GeneCards [52]).

Datasets can be also analyzed for overlapping by automatically plotting Venn diagrams, while tables and ranking functions are also available. A similar effort has been made in developing canSAR, an integrated knowledge database that collects and organizes multidisciplinary data, comprising biological, chemical, pharmacological and clinical annotation data. Machine learning approaches such as polynomial regression and neural networks are then applied to provide drug discovery predictions [53]. Users can enter proteins/diseases or even structures of interest in the query form, which returns several results that can be further refined by setting the advanced search options. CORDITE (CORona Drug InTERactions database) is another database collecting curated data on potential druggable targets coming from published articles and preprints considering both original research articles, reviews and comments [54]. CORDITE makes collected data accessible via a webserver and an open API to programmatically access to the database. Users can easily browse across interactions, targets, drugs, publications and clinical trials, to carry out meta-analyses on potential drug candidates for clinical trials or to identify effective treatments by integrating the information coming from this curated database into other software or apps. A similar attempt guides the COVID-19 Disease Map, a collective effort bringing together clinical researchers, life scientists, pathway curators, computational biologists and data scientists to build a knowledge repository of molecular mechanisms of COVID-19 summarizing host-pathogen interactions specific to SARS-CoV-2. The COVID-19 Disease Map offers a platform for visual exploration of molecular processes involved in SARS-CoV-2 entry, replication and host-pathogen interactions, as well as immune response and host repair mechanisms, suggesting potential targets of intervention, together with possible interacting drugs [55]. The same effort brought to the development of CoV-Hipathia (COVID-19 Pathway Interpretation and Analysis), a simple method to investigate rewiring of human cell signaling pathways following SARS-CoV-2 infection. The CoV-Hipathia web tool shows which druggable pathways are relevant for the viral cycle and are affected in user-provided gene expression data [56].

More clinically oriented tools also exist. One is Chemical Checker, a daily updated database collecting a literature-curated list of bioactive chemical compounds potentially effective in treating COVID-19. Chemicals are catalogued via evidence levels: text mining search, computational, preclinical and clinical. Furthermore, Chemical Checker categorizes compounds via their biochemical/biological mode of action, cataloguing them as immunomodulators, protease inhibitors, polymerase inhibitors, virus entry blockers and so on. A full list of currently ongoing or completed COVID-19 clinical trials (~1800) has been compiled by the WHO and can be searched via the *ad hoc* extension of the US National Institute of Health's ClinicalTrials.gov website. This allows the sorting of clinical studies by drug or procedure name, recruitment status, date, among other criteria.

Discussion

Despite early and constant warnings about the risk of bat-derived zoonosis viruses after the 2003 SARS epidemic [57] SARS-CoV-2 has found the world health systems unprepared to deal with it. SARS-CoV-2 rapidly evolved into a global health and economic challenge no country was prepared to face, from a Public Health Emergency of International Concern on 30 January 2020 to a global pandemic on 11 March 2020 by the WHO.

As the emergence of a novel coronavirus arose quickly, data scientists and bioinformaticians have promptly responded to the COVID-19 pandemic by developing SARS-CoV-2 specific tools to support research and advancing knowledge on biology, pathology, epidemiology and treatment of COVID-19, building on and taking inspiration from preexisting platforms like Nextstrain and the GISAID Initiative.

In this review, we bear testimony of the current effort performed by bioinformatics in fighting SARS-CoV-2 in the form of online tools. We were surprised by the quality, speed of development and abundance of the current resources, showing a vital field pervaded by a spirit of sharing of data, code and ideas. Virtually all web tools currently available freely provide the underlying data, in the form of genomic sequences, epidemiological tables, interaction maps and clinical information and the majority of the tools also provide the underlying code for data visualization. We noticed also a commendable effort by all developers to provide tools that are easy to use also for non-computational users, allowing a vast audience to quickly analyze COVID-19 data. This is having and will have indubitable positive effects in the current research effort against the pandemic, as it allows to access information that can have immediate political and social applications (based on epidemiology data and models) and on clinical decisions (based on the integrated mutational/pharmacological/Interactomics data available).

