*Briefings in Bioinformatics*, 2025, **26(2)**, bbaf084
https://doi.org/10.1093/bib/bbaf084
**Review**

# VITALdb: to select the best viroinformatics tools for a desired virus or application

Mira Koul‡, Shalini Kaushik‡, Kavya Singh, Deepak Sharma*

Computational Biology and Translational Bioinformatics (CBTB) Laboratory, Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India

*Corresponding author. Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India.
E-mail: deepak.sharma@bt.iitr.ac.in; deepak.aiims@gmail.com.

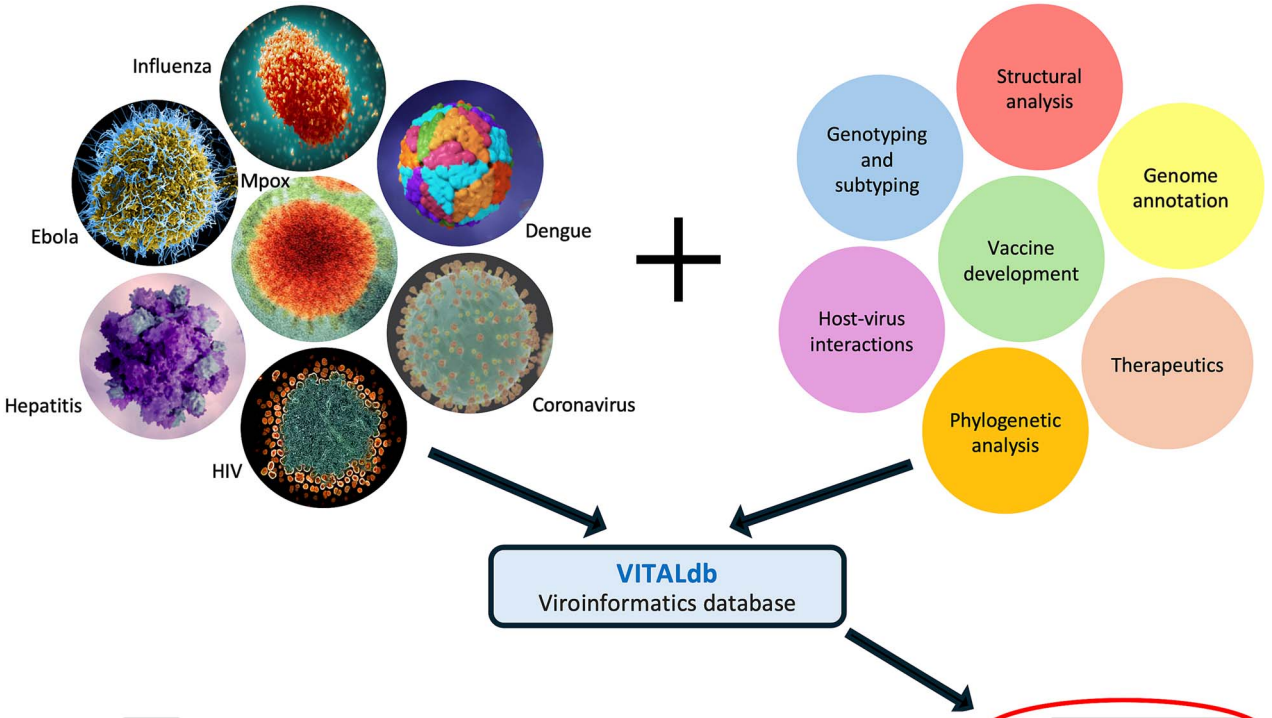
‡Mira Koul and Shalini Kaushik equal contribution.

## Abstract

The recent pandemics of viral diseases, COVID-19/mpox (humans) and lumpy skin disease (cattle), have kept us glued to viral research. These pandemics along with the recent human metapneumovirus outbreak have exposed the urgency for early diagnosis of viral infections, vaccine development, and discovery of novel antiviral drugs and therapeutics. To support this, there is an armamentarium of virus-specific computational tools that are currently available. VITALdb (VIroinformatics Tools and ALgorithms database) is a resource of ∼360 viroinformatics tools encompassing all major viruses (SARS-CoV-2, influenza virus, human immunodeficiency virus, papillomavirus, herpes simplex virus, hepatitis virus, dengue virus, Ebola virus, Zika virus, etc.) and several diverse applications [structural and functional annotation, antiviral peptides development, subspecies characterization, recognition of viral recombination, inhibitors identification, phylogenetic analysis, virus–host prediction, viral metagenomics, detection of mutation(s), primer designing, etc.]. Resources, tools, and other utilities mentioned in this article will not only facilitate further developments in the realm of viroinformatics but also provide tremendous fillip to translate fundamental knowledge into applied research. Most importantly, VITALdb is an inevitable tool for selecting the best tool(s) to carry out a desired task and hence will prove to be a vital database (VITALdb) for the scientific community.
Database URL: https://compbio.iitr.ac.in/vitaldb

