



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

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



An innovative user-friendly platform for Covid-19 pandemic databases and resources

Shahid Ullah^{b,*}, Anees Ullah^d, Wajeeha Rahman^b, Farhan Ullah^b, Sher Bahadar Khan^c, Gulzar Ahmad^b, Muhammad Ijaz^b, Tianshun Gao^{a,**}

^a Research Center, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen, Guangdong, China

^b S Khan Lab Mardan, Khyber Pakhtunkhwa, Pakistan

^c Department of Animal Health, The University of Agriculture, Peshawar, Pakistan

^d Kyrgyz State Medical Academy (KSMA), Kyrgyzstan

ARTICLE INFO

Keywords:

CO-19 PDB
COVID-19
Digital image
Genomic
Chemical structure

ABSTRACT

Background: The current coronavirus disease-19 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a global outbreak of a disease from a new coronavirus. Several databases have been published on this pandemic, but the research community still needs an easy way to get comprehensive information on COVID-19.

Objectives: COVID-19 pandemic database (CO-19 PDB) aims to provide wonderful insights for COVID-19 researchers with the well-gathered of all the COVID-19 data to one platform, which is a global challenge for the research community these days.

Methods: We gathered 59 updated databases since December-2019 until May 2021 and divided them into six categories: digital image database, genomic database, literature database, visualization tools database, chemical structure database, and social science database. These categories focus on taking number of functions from the images, information from gene sequences, updates from relevant papers, essays, reports, articles, and books, the data or information in the form of maps, graphs, and charts, information of bonds between atoms, and updates about events of the physical and social environment, respectively.

Results: Users can search the information of interest in two ways including typing the name of the database in the search bar or by clicking the right category directly. Computer languages such as CSS, PHP, HTML, Java, etc. are utilized to construct CO-19 PDB.

Conclusion: This article attempts to compile up-to-date appropriate COVID-19 datasets and resources that have not been compiled and given in such an accessible and user-friendly manner. As a result, the CO-19 PDB offers extensive open data sharing for both worldwide research communities and local people. Further, we have planned future development of new features, that will be awesome for future study.

Introduction

Background

COVID-19 pandemic broke out in the end of the year 2019 [1], and spread to 211 countries quickly [2,3]. It is caused by a novel virus

known as severe acute respiratory syndrome coronavirus 2. (SARS-CoV-2) [4], which is responsible for the violent (COVID-19) pandemic [5]. Researchers recently clearly stated that COVID-19 patients getting angiotensin-converting-enzyme inhibitors (ACEI) could experience better health [5,6]. The COVID-19 vaccine, considered to be an important prophylactic safety and prevention technique, is being

* Corresponding author at: S Khan Lab, Mardan, KPK, Pakistan, PhD From Huazhong University of Science and technology, Wuhan, China. Postdoctoral. Chinese academy of Science & Shenzhen University Institute of Low-dimensional Materials Genome Initiative, College of Chemistry and Environmental Engineering, Shenzhen, Guangdong 518060, P. R. China

** Corresponding author at: Currently working as an associate Prof in. Research Center, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen, Guangdong, China PhD: Department of Bioinformatics & Systems Biology, Collage of life science and technology Huazhong University of Science and Technology (HUST) Wuhan 430074, China

E-mail addresses: dridechina@gmail.com, drsbioch@yahoo.com, drskbioch@gmail.com (S. Ullah).

<https://doi.org/10.1016/j.cmpbup.2021.100031>

Received 18 May 2021; Received in revised form 18 September 2021; Accepted 24 September 2021

Available online 26 September 2021

2666-9900/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

manufactured in about 90 institutions around the world [2]. Previously we have published several databases on different research areas in well-cited international journals, such as, databases on phosphorylation animal and fungi (DBPAF), in Scientific Report journal [7], a database of circadian genes in eukaryotes (CGDB), in Nucleic Acids Research journal [8], database for protein phosphorylation sites in prokaryotes (dbPSP), in Oxford Database journal [9], Emergency Data Base of COVID-19 (EDBCO-19) [10], Database of plant Research (DBPR) [11] and so on. Taken together, we have provided a huge platform named HABDSK to scientific community in the form of eleven databases of a different research field that can be freely accessed [12] with timely updates.

Aim of the study

This article attempts to gather all the knowledge needed for the COVID-19 pandemic worldwide by analyzing papers and databases that have been published by a recent global scientists, Like the regular publishing of articles reported in the Centers for Disease Control and Prevention (CDC) database [13], statistics are reported by World Health Organization (WHO) [14], and several other databases have been online on the COVID-19 pandemic in various study areas and facets of COVID-19, each and every one has its own information and importance, however, in this crucial time, we have gathered all this data on a friendly and faster-finding platform for the global scientific community, and have grouped them into 6 categories so researchers can query for the knowledge they need in a short time. In short, we have provided a new way of searching in the form of a new feature (Fig. 2A), and will be useful for future research.