A shared limitation of all tools described here is the lack of full data coverage on the COVID-19 pandemic: except for a few countries, testing is not performed on the totality of the population and genomic data are even scarcer, especially in less developed countries [58]. While some web tools are more developed and providing a higher quantity of information than others, we believe that each resource presented here (Table 1) is uniquely useful in a specific task (summarized in Table 2) and we will leave to the readers the decision of which tool is 'best' for their particular scientific enquiry.

It is notable that, while the majority of currently existing web tools to analyze SARS-CoV-2 utilize custom JavaScript frameworks, roughly a third of them are written using R/Shiny solutions [24]. Apart from the technical advantages of the Shiny server (speed, ease of deployment, stability, integration with JavaScript), this is probably the result of the current worldwide popularity of the R language. R provides tools to perform biological sequence analysis, through the Bioconductor initiative [59], excellent data visualization methods and high-performance machine learning algorithms. Beyond our analysis, several non-web R packages have been developed in the first months after COVID-19 outbreak beyond webtools, such as the R packages (deposited on the CRAN repository [60]) coronavirus, COVID19, covid19italy and covid19.analytics, which provide daily updates on COVID-19 case numbers from several official sources. An exhaustive list drawn for COVID-19 collected more than a 100 R resources already in March 2020 (https://www.statsandr.com/blog/top-r-resources-on-covid-19-coronavirus/#coronavirus).

While many of these resources are prototypes and roughly 50% are not frequently updated, their quick deployment stands as proof of the very vital status of the R bioinformatics community.

Furthermore, the web development effort of the bioinformatics community against COVID-19 is not limited to the generation of data analysis and visualization tools. For example, one popular resource that goes beyond the scope of our manuscript is PlatCOVID (https://platcovid.com/), an online website/community where scientists share literature and information about COVID-19. The PlatCOVID web portal collects, classifies, clusters and ranks all literature related to SARS-CoV-2, allowing users to actively comment on each new publication.

While peer-reviewed publications and clinical trials must remain the prime source for clinical decisions and research facts, online tools are speeding up the sharing of COVID-19 real-time information, thus allowing for quick situational assessments and a faster clinical and scientific response to e.g. mutational events. In this historical conjuncture, the combination of the internet, large data collection, worldwide bioinformatical analysis efforts and machine learning algorithms are improving our ability to manage the current COVID-19 pandemic. The beneficial effects of web tools will limit the economic and clinical burden in every country and will constitute a technological basis to fight more COVID-19 outbreaks and future epidemics.

Key Points

- COVID-19 has stimulated the fast creation of dedicated web tools.
- COVID-19 web tools are mostly providing the means to analyze data from epidemiology, genomics, interactomics and pharmacology.
- We provide here a selection of popular and useful tools to analyze publicly available (and, in some cases, user-provided) COVID-19 data.

Funding

This work was supported by the Italian Ministry of University and Research, Montalcini 2016 grant.