**Received:** October 14, 2024. **Revised:** January 14, 2025. **Accepted:** February 17, 2025

## Graphical Abstract




| Tools | Key features | CI | Access | Virus/Disease | Application |
|---|---|---|---|---|---|
| GISAID | Data-sharing platform for Influenza, Dengue, Mpox and hCoV-19 viruses | 381.99 | Web | Influenza, Dengue, Mpox and hCoV-19 viruses | Database |
| NCBI-IVR | Repository about Influenza viruses | 69.39 | Web | Influenza viruses | Database |
| IRD | Comprehensive resource for the study on the influenza virus | 45.85 | Web | Influenza viruses | Database |
| FluTE* | Simulation model for influenza pandemic | 30.5 | Install | Influenza virus | Epidemiological analysis |
| INSaFLU | Server used for WGS-based surveillance of influenza virus | 12.6 | Web, Install | Influenza viruses | Genome annotation |

**Keywords**: viroinformatics; antiviral drugs; virus detection; vaccine development

## Introduction

The term 'viroinformatics' was coined as a result of the inseparable intertwining of virology with bioinformatics [1]. The bond between these two fields has become even stronger (indissoluble) due to the COVID-19 pandemic resulting in a boom of viroinformatics tools; from ~100 about a decade ago [1] to ~360 presently. The development of computational tools in virology starts with the basic analyses (comparative sequence analysis, assigning gene/protein functions, and designing primers for diagnostics or subtype identification) and the data generated is subsequently utilized to develop more complex programs (genome annotation, metagenomic analysis, genotype–phenotype correlation studies, and virus–host protein–protein interactions) [1, 2]. There is a dire need of virus-related databases and bioinformatic tools for several reasons: to find potential targets for therapeutics to cure viral infections, availability of limited treatment options, understanding the genetic diversity in viruses, developing efficient strategies to control viral diseases, and many more. The vast field of viroinformatics addresses all these issues.

Biochemical pathway analysis and functional/structural annotation are crucial to understand the structure–function relationship of proteins encoded in viral genomes that ultimately lead to detection of binding sites as novel drug targets. In addition, identifying virus–virus and virus–host interactions can also identify potential drug targets. Prediction of antiviral peptides (AVPs) can accelerate the development of antiviral drugs. Due to frequent emergence of new viruses, it is important to conduct taxonomic classification for characterizing viral composition, determining the viral strain characteristics and hierarchal clustering of viruses. Moreover, viruses undergo mutations and recombinations to generate genetic diversity. Analysing recombination patterns and investigating recombination events in viral genomes helps to minimize errors in constructing phylogenetic trees. Novel virus discovery using next-generation sequencing (NGS) and metagenomic data analysis helps in understanding the mysterious infections and controlling the spread of viral diseases.

We had earlier enumerated ~100 viroinformatics resources and categorized them based on their uniqueness for a certain virus and/or application [1]. In this article, we have immensely expanded this list to 358 viroinformatics tools (Suppl. Table 1), and subsequently built VITALdb (VIroinformatics Tools and

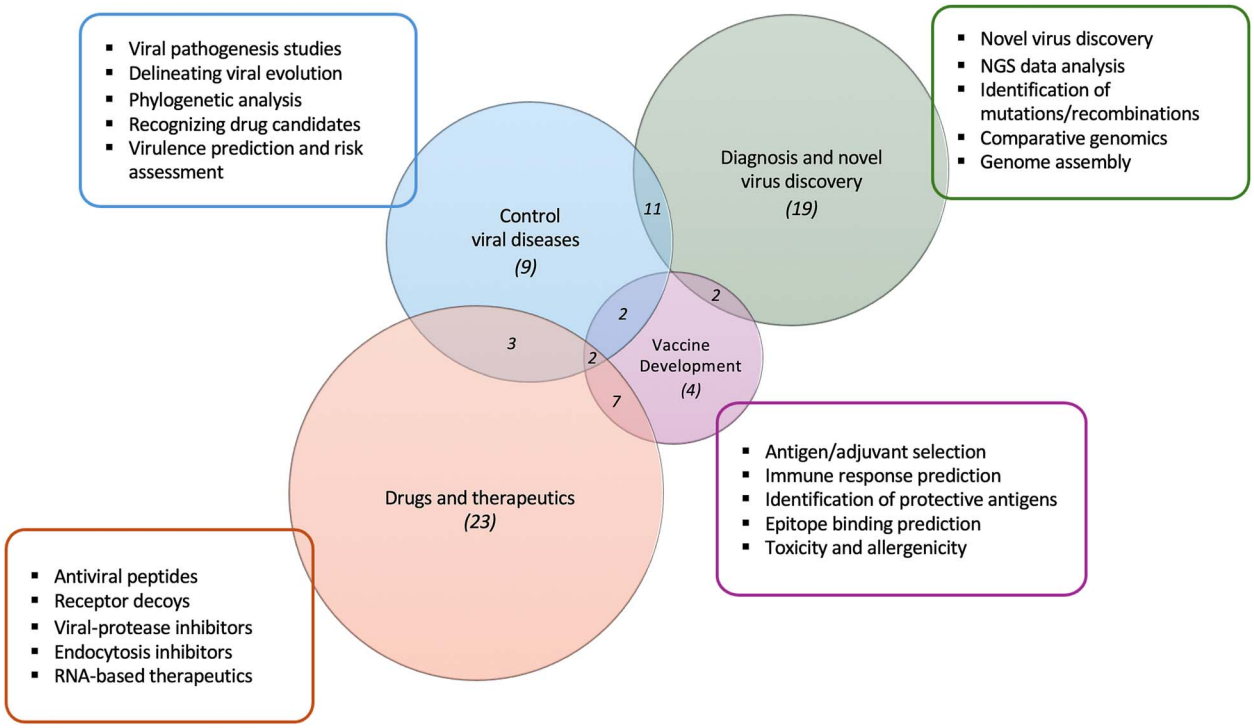



Figure 1. Utility of viroinformatics in diverse antiviral strategies (the numerals specify the number of tools).