Materials and methods

Construction of database

CO-19 PDB, data is collected and constructed on the friendly finding way (Fig. 1), many keywords have been used for searching the data such as; COVID-19 database, Corona database, Virus database, etc. in several searching engines like Google, Google Scholar and especially PubMed. Computer programming platforms have been used in the construction of the database. Finally, CO-19 PDB is available for researchers to be operated easily.

Use of the CO-19 PDB

Our database offers two ways to search. Users can search by clicking on the categories that lead to the proper table of the categories with small description which is shown in the (Fig. 2A). Further clicking will lead to the official link with a small description of the needed query, shown in (Fig. 2B), furthermore clicking will give the new window of needed database which we have placed the tickmark sign as an example. In addition, insert the required database name in the search bar that is mentioned at the top of the main page and has been highlighted in the (Fig. 2C).

Results

Statistic of the database

In this work, we have focused on COVID-19 databases, and have

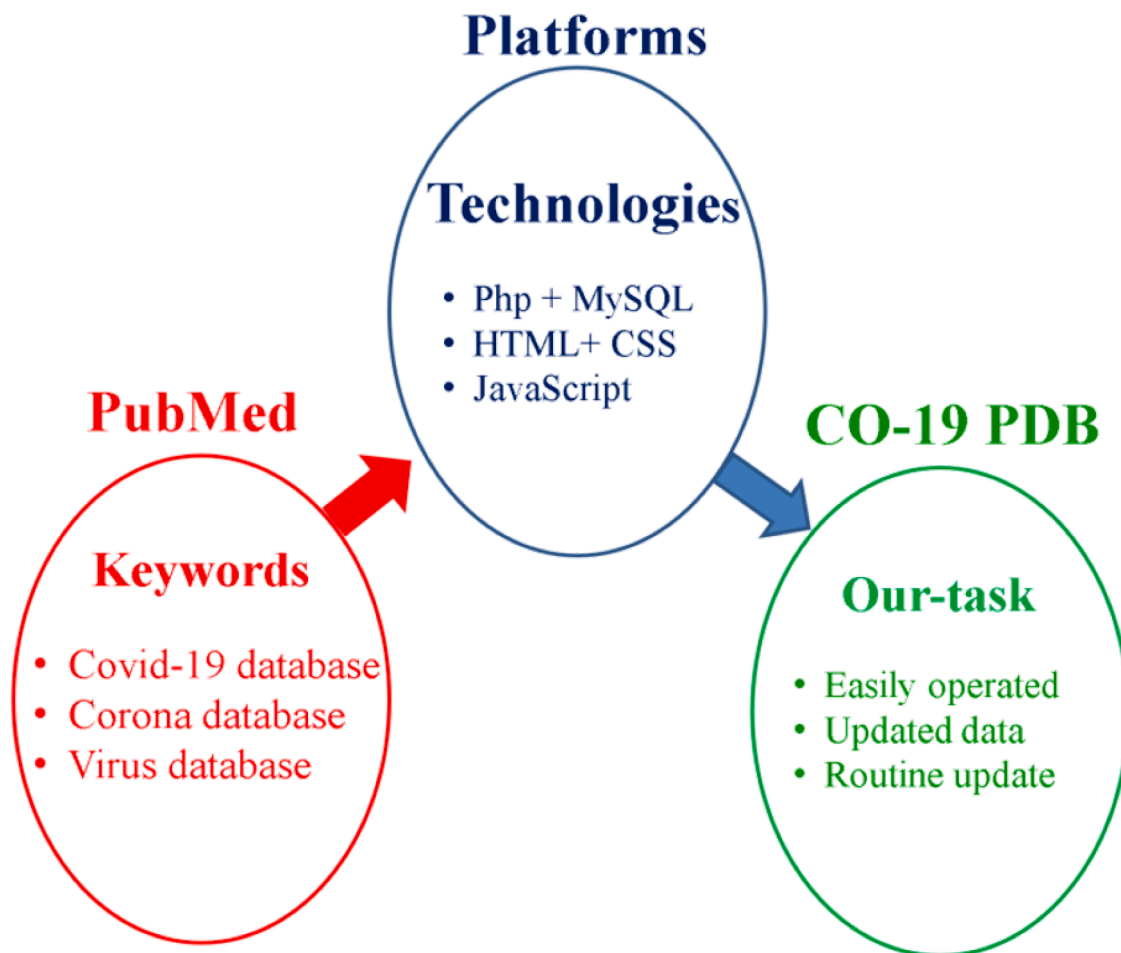


Fig. 1. The data collection and construction procedure of the CO-19 PDB.