References

1. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol* 2020;**92**:522–8.
2. Phua J, Weng L, Ling L, *et al*. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *Lancet Respir Med* 2020;**8**:506–17.
3. Coronavirus disease (COVID-19) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
4. Gerli AG, Centanni S, Miozzo MR, *et al*. COVID-19 mortality rates in the European Union, Switzerland, and the UK: effect of timeliness, lockdown rigidity, and population density. *Minerva Med* 2020.
5. Chen J, Lu H, Melino G, *et al*. COVID-19 infection: the China and Italy perspectives. *Cell Death Dis* 2020;**11**:1–17.
6. Le TT, Andreadakis Z, Kumar A, *et al*. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* 2020;**19**:305–6.
7. Grubaugh ND, Ladner JT, Lemey P, *et al*. Tracking virus outbreaks in the twenty-first century. *Nat Microbiol* 2019;**4**:10–9.
8. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill* 2017;**22**.
9. Gopalakrishna G, Choo P, Leo YS, *et al*. SARS transmission and hospital containment. *Emerg Infect Dis* 2004;**10**:395–400.
10. Knyazev S, Hughes L, Skums P, *et al*. Epidemiological data analysis of viral quasispecies in the next-generation sequencing era. *Brief Bioinform* 2020.
11. Xiao M, Liu X, Ji J, *et al*. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med* 2020;**12**:57.
12. Sánchez-Pacheco SJ, Kong S, Pulido-Santacruz P, *et al*. Median-joining network analysis of SARS-CoV-2 genomes is neither phylogenetic nor evolutionary. *Proc Natl Acad Sci USA* 2020;**117**:12518–9.
13. Forster P, Forster L, Renfrew C, *et al*. Reply to Sánchez-Pacheco *et al*., Chookajorn, and Mavian *et al*.: explaining phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci USA* 2020;**117**:12524–5.
14. Bauer DC, Tay AP, Wilson LOW, *et al*. Supporting pandemic response using genomics and bioinformatics: a case study on the emergent SARS-CoV-2 outbreak. *Transbound Emerg Dis* 2020;**67**:1453–1462.
15. Irwin KK, Renzette N, Kowalik TF, *et al*. Antiviral drug resistance as an adaptive process. *Virus Evol* 2016;**2**:vew014.
16. Guzzi PH, Mercatelli D, Ceraolo C, *et al*. Master regulator analysis of the SARS-CoV-2/human Interactome. *J Clin Med* 2020;**9**:982–996.
17. Gordon DE, Jang GM, Bouhaddou M, *et al*. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020;**583**:459–68.
18. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol* 2020;**11**:1800–1812.
19. Wang J. Fast identification of possible drug treatment of coronavirus Disease-19 (COVID-19) through computational drug repurposing study. *J Chem Inf Model* 2020;**60**:3277–86.
20. Srinivasan S, Cui H, Gao Z, *et al*. Structural genomics of SARS-CoV-2 indicates evolutionary conserved functional regions of viral proteins. *Viruses* 2020;**12**:360–376.
21. Hoffmann M, Kleine-Weber H, Schroeder S, *et al*. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020;**181**:271–280.e8.
22. Yozwiak NL, Schaffner SF, Sabeti PC. Data sharing: make outbreak research open access. *Nature News* 2015;**518**:477.
23. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;**20**:533–4.
24. Beeley C. *Web Application Development with R Using Shiny*, 3rd edn. Birmingham, UK: Packt Publishing, 2016.
25. Noll NB, Aksamentov I, Druelle V, *et al*. COVID-19 Scenarios: an interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2. *medRxiv* 2020. https://doi.org/10.1101/2020.05.05.20091363.
26. Getz WM, Salter R, Muellerklein O, *et al*. Modeling epidemics: a primer and numerus model builder implementation. *Epidemics* 2018;**25**:9–19.
27. Wilson N, Barnard LT, Kvalsig A, *et al*. Modelling the potential health impact of the COVID-19 pandemic on a hypothetical European country. *medRxiv* 2020. https://doi.org/10.1101/2020.03.20.20039776
28. Loeffler-Wirth H, Schmidt M, Binder H. Covid-19 transmission trajectories-monitoring pandemic in the worldwide context. *Viruses* 2020;12:777–797.