ALgorithms database) to provide users a one-stop platform for interactive and easy access to these resources (https://compbio.iitr.ac.in/vitaldb). Of these, 116 tools have ≥10 citation index (numbers of citations per year) highlighting their importance and wide usage (Table 1). The aim of this study has many aspects that include generating awareness regarding all the diverse tools, classification of these assets according to their functionality and uniqueness for a certain virus, enabling virologists/researchers to get an overview of the important tools for specific application (such as developing novel antivirals, vaccine development, diagnosis and novel viral discovery, controlling viral diseases, and combating COVID-19), and emphasizing useful tools that remain currently unexplored (Fig. 1).

## Drugs and therapeutics

The rapid development of antiviral compounds/agents or chemotherapy offers a decisive approach to the control of any virus. They fall primarily in three classes: anti-herpes virus, anti-retrovirus, and, to a lesser extent, anti-rhino virus compounds. Ongoing viral replication, prolonged drug exposure, narrow antiviral spectrum, and ineffectiveness against the latent virus are few reasons that antiviral drugs are not sufficiently efficacious. Additionally, viruses can continuously evolve into new variants/subvariants and become resistant to existing treatments. Thus, the research for designing safe and effective drugs is constantly needed to keep up with the genetic diversity of viruses.

Therapeutic databases offer the potential to boost research in the fields of virology and drug discovery (Table 2). Resources like Virus-CKB 2.0 is the knowledgebase of 65 antiviral drugs, 178 viral-related targets and 3766 compounds identified for these target molecules [3]. On the other hand, CORDITE is a specific database of drug interactions for SARS-CoV-2 [4] while VDA-RWLRLS framework helps in prioritizing drugs [5]. AVCpred server uses the quantitative structure–activity relationship (QSAR) for predicting and designing antiviral therapeutics [6].

Natural or synthetic AVPs are novel antiviral compounds that exhibit huge potential in controlling emerging viral pathogens. These classes of molecules show less toxicity with minor side effects. miRNAs are also important targets for developing novel antiviral therapeutics. VIRmiRNA is a comprehensive repository of viral miRNAs (including 1308 experimentally validated miRNAs) and their targets [7]. AI4AVP, AntiVPP 1.0, iDVIP, AVPIden, AVPpred, ENNAVIA, and Meta-iAVP are the tools for prediction of AVPs based on machine learning algorithms [8–14]. In comparison, Deep-AVPpred and PandoraGAN, identifies novel AVPs in protein sequences (predicts whether peptide will function as AVP or not) and designs bioactive AVPs, respectively [15, 16]. AVPdb is a resource of experimentally verified AVPs to help researchers in the development of antiviral therapeutics. It targets more than 60 medically important viruses including SARS, influenza, hepatitis B virus (HBV), HCV, DENV, RSV, and HSV [17]. For developing new drugs/therapeutics, it is crucial to understand the drug binding sites in viral proteins, drug resistance mutations in viral genes and molecular mechanism of pathogenesis (for instance, identification of phosphorylation sites). RVDB-prot provides a well-curated collection of protein sequences (Hidden Markov Model profiles) equivalent of nucleic acid data from RVDB [18]. Geno2pheno predicts phenotypic drug resistance in HIV-1, HBV, and HCV [19] while Stanford HIV Drug Resistance DB monitors drug resistance against antiretroviral drugs against HIV [20]. ViralPhos predicts potential phosphorylation sites in viral proteins [21]. Conversely, DeepIPs uses a deep-learning algorithm to identify phosphorylation sites in host cells infected with SARS-CoV-2 [22].

Destructive tendencies and consequences of virus proteins are closely related to their subcellular localization in host cell. iLoc-Virus predicts the subcellular localization of viral proteins using a multilabel learning classifier [23]. Protein–protein interactions (PPIs) between human and virus play an important role in