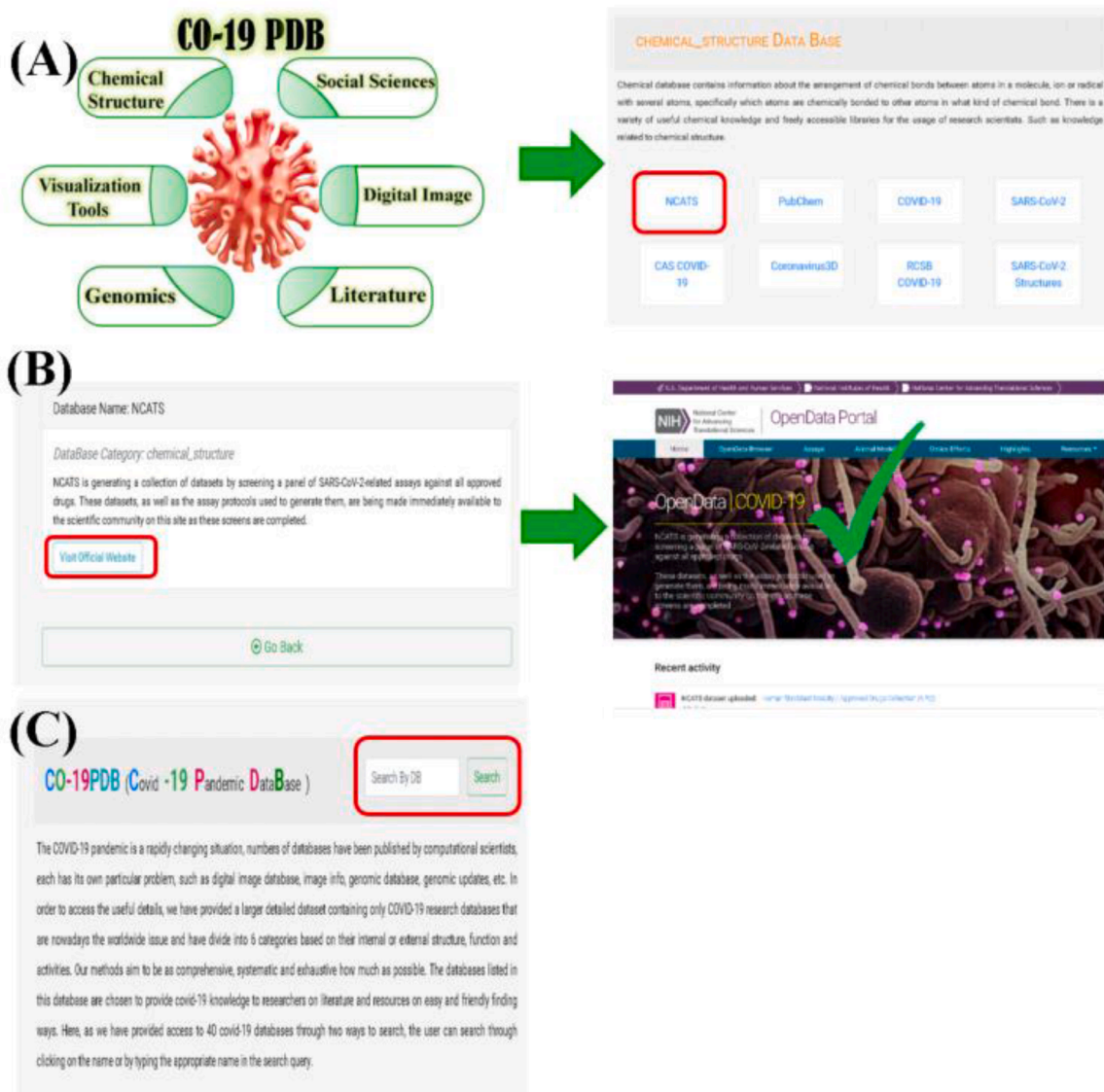


Fig. 2. The usage and statistic of CO-19 PDB, (A) Searching by clicking the name of the category. (B) Can search the name of needed database. (C) The percentage of the collected data.

collected all the database from December 2019 till May 2021, (Fig. 3A) is the growth of monthly wise databases, which show the tremendous growth with the passage of time, while (Fig. 3B) is the category wise growth of databases, in which, the literature category shown the highest value, due to vast amount of globally published literature in COVID-19 pandemic, in (Fig. 3C) the clear percentile of the data is shown, which can be of great help to further researchers, (Fig. 3D) depicts the overall number of new cases, while (Fig. 3E) depicts the overall number of deaths from December 2019 till 5 May 2021, in which the United States and India are at the top, and (Fig. 3F) depicts the top 6 confirmed cases by area, with America and Europe at the top. For more updated information and research we have provided this separate platform. It is to be noted, that all the redundant, disabled, and non-assessable database links have been updated or removed and provided new and updated COVID-19 databases which are in the form of database and table, (Table S1).