29. Tobias A, Valls J, Satorra P, *et al*. COVID19-*Tracker: una aplicación Shiny para analizar datos de la epidemia de SARS-CoV-2 en España* [COVID19-Tracker: a shiny app to analise data on SARS-CoV-2 epidemic in Spain]. *Gac Sanit* 2020;**S0213-9111**(20)30085–30086.
30. Wu F, Zhao S, Yu B, *et al*. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;**579**:265–9.
31. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pract* 2020.
32. Hadfield J, Megill C, Bell SM, *et al*. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;**34**:4121–3.
33. Wang J-T, Lin Y-Y, Chang S-Y, *et al*. The role of phylogenetic analysis in clarifying the infection source of a COVID-19 patient. *J Infect* 2020;**81**:147–78.
34. Dewey CN. Whole-genome alignment. In: *Evolutionary Genomics: Statistical and Computational Methods*, . New York, USA: Springer, 2019, 121–47.
35. Zhang E, Gupta N, Nogueira R, *et al*. Rapidly deploying a neural search engine for the COVID-19 open research dataset: preliminary thoughts and lessons learned. 2020. https://arxiv.org/abs/2004.05125.
36. Mercatelli D, Triboli L, Fornasari E, *et al*. Coronapp: a web application to annotate and monitor SARS-CoV-2 mutations. *bioRxiv* 2020. https://doi.org/10.1101/2020.05.31.124966.
37. McInnes L, Healy J, Melville JUMAP. Uniform manifold approximation and projection for dimension reduction 2018; *arXiv*1802.03426 [cs, stat].
38. Teixeira AS, Monteiro PT, Carriço JA, *et al*. Not seeing the forest for the trees: size of the minimum spanning trees (MSTs) Forest and branch significance in MST-based phylogenetic analysis. *PLoS One* 2015;**10**:e0119315.
39. Maan H, Mbareche H, Raphenya AR, *et al*. Genotyping SARS-CoV-2 through an interactive web application. *Lancet Digital Health* 2020;**2**:e340–e341.
40. Rambaut A, Holmes EC, Hill V, *et al*. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020. https://doi.org/10.1038/s41564-020-0770-5.
41. Argimón S, Abudahab K, Goater RJE, *et al*. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genomics* 2016;**2**:e000093.
42. van Dorp L, Acman M, Richard D, *et al*. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol* 2020;**83**:104351.
43. Singer J, Gifford R, Cotten M. *et al*., CoV-GLUE: a web application for tracking SARS-CoV-2 genomic variation. 2020. doi: 10.20944/preprints202006.0225.v1.
44. Sedova M, Jaroszewski L, Alisoltani A, *et al*. Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence. *Bioinformatics* 2020. doi: 10.1093/bioinformatics/btaa550.
45. Zhou Y, Hou Y, Shen J, *et al*. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Disc* 2020;**6**:1–18.
46. Aouidate A, Ghaleb A, Chtita S, *et al*. Identification of a novel dual-target scaffold for 3CLpro and RdRp proteins of SARS-CoV-2 using 3D-similarity search, molecular docking, molecular dynamics and ADMET evaluation. *J Biomol Struct Dyn* 2020;1–14.
47. Sadegh S, Matschinske J, Blumenthal DB, *et al*. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *Nat Commun* 2020;**11**:3518.
48. Guirimand T, Delmotte S, Navratil V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res* 2015;**43**:D583–7.
49. Lasso G, Mayer SV, Winkelmann ER, *et al*. A structure-informed atlas of human-virus interactions. *Cell* 2019;**178**:1526–1541.e16.
50. The COVID-19 Gene and Drug Set Library. *Patterns (N Y)* 2020;
51. Wishart DS, Feunang YD, Guo AC, *et al*. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
52. Stelzer G, Rosen N, Plaschkes I, *et al*. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics* 2016;**54**:1.30.1–1.30.33.
53. Coker EA, Mitsopoulos C, Tym JE, *et al*. canSAR: update to the cancer translational research and drug discovery knowledgebase. *Nucleic Acids Res* 2019;**47**:D917–22.
54. Martin R, Löchel HF, Welzel M, *et al*. CORDITE: the curated CORona drug InTERactions database for SARS-CoV-2. *iScience* 2020;**23**:101297.
55. Ostaszewski M, Mazein A, Gillespie ME, *et al*. COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Sci Data* 2020;**7**:136.
56. Peña-Chilet M, Esteban-Medina M, Falco MM, *et al*. Using mechanistic models for the clinical interpretation of complex genomic variation. *Sci Rep* 2019;**9**:18937–18948.
57. Field HE. Bats and emerging zoonoses: henipaviruses and SARS. *Zoonoses Public Health* 2009;**56**:278–84.
58. Tanne JH, Hayasaki E, Zastrow M, *et al*. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. *BMJ* 2020;**368**:m1090.
59. Huber W, Carey VJ, Gentleman R, *et al*. Orchestrating high-throughput genomic analysis with bioconductor. *Nat Methods* 2015;**12**:115–21.
60. Hornik K. The comprehensive R archive network. *WIREs Comput Stat* 2012;**4**:394–8.