Table 1. Databases, web servers, and tools for virus research with CI ≥ 10

| Tools | Key feature(s) | CI[a] | Access[b] | URL [reference] |
|---|---|---|---|---|
| AI4AVP | Antiviral peptides prediction tool | 12 | W,D | http://axp.iis.sinica.edu.tw/AI4AVP/<br>https://github.com/lsbnb/amp_gan [8] |
| AVPdb | Resource of experimentally validated antiviral peptides targeting medically important viruses | 21.8 | W | http://crdd.osdd.net/servers/avpdb [17] |
| AVPIden | Identification and functional prediction of antiviral peptides | 18.6 | W | http://awi.cuhk.edu.cn/AVPIden/ [13] |
| AVPpred | Antiviral peptide prediction algorithm | 20.16 | W | http://crdd.osdd.net/servers/avppred [10] |
| CATNAP | Tool to compile, analyse, and tally neutralizing antibody panels | 16.6 | W,D | https://www.hiv.lanl.gov/components/sequence/HIV/neutralization/ [46] |
| Cenote-Taker 2 | Virus discovery and annotation tool | 29 | D | https://github.com/mtisza1/Cenote-Taker2 [61] |
| CheckV | Pipeline for assessing the quality and completeness of metagenome-assembled viral genomes | 349.6 | D | https://portal.nersc.gov/CheckV/ [138] |
| CliqueSNV | Reconstruction technique for viral variants | 14.2 | D | https://github.com/vyacheslav-tsivina/CliqueSNV [65] |
| CoronaSPAdes | Assembler for RNA viral species recovery | 21 | W,D | http://cab.spbu.ru/software/spades [66] |
| CoV-GLUE | For tracking SARS-CoV-2 genomic variation | 40.66 | W | https://cov-glue.cvr.gla.ac.uk/index.php [102] |
| CoVex | Platform for SARS-CoV-2 host interactome exploration and drug (target) identification | 40.52 | W | https://exbio.wzw.tum.de/covex/ [37] |
| CRISPR | Tool for studying Virus–host interactions | 27.33 | D | https://www.addgene.org/crispr/libraries/ [173] |
| DBatVir | Database of bat-associated viruses | 17.58 | W,D | http://www.mgc.ac.cn/DBatVir/ [78] |
| Deep-AVPpred | Classifier for discovery of peptide drugs for viral infections | 16.6 | W,D | https://deep-avppred.anvil.app/ [15] |
| DeepHost | Tool for phage host prediction | 14.22 | D | https://github.com/deepomicslab/DeepHost [32] |
| DeepIPs | Architecture to find phosphorylation sites in SARS-CoV-2-infected host cells | 26.59 | W,D | https://github.com/linDing-group/DeepIPs<br>http://lin-group.cn/server/DeepIPs/ [22] |
| DeepVirFinder | For viral identification using metagenomic data | 94.37 | D | https://github.com/jessieren/DeepVirFinder [57] |
| ENNAVIA | Prediction of antiviral and anti-coronavirus effects for therapeutic peptides | 21.4 | W,D | https://research.timmons.eu/ennavia [11] |
| ERVmap | Tool to analyse transcription of human endogenous retroviruses | 30.8 | W,D | https://www.ncbi.nlm.nih.gov/ [136] |
| FluTE[c] | Influenza epidemic simulation model | 30.5 | D | http://www.cs.unm.edu/~dlchao/flute/ [148] |
| GATU | Annotation of viral genomes | 14.9 | W | http://athena.bioc.uvic.ca/virology-ca-tools/gatu/ [149] |
| Geno2pheno | For prediction of drug resistance in HIV-1, HBV and HCV | 15.37 | W | http://www.geno2pheno.org/ [19] |
| Genome Detective | Identification of viruses through high-throughput sequencing data | 62.6 | W,D | http://www.genomedetective.com/app/typingtool/virus/ [62] |
| GISAID | Data-sharing platform for all influenza viruses | 381.99 | W | https://gisaid.org/ [131] |
| GLUE | Software for sequence data of viruses | 18 | W,D | http://glue-tools.cvr.gla.ac.uk [76] |
| GVD | Identifying age-based patterns of virome diversity in the gut of humans | 131.5 | W,D | https://bitbucket.org/MAVERICLab/gvd/ [172] |
| HAVoC | Pipeline for reference-based consensus assembly and lineage assignment for SARS-CoV-2 sequences | 12.8 | W,D | https://www.helsinki.fi/en/projects/havoc<br>https://bitbucket.org/auto_cov_pipeline/havoc [96] |
| HBVdb | Knowledgebase for Hepatitis B virus | 17.11 | W | http://hbvdb.ibcp.fr [133] |
| HERVd | Human intrinsic retroviruses database | 10.4 | W | http://herv.img.cas.cz/ [68] |
| HEVnet | Library for improved hepatitis E virus molecular typing, characterization, and epidemiological studies | 12.69 | W | https://www.rivm.nl/mpf/typingtool/hev [134] |
| HVIDB | Repository for human–viral protein–protein interactions | 16.2 | W,D | http://zzdlab.com/hvidb/ [26] |
| iACVP | Tool for identification of anti-coronavirus peptides | 15.3 | W,D | http://kurata35.bio.kyutech.ac.jp/iACVP [111] |
| iLoc-Virus[c] | Classifier for finding the intracellular localization of viral proteins with both single and numerous sites | 20.98 | W | http://icpr.jci.edu.cn/bioinfo/iLoc-Virus [23] |
| IMG/VR v3 | Ecological and evolutionary structure for examining of the uncultivated viral genomes | 98.59 | W,D | https://img.jgi.doe.gov/vr<br>https://genome.jgi.doe.gov/portal/IMG_VR [139] |
| INSaFLU | WGS-based surveillance of influenza virus | 12.6 | W,D | https://insaflu.insa.pt/ [132] |