Previous published work and development of new features

Previously many articles have been published in well-known journals [15–18](Tab: 1), which have collected the databases of different

organisms and different research area, e.g. Biological databases for human research [19] which have collected 74 human databases, “Online Databases for Taxonomy and Identification of Pathogenic Fungi and Proposal for a Cloud-Based Dynamic Data Network Platform [20]” have collected 24 fungi databases, so that a well comprehensive COVID-19 database is also needed for the research community to sort and save all the COVID-19 data for future researchers [21], because nowadays database of COVID-19 has been an integral part of modern biology. Further, published work has collected the databases and has presented them in the form of a Table, while in our work we have provided the table as well as the database of the databases, named COVID-19 PD, the Comparison is given in (Table 1). Furthermore, to make easier and clearer, we have categorized the databases on the basis of their properties such as, digital image database, genomic database, literature database, visualization tools database, chemical structure database, and social science database, and have planned future development of new features, which are shown in (Fig. 2A). that have not been provided before.

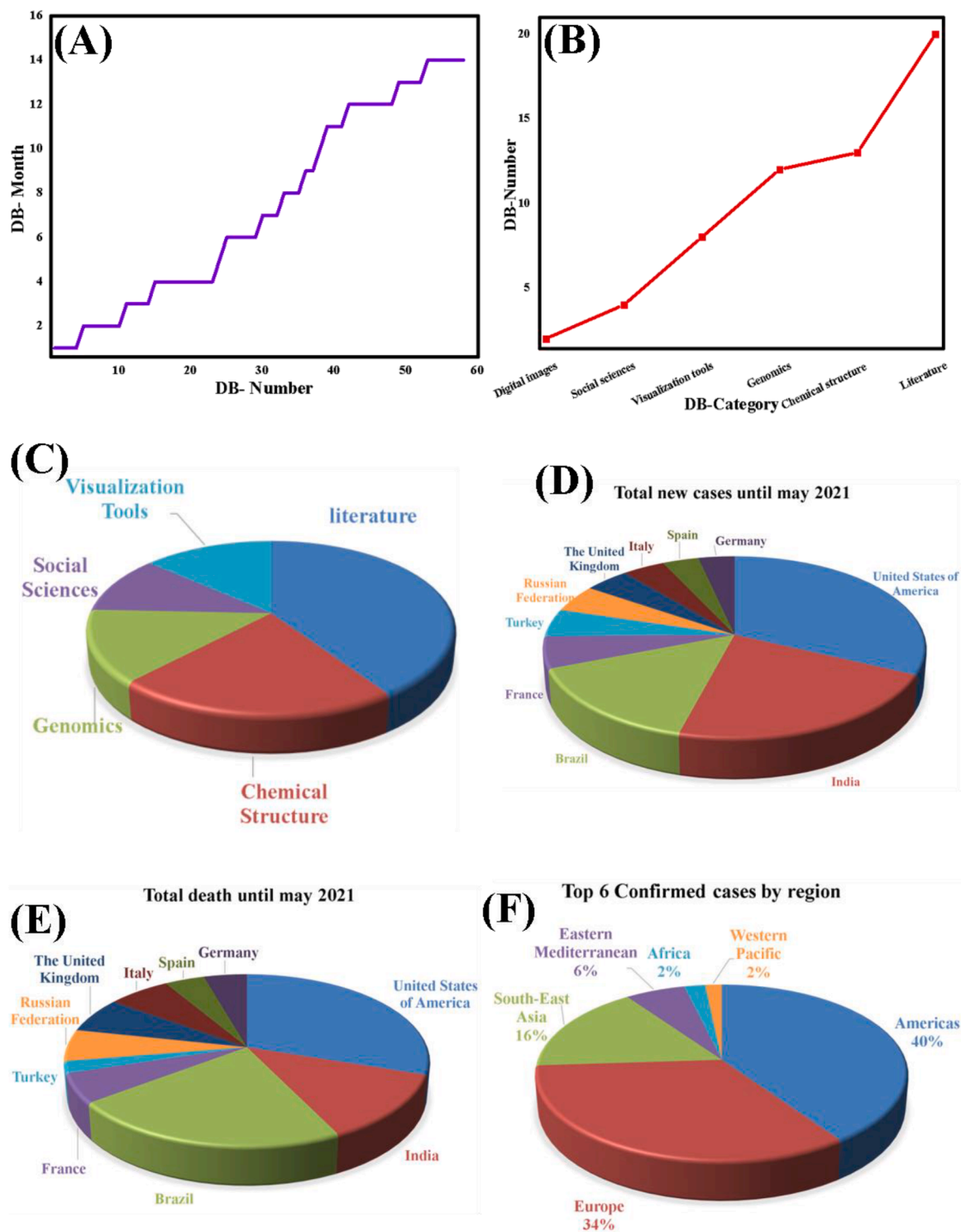


Fig. 3. The statistic of CO-19 PDB, (A) Monthly wise growth, (B) Category-wise growth, (C) The percentage of the collected data, (D) Total new cases, (E) Total death, (F) Top 6 confirmed cases by region.

Discussion

CO-19 PDB classification

The classification of various viruses is based on the selection and contrast displayed by the viruses, them to distinguish [23] based on sequence similarity [24], the molecular structure of the genome [25], pathogenicity [26], structural similarities [27,28] and host range [29]. Huge research has been done on different viruses and has been saved in

the form of published literature or databases, for an easy access and using this huge data, several databases have been released previously such as NIH, COVID-19 Data Portal and EDBCO-19. Currently on COVID-19 Pandemic, number of databases have been published, but a simplified and easier-to-use research is still needed for the conformant of the scientific community, so we have gathered revised datasets and grouped them into several categories based on their external and internal structure and function as described below, that can be access through this link <https://www.habdsk.org/co-19pdb.php>

Table 1
Comparison of CO-19 PDB with other published work.

Authors	Year	Category	Form of	DB. No	Journal name	Ref.
Our work	2021	COVID-19	DB+Table	59
Rigden and Fernández 2021	2021	Covid+other	Table	89	Nucleic Acids Res.	[22]
Rigden and Fernández 2020	2020	Nucleic acid	Table	70	Nucleic Acids Res.	[18]
Xu 2012	2012	Protein	Table	121	Curr. Protoc. Mol. Biol.	[15]
Zou, Ma et al. 2015	2015	Human	Table	74	Genomics, Proteomics Bioinf.	[19]
Harper 1994	1994	DNA+Protein	Table	50	Curr. Opin. Biotechnol.	[17]