*(continued)*

Table 1. Continued

| Tools | Key feature(s) | CI[a] | Access[b] | URL [reference] |
|---|---|---|---|---|
| iPHoP | Host prediction for metagenome-derived viruses | 106.2 | D | https://bitbucket.org/srouxjgi/iphop/src/main/ [33] |
| IRD | Resource for influenza virus research | 45.85 | W | https://legacy.fludb.org/brc/home.spg?decorator=influenza [44] |
| IVA | *De novo* assembler for RNA virus genomes | 22.3 | W,D | http://sanger-pathogens.github.io/iva [124] |
| LANL HCV Database | Includes sequence information and immunological sites in HCV (now replaced by ViPR) | 29.6 | W | http://hcv.lanl.gov [150] |
| LSTM-PHV | Model to predict human–virus protein–protein interactions | 26.9 | W,D | http://kurata35.bio.kyutech.ac.jp/LSTM-PHV [24] |
| Meta-iAVP | Tool to improve the antiviral peptides prediction | 21.8 | W | http://codes.bio/meta-iavp/ [12] |
| Metavir 2 | For viral metagenome comparison and assembled virome analysis | 24.2 | W | http://metavir-meb.univ-bpclermont.fr [151]<br>https://www.protocols.io/view/MetaVir-Analysis-of-viromes-261ge4rwv479/v1 |
| MetaviralSPAdes | Assembly of viruses from metagenomic data | 51.4 | D | https://github.com/ablab/spades/tree/metaviral_publication [50] |
| MVP | Microbe–phage interaction database | 10.98 | W | http://mvp.medgenius.info [30] |
| Naomi | Program for calculating HIV pandemic indicators in sub-Saharan African regions | 12.3 | D | https://github.com/jeffeaton/naomi-model-paper;<br>https://github.com/mrc-ide/naomi/tree/master/inst/extdata [86] |
| NCBI Genotyping Tool | For genotyping viral sequences | 14.5 | W | http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi [152] |
| NCBI Viral Genomes Resource | Includes sequences for ∼3600 genomes of viruses | 66.9 | W | http://www.ncbi.nlm.nih.gov/genomes/VIRUSES/viruses.html [153] |
| NCBI-IVR | Offers accessibility to the influenza virus sequence data, BLAST, alignment, phylogenetic, and genome annotation tools | 69.39 | W | http://www.ncbi.nlm.nih.gov/genomes/FLU/ [154] |
| NCBI-VVR | Database about viral variations | 57.97 | W | http://www.ncbi.nlm.nih.gov/genomes/VirusVariation/ [164] |
| OncoDB | For analysing the expression of genes and infections caused by viruses in cancer | 66.17 | W | http://oncodb.org [140] |
| P-HIPSTer | Framework for pathogen host interactome prediction | 27.77 | W,D | http://phipster.org/ [25] |
| PandoraGAN | Tool to design bioactive antiviral peptides | 20.25 | D | https://gitlab.com/shraddha.surana/antiviral-peptide-predictions-using-gan [16] |
| Pangolin | For allocating the most possible lineage to a particular SARS-CoV-2 genomic sequence | 298.8 | D | https://cov-lineages.org/resources/pangolin.html [95] |
| PASC | Classification of viruses using a genome-based online tool called PASC | 18.9 | W | http://www.ncbi.nlm.nih.gov/sutils/pasc/ [81] |
| PaVE | Database of curated papillomavirus genomic sequences | 36.2 | W | https://pave.niaid.nih.gov/ [55] |
| PHACCS[c] | Assessing the diverseness of uncultured/environmental communities of viruses | 11.24 | W | http://phage.sdsu.edu/phaccs [155] |
| PhaGCN2 | Tool for virus classification for viral genomic fragments | 26.05 | D | https://github.com/KennthShang/PhaGCN2.0 [52] |
| PHIAF | For prediction of phage–host interactions | 18 | D | https://github.com/mengluli-web/PHIAF [29] |
| Phigaro | Application for high-throughput prophage sequence annotation | 30 | D | https://github.com/bobeobibo/phigaro [128] |
| pVOGs | Database for comparative genome analysis and protein families annotation | 45.42 | D | ftp://ftp.ncbi.nlm.nih.gov/pub/kristensen/pVOGs/home.html [144] |
| PVPred-SCM | Tool used for the assessment and prediction of phage virion proteins | 10.9 | W | http://camt.pythonanywhere.com/PVPred-SCM [123] |
| RaFAH | Host prediction tool for viruses of bacterial and archaeal communities based upon protein content | 24.7 | W,D | https://sourceforge.net/projects/rafah/ [120] |
| RDP5 | For studying recombination in, and eliminating indications of recombination from, nucleotide sequence datasets | 114 | W,D | http://web.cbio.uct.ac.za/~darren/rdp.html [73] |
| RdRp-scan | Identification and annotation of different RNA viruses within metagenomic datasets | 26.22 | D | https://github.com/shenwei356/taxonkit [147] |
| RotaC | Genotyping tool for group A rotaviruses | 29.83 | W | http://rotac.regatools.be [156] |
| RVDB | Contains all viral, virus-related, and virus-like nucleotide sequences | 24.2 | W,D | https://rvdb.dbi.udel.edu/ [141] |
| RVDB-prot | Resource of viral proteins together with its HMM analysis | 12.1 | W,D | https://rvdb-prot.pasteur.fr/ [18] |
| RVRD[c] | Database to enhance virus detection in metagenomic data | 24.2 | D | https://bench.cs.vt.edu/FastViromeExplorer/RNA_virus_database/ [67] |
| SCUEAL[c] | Prediction of HIV-1 subtypes | 12.92 | W | http://www.datamonkey.org/dataupload_scueal.php [157] |
| SDT | Virus classification based upon pairwise alignment of sequences and identity calculation | 160.3 | W,D | www.cbio.uct.ac.za/SDT [82] |
| SeqMap[c] | For mapping short sequences to genome | 40.89 | W | http://seqmap.compbio.iupui.edu/ [170] |

*(continued)*

Table 1. Continued

| Tools | Key feature(s) | CI[a] | Access[b] | URL [reference] |
|---|---|---|---|---|
| SGN VIGS | Virus-Induced Gene Silencing (VIGS) frameworks for functional genomics | 12.4 | D | https://github.com/solgenomics/VIGS [174] |
| SSE | Analysis of amino acid and nucleotide sequences | 20.8 | D | www.virusevolution.org/Downloads/Software [158] |
| Stanford HIV Drug Resistance DB | Contains genotype-treatment, genotype–phenotype, and genotype-outcome correlation data | 12 | W | http://hivdb.stanford.edu [146] |
| V-pipe | Pipeline for estimating the genetic diversity of viruses by utilizing high-throughput dataset | 21.19 | D | https://github.com/cbg-ethz/V-pipe [49] |
| VADR | System for validation and annotation of viral sequences in GenBank submissions | 13.09 | D | https://github.com/nawrockie/vadr [127] |
| VAPiD | Detection and annotation tool for viruses | 10.62 | D | https://github.com/rcs333/VAPiD [130] |
| vConTACT | Classifying dsDNA viruses that infects bacterial and archaeal populations | 35.2 | D | https://bitbucket.org/MAVERICLab/vcontact/src/master/ [83] |
| VDA-RWLRLS[c] | Anti-SARS-CoV-2 drug prioritizing framework | 20.9 | D | https://github.com/plhhnu/VDA-RWLRLS/ [5] |
| VIBRANT | Method for annotation of viral genomes and metagenomic assemblies | 165.99 | W, D | https://de.cyverse.org/de/ [58] |
| VICTOR | For classification of viruses | 70.11 | D | https://victor.dsmz.de/ [84] |
| VICUNA | Algorithm for *de novo* assembly of extremely heterogeneous population of viruses | 19.62 | W,D | http://www.broadinstitute.org/scientific-community/science/projects/viral-genomics/viral-genomics-analysis-software [125] |
| VIP | Pipeline for identification and discovery of viruses | 14.86 | D | https://github.com/keylabivdc/VIP [47] |
| VIPERdb v3.0 | Structure-derived data analytics resource for virus capsids | 14.75 | W | http://viperdb.scripps.edu [171] |
| ViPR | Integrated library of data for several viral families | 59.46 | W | http://www.viprbrc.org [160] |
| ViralZone | Fact sheets offering sequence information of all known viral families and genera | 39.74 | W | http://www.expasy.org/viralzone/ [161] |
| ViraMiner | Method to identify viruses in various human biospecimens | 24.95 | D | https://github.com/NeuroCSUT/ViraMiner [63] |
| VirClust | Platform for viral annotations, gene identification, and classification by hierarchy | 20.57 | W,D | http://rhea.icbm.uni-oldenburg.de/virclust/ [56] |
| VirFinder | Finding sequences of viruses through assembled metagenomic data | 78.74 | D | https://github.com/jessieren/VirFinder [59] |
| VirHostMatcher-Net | Predicting virus–prokaryote interactions | 25.33 | W,D | https://anaconda.org/hcc/virhostmatcher-net [38] |
| VirHostNet 2.0 | Knowledgebase devoted to exploration of virus–host protein–protein interactions | 22.51 | W | http://virhostnet.prabi.fr [35] |
| VIRIDIC | Computing intergenomic resemblance of prokaryote-infecting viruses | 123.92 | W,D | https://rhea.icbm.uni-oldenburg.de/viridic/ [77] |
| VIRify | Method for identification, annotation, and taxonomy categorization of viruses | 16.5 | D | https://github.com/EBI-Metagenomics/emg-viral-pipeline [129] |
| VIRmiRNA | Database for experimentally verified viral miRNAs and its targets | 11.5 | W,D | http://crdd.osdd.net/servers/virmirna [7] |
| ViroBLAST | Extends utility of BLAST to query against multiple databases and user-customized datasets | 16.06 | W, D | http://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php [162] |
| VIROME | Tool for classifying the metagenomic sequences of viruses | 17.56 | W | http://virome.dbi.udel.edu [163] |
| ViromeScan | Profiling metagenomic virus community | 15.43 | W,D | http://sourceforge.net/projects/viromescan/ [80] |
| VirSorter2 | Method to identify distinct DNA and RNA viruses | 204.52 | D | https://bitbucket.org/MAVERICLab/virsorter2 [91] |
| VIRULIGN | Tool for rapid annotation and alignment of viral genomes | 10.7 | D | https://github.com/rega-cev/virulign [126] |
| Virus taxonomy | Database of the ICTV | 170.65 | W | https://ictv.global/ [137] |
| Virus Variation Resource | Viral sequence data resource | 56.783 | W | http://www.ncbi.nlm.nih.gov/genome/viruses/variation/ [164] |
| Virus-Genotyping Tools[c] | High-throughput genotyping of recombinant and nonrecombinant viruses | 11.68 | W | http://bioafrica.mrc.ac.za/rega-genotype/html/ [165] |
| Virus-Host DB | Linking virus genomes with host taxonomy | 44.34 | W | https://www.genome.jp/virushostdb/ [92] |
| Virus-PLoc | Prediction of subcellular location of viral proteins | 10.23 | W | http://www.csbio.sjtu.edu.cn/bioinf/virus [166] |
| VirusCircBase | Database of virus circular RNAs | 12.8 | W,D | http://www.computationalbiology.cn/ViruscircBase/home.html [145] |
| Virus-CKB 2.0 | Viral-associated disease-specific chemogenomics knowledgebase | 7.2 | W | https://www.cbligand.org/g/virus-ckb [3] |
| VirusDetect[c] | Pipeline for efficient virus discovery | 29.30 | W,D | http://bioinfo.bti.cornell.edu/tool/VirusDetect/ [51] |
| Viruses.STRING | Virus–host protein–protein interaction database | 22.88 | W,D | http://viruses.string-db.org/ [36] |
| VirusFinder[c] | Software for detecting integration sites by blending SVdetect with CREST | 15.45 | D | http://bioinfo.mc.vanderbilt.edu/VirusFinder/ [167] |
| VirusMINT[c] | Catalogues interactions between viral and human proteins | 14.35 | W | http://mint.bio.uniroma2.it/virusmint/ [168] |
| VirusSeeker | Pipeline for virus discovery and virome composition analysis | 18.97 | W,D | http://pathology.wustl.edu/virusseeker/index.htm [48] |
| VirusSeq[c] | Finding integration events using discordant read-pair information | 14.35 | D | http://odin.mdacc.tmc.edu/~xsu1/VirusSeq.html [169] |
| ViruSurf | Investigating viral sequences | 12.24 | W | http://gmql.eu/virusurf_gisaid/ [142] |
| VISDB | Database of viral integration sites in the human genome | 10.26 | W,D | https://bioinfo.uth.edu/VISDB [143] |
| VPF-Class | Taxonomic classification and host prediction of viral contigs | 24.25 | D | https://github.com/biocom-uib/vpf-tools [122] |
| WIsH | Prediction of prokaryotic hosts of phages from their genomic sequences | 34.52 | D | https://github.com/soedinglab/wish [121] |

[a]CI, citation index (number of citations per year and as of January 2025). [b]W, web based. D, downloadable. W, D, both Web based and downloadable. All the tools mentioned are freely available. [c]Currently inaccessible.

Table 2. Categorization of resources based on specific application

| Application | Resource [reference] | Specific feature(s) | CI[a] | Access[b] |
|---|---|---|---|---|
| Drugs and therapeutics | AI4AVP [8] | Tool for the prediction of antiviral peptides | 12 | W,D |
| | AntiVPP 1.0 [9] | Predict antiviral peptides | 9.94 | D |
| | AVCpred [6] | For prediction and design of antiviral compounds | 9 | W,D |
| | AVPdb [17] | Resource of experimentally validated antiviral peptides targeting medically important viruses | 21.8 | W |
| | AVPIden [13] | Identifies and predict functionality of antiviral peptides | 18.16 | W |
| | AVPpred [10] | Antiviral peptide prediction algorithm | 20.16 | W |
| | CORDITE [4] | Provides knowledge on potential drugs for SARS-CoV-2 | 6.33 | W |
| | CoVex [37] | Platform for SARS-CoV-2 host interactome exploration and drug (target) identification | 40.52 | W |
| | Deep-AVPpred [15] | For discovery of peptide drugs for viral infections | 16.66 | W,D |
| | DeepHost [32] | Tool for phage host prediction | 14.2 | D |
| | DeepIPs [22] | Identifies phosphorylation sites in host cells infected with SARS-CoV-2 | 26.59 | W,D |
| | ENNAVIA [11] | Method for antiviral and anti-coronavirus activity prediction for therapeutic peptides | 21.4 | W,D |
| | Geno2pheno [19] | For predicting drug resistance in HIV-1, HBV, and HCV | 15.37 | W |
| | HCVpro [27] | Database of hepatitis C virus protein interaction | 6.99 | W,D |
| | HVIDB [26] | Catalogues human–virus protein–protein interactions | 16.2 | W,D |
| | iDVIP [14] | Model for identification and characterization of viral integrase inhibitory peptides | 0.99 | W |
| | iLoc-Virus[c] [23] | For identifying the subcellular localization of virus proteins | 20.98 | W |
| | LSTM-PHV [24] | Predicts human–virus protein–protein interactions | 26.9 | W,D |
| | Meta-iAVP [12] | For improving the prediction of antiviral peptides | 21.8 | W |
| | MVP [30] | Microbe–phage interaction database | 10.98 | W |
| | NCBI-HHPID [28] | Catalogues all interactions between HIV-1 and human proteins | 5.46 | W |
| | PandoraGAN [16] | Tool to design bioactive antiviral peptides | 20.25 | D |
| | P-HIPSTer [25] | Pathogen–host interaction prediction tool | 27.77 | W,D |
| | PHIAF [29] | For prediction of phage–host interactions | 18 | D |
| | PHISDetector [31] | Tool to detect diverse *in silico* phage–host interaction signals | 9.9 | D |
| | Standford HIV Drug Resistance DB [20] | Includes genotypic-treatment and genotypic-phenotypes | 12 | W |
| | VDA-RWLRLS[c] [5] | Anti-SARS-CoV-2 drug prioritizing framework | 20.9 | D |
| | ViralPhos[c] [21] | Phosphorylation sites prediction on virus proteins | 1.70 | W |
| | ViRBase v3.0 [34] | Virus and host ncRNA-associated interaction repository | 5.88 | W,D |
| | VirHostMatcher-Net [38] | Framework for predicting virus–prokaryote interactions | 25.33 | W,D |
| | VirHostNet 2.0 [35] | Knowledgebase of network-based exploration of virus–host protein–protein interactions | 22.51 | W |
| | VIRmiRNA [7] | Resource for experimentally validated viral miRNAs and their targets | 11.53 | W,D |
| | Virus-CKB 2.0 [3] | Viral-associated disease-specific chemogenomics knowledgebase | 7.02 | W |
| | Viruses.STRING [36] | Virus–host protein–protein interaction database | 22.8 | W,D |
| Vaccine development | CATNAP [46] | Tool for compiling, analysing, and tally neutralizing antibody panels | 16.6 | W,D |
| | DBCOVP [41] | Database of coronavirus virulent glycoproteins | 5.87 | W |
| | IRD [44] | Resource for influenza virus research | 45.85 | W |
| | ISU *FLU*ture [42] | For monitoring the temporal genetic patterns of influenza A virus in swine | 9.8 | W |
| | IRAT [43] | Pandemic preparedness and influenza risk assessment tool | 6.81 | W |
| | MMV-db [45] | Database about vaccinomics and RNA-based treatments for mammarenaviruses that causes contagious hemorrhagic fever | 4.1 | W,D |
| Diagnosis and novel virus discovery | Cenote-Taker 2 [61] | Virus discovery and annotation tool | 29 | D |
| | CliqueSNV [65] | Reconstruction of viral variants | 14.2 | D |
| | coronaSPAdes [66] | Assembler for RNA viral species recovery | 21 | W,D |
| | DisCVR [64] | Tool for rapid viral diagnosis | 1.84 | W |
| | DeepVirFinder [57] | Tool for virus identification using metagenomic data | 94.37 | D |
| | Genome Detective [62] | For identifying viruses using high-throughput sequencing data | 62.6 | W,D |
| | GLUE [76] | Software for viral sequence data | 18 | W,D |
| | HERVd [68] | Human intrinsic retrovirus database | 10.4 | W |
| | MetaviralSPAdes [50] | For assembling viruses using metagenomic data | 51.4 | D |
| | MicroGMT [71] | Mutation tracking system for SARS-CoV-2 and other genomic sequences of microbes | 6.44 | D |
| | PaVE [55] | Database of curated papillomavirus genomic sequences | 36.2 | W |
| | PoSeiDon [72] | For identification of recombinational events and natural selection | 3.62 | D |
| | PuMA [53] | Papillomavirus genome annotation tool | 3.2 | D |
| | RAT [74] | High-throughput detection of recombination | 7.2 | D |

*(continued)*

Table 2. Continued

| Application | Resource [reference] | Specific feature(s) | CI[a] | Access[b] |
|---|---|---|---|---|
| | RDP5 [73] | Program for analysing recombination | 114 | W,D |
| | SimFlu[c] [54] | For forecasting the influenza A virus variation patterns | 9.58 | W |
| | V-pipe [49] | For analysing the genomic diversity of viruses using high-throughput dataset | 21.19 | D |
| | ViMIC [69] | Database of human disease-related virus mutations, integration sites and *cis*-effects | 6.67 | W,D |
| | ViMRT [70] | Recognizes virus mutation | 3.36 | W |
| | VIP [47] | Pipeline for metagenomics of virus identification and discovery | 14.86 | D |
| | ViraMiner [63] | Method to identify viruses in various human biospecimens | 24.95 | D |
| | VirClust [56] | Tool for clustering, protein identification, and annotation of viruses (prokaryotes) | 20.57 | W |
| | VIRIDIC [77] | To compute the intergenomic resemblance of prokaryotes infecting viruses | 123.92 | W,D |
| | ViReMa [75] | Detects recombination junctions in viral genomes | 8