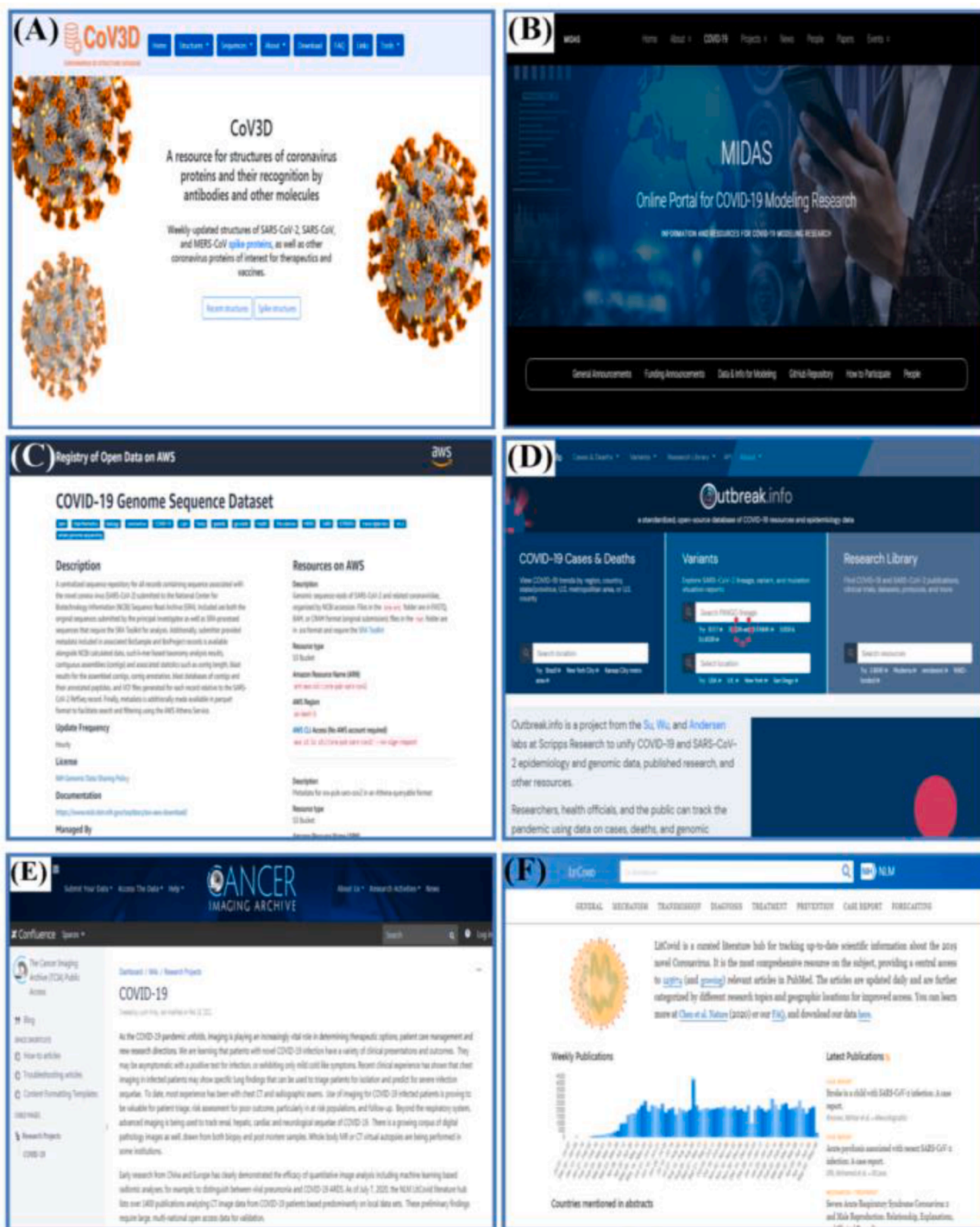


Fig. 4. The main pages of some common using databases in COVID-19 pandemic.

Chemical structure database

A chemical database contains information about the arrangement of chemical bonds between atoms in a molecule, ion, or radical with several atoms, specifically which atoms are chemically bonded to other atoms in what kind of chemical bond [30]. There is a variety of useful chemical knowledge and freely accessible libraries for the usage of research scientists. Such as knowledge related to the chemical structure [31] (Fig. 4A) shown COV3D is a weekly modified database that provides a detailed annotated collection of coronavirus protein structures and their identification by antibodies and other molecules.

Visualization tools database

Visualization tool database contains a graphical representation of any data or information. Visual elements such as maps, graphs, and charts are some of the few data visualization tools that provide the audience with an easy and accessible way of understanding the represented information. Using Visualization tools, the COVID-19 data and information can be read and generated easily and quickly [32]. (Fig. 4B) shown MIDAS is a global network of scientists who developed and applied theoretical, methodological, and mathematical models to help explain the complexities of infectious diseases in terms of pathogenesis, dissemination, efficient management methods, and forecasting.

Genomic database

Genomics is an interdisciplinary biological field that focuses on the structure, work, development, and edition of the genome. It is a complete set of DNA of an organism including all its genes [33]. Mostly it includes gene sequencing and analysis by using high-volume DNA sequencing and bioinformatics to evaluate the role and composition of whole genomes [34]. (Fig. 4C) shown COVID-19 genomic sequence database is a consolidated sequence database for all records containing sequences associated with the novel corona virus (SARS-CoV-2) that have been sent to the Sequence Read Archive (SRA) at the National Center for Biotechnology Information.

Social science database

Social science is the study of the physical and modern sphere. It is basically an analysis of international relations or the scientific study of human culture. Simply, it deals with the humans-their development, behavior, relationship, and resources they use and many kinds of organizations such as family, school, workplace, etc. it has brought awareness among the people about our environments and the incidents that happened in the past [35]. (Fig. 4D) shown main page named "Outbreak.info" is a web initiative that seeks to collect COVID-19 and SARS-CoV-2 epidemiology and genomic evidence, as well as published research and other materials.

Digital image database

A digital picture is consisting of graphic components, also known as pixels, each with a limited, distinct quantity of numerical representation due to its strength or gray color from its two-dimensional functions [36]. A digital image has the potential to improve a number of functions, such as the understanding of information for low-contrast films, the electronic transfer of images to remote installations, and the storage space requirements for archiving once treatment has been completed [37]. (Fig. 4E) shown, the cancer image archive (TCIA) is a service that de-identifies and hosts a large archive of medical image of cancer, that plays an important role in Covid-19 pandemic.

Literature database

The scientific literature contains publications that report a novel, experimental and theoretical work in the natural and social sciences, and is often abbreviated as literature within a scientific field [38]. Literature Database is an on-line, searchable bibliographic database of selected papers, essays, report, articles, and books relevant to the study of arts and cultural policy [39]. (Fig. 4F) shown "LitCovid" is a curated literature hub for monitoring up-to-date research knowledge about the newly discovered Corona virus in 2019. It is the most extensive resource on the topic, providing unified access to 125674 related PubMed articles.

Conclusion

People around the world are widely infected with COVID-19. Global scientists are working on COVID-19 and nearly 90 well-known research institutions are on record. Computational work has been done, and many datasets have been released and updated on a regular basis. Therefore, we have created a database of databases that contains all updated COVID-19 data on an easy platform and is accessible to all global researchers to aid in their research and study. To prevent wastage of time and to make it easier to find, we have classified the updated data into six categories according to their physical and chemical properties and have two ways of accessing to it, users can search by clicking on the category or enter the name of the database they need in the search bar. In short, the goal of this article is to put together up-to-date applicable COVID-19 datasets and resources that have not been gathered and provided before in such an easy and friendly finding ways. As a result, CO-19 PDB provides wide-ranging open data sharing for both global research communities and local people.

Author's contribution

Dr. Shahid Ullah and Prof. Tianshun Gao supervised the project. Dr. Anees Ullah, Ms. Wajeeha Rahman, Mr. Farhan Ullah, Dr. Sher Bahadar Khan, Mr. Gulzar Ahmad, and Mr. Muhammad Ijaz collected and verified the data carefully, all authors reviewed the manuscript and agreed to submit.

Declaration of Competing Interest

To avoid future conflict, Co-19PDB database is uploaded on (<http://www.habdsk.org/co-19pdb.php>) so that we have provided some content in this article.

Acknowledgment

To avoid future conflict, CO-19 PDB database is uploaded on (<http://www.habdsk.org/co-19pdb.php>) so that we have provided some content in this article. This project is supported by National Natural Science Foundation of China [32100434] and Shenzhen's introduction of talents and research start-up [392020].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.cmpbup.2021.100031](https://doi.org/10.1016/j.cmpbup.2021.100031).

References

- [1] H. Li, S.M. Liu, X.H. Yu, S.L. Tang, C.K. Tang, Coronavirus disease 2019 (COVID-19): current status and future perspectives, *Int. J. Antimicrob. Agents* 55 (2020), 105951, <https://doi.org/10.1016/j.ijantimicag.2020.105951> <https://doi.org/>
- [2] L. Yang, D. Tian, W. Liu, Strategies for vaccine development of COVID-19, *Sheng Wu Gong Cheng Xue Bao* (2020) 593–604.

- [3] M. Gbetuwa, U. Iqbal, RE: COVID19 impact on human behaviour in Sierra Leone, (2020).
- [4] C. Contini, M. di Nuzzo, N. Barp, A. Bonazza, R. de Giorgio, M. Tognon, S. Rubino, The novel zoonotic COVID-19 pandemic: an expected global health concern, *J. Infect. Dev. Countries* 14 (2020) 254–264.
- [5] J.S. Rico-Mesa, A. White, A.S. Anderson, Outcomes in patients with COVID-19 infection taking ACEI/ARB, *Curr. Cardiol. Rep.* 22 (2020) 1–4.
- [6] A. Humayun, M.I. Anwar, S.S. Alam, U. Iqbal, Epidemic to Pandemic: Just a Matter of Time, in: *Proceedings*, 2020, pp. 1–3.
- [7] S. Ullah, S. Lin, Y. Xu, W. Deng, L. Ma, Y. Zhang, Z. Liu, Y. Xue, dbPAF: an integrative database of protein phosphorylation in animals and fungi, *Sci. Rep.* 6 (2016) 1–9.
- [8] S. Li, K. Shui, Y. Zhang, Y. Lv, W. Deng, S. Ullah, L. Zhang, Y. Xue, CGDB: a database of circadian genes in eukaryotes, *Nucleic Acids Res.* (2016) gkw1028.
- [9] Z. Pan, B. Wang, Y. Zhang, Y. Wang, S. Ullah, R. Jian, Z. Liu, Y. Xue, dbPSP: a curated database for protein phosphorylation sites in prokaryotes, *Database* (2015) 2015.
- [10] S. Ullah, F. Ullah, I.M. Rahman, G. Ahmad, W. Ullah, EDBCO-19: Emergency Data Base of COVID-19, *J. Clin. Med. Res.* 2 (2020) 1–4.
- [11] S. Ullah, W. Rahman, F. Ullah, G. Ahmad, M. Ijaz, T. Gao, DBPR: DataBase of Plant Research, (2021).
- [12] J.M. KhanSA, V. Anand, K.A. Soomro, M. Junaid, Tremendous Contribution of Dr. Shahid Ullah to Scientific Community during COVID-19 Pandemic in the Form of Scientific Research, *J. Clin. Med. Res.* 2 (2020) 1–7.
- [13] S.B. Thacker, C.D.C. Library. COVID-19 Research Articles Downloadable Database., (2020). <https://www.cdc.gov/library/researchguides/2019novelcoronavirus/researcharticles.html>. (Accessed September 14, 2021).
- [14] World Health Organization, Global research on coronavirus disease (COVID-19), (2021). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>. (Accessed September 14, 2021).
- [15] D. Xu, Y. Xu, Protein databases on the internet, *Curr. Protocols Mol. Biol.* 68 (2004) 19.4.1–19.4.15.
- [16] G.D. Bader, M.P. Cary, C. Sander, Pathguide: a pathway resource list, *Nucleic Acids Res.* 34 (2006) D504–D506.
- [17] R. Harper, Access to DNA and protein databases on the Internet, *Curr. Opin. Biotechnol.* 5 (1994) 4–18.
- [18] D.J. Rigden, X.M. Fernández, The 27th annual Nucleic Acids Research database issue and molecular biology database collection, *Nucleic Acids Res.* 48 (2020) D1–D8.
- [19] D. Zou, L. Ma, J. Yu, Z. Zhang, Biological databases for human research, *Genomics Proteomics Bioinformatics* 13 (2015) 55–63.
- [20] P.Y. Prakash, L. Irinyi, C. Halliday, S. Chen, V. Robert, W. Meyer, Online databases for taxonomy and identification of pathogenic fungi and proposal for a cloud-based dynamic data network platform, *J. Clin. Microbiol.* 55 (2017) 1011–1024.
- [21] M.-C. Lin, U. Iqbal, Y.-C. Li, AI in medicine: big data remains a challenge, (2018).
- [22] D.J. Rigden, X.M. Fernández, The 2021 Nucleic Acids Research database issue and the online molecular biology database collection, *Nucleic Acids Res.* 49 (2021) D1–D9.
- [23] A.M.Q. King, E. Lefkowitz, M.J. Adams, E.B. Carstens, Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses, Elsevier, 2011.
- [24] P. Simmonds, Methods for virus classification and the challenge of incorporating metagenomic sequence data, *J. Gen. Virol.* 96 (2015) 1193–1206.
- [25] D. Baltimore, Expression of animal virus genomes, *Bacteriol. Rev.* 35 (1971) 235.
- [26] W.S.M. Wold, M.G. Ison, Adenovirus, in: *Fields Virology 6th Edition*, Lippincott, Williams & Wilkins, 2013; pp. 1732–1767.
- [27] C. Büchen-Osmond, Further progress in ICTVdB, a universal virus database, *Arch. Virol* 142 (1997) 1734–1739.
- [28] C. Büchen-Osmond, M. Dallwitz, Towards a universal virus database—progress in the ICTVdB, *Arch. Virol.* 141 (1996) 392–399.
- [29] J.D. Hare, H.B. Short, Manual of clinical microbiology, *Am. J. Dis. Child.* 129 (1975) 754.
- [30] L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, The chemical structure of a molecule resolved by atomic force microscopy, *Science* 325 (2009) 1110–1114.
- [31] S.S. Jick, J.A. Kaye, C. Vasilakis-Scaramozza, L.A.G. Rodríguez, A. Ruigómez, C. R. Meier, R.G. Schlienger, C. Black, H. Jick, Validity of the general practice research database, *Pharmacotherapy* 23 (2003) 686–689.
- [32] E. Graham, Introduction: Data visualisation and the Humanities, (2017).
- [33] O.M. de la S. Advisory, W.H. Organization, Genomics and world health: report of the Advisory Committee on Health Research, World Health Organization, 2002.
- [34] R.J. Ogbe, D.O. Ochalefu, O.B. Olaniru, Bioinformatics advances in genomics-a review, *Int. J. Curr. Res. Rev.* 8 (2016) 5–11.
- [35] A. Giorgi, The relationships among level, type, and structure and their importance for social science theorizing: a dialogue with Schütz, *Duquesne Stud. Phenomenol. Psychol.* 3 (1979) 81–92.
- [36] F. Patin, An introduction to digital image processing. online]: http://www.programmersheaven.com/articles/patin_ImageProc.Pdf. (2003).
- [37] R.L. Cumberlin, J.E. Rodgers, F.H. Fahey, Digital image processing of radiation therapy portal films, *Comput. Med. Imaging Graph.* 13 (1989) 227–233.
- [38] A.H. Renear, S. Sacchi, K.M. Wickett, Definitions of dataset in the scientific and technical literature, *Proc. Am. Soc. Inf. Sci. Technol.* 47 (2010) 1–4.
- [39] B.K. Lin, M. Clyne, M. Walsh, O. Gomez, W. Yu, M. Gwinn, M.J. Khoury, Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database, *Am. J. Epidemiol.* 164 (2006) 1–4.
- Anees Ullah: Bachelor in medicine: Kyrgyz state medical University, Email: anes-ullah923@gmail.com
- Wajeeha Rahman: done Master: now working as a researcher in S-Khan Lab Mardan KPK, Pakistan <http://www.habdisk.org/students.php>, Email: wajeeharehman7464@gmail.com
- Farhan Ullah: done Master: now working as a researcher in S-Khan Lab Mardan KPK, Pakistan <http://www.habdisk.org/students.php>, Email: Farhanu686@gmail.com
- Sher Bahadr khan: Assistant professor: Department of Animal Health, The University of Agriculture, Peshawar, Pakistan, Email: drsher@aup.edu.pk
- Gulzar Ahmad: done Master: now working as a researcher in S-Khan Lab Mardan KPK, Pakistan <http://www.habdisk.org/students.php>, Email: gulzarbiotech@gmail.com
- Muhammad Ijaz: done Master: now working as a researcher in S-Khan Lab Mardan KPK, Pakistan <http://www.habdisk.org/students.php>, Email: m.ijaz.biotech@gmail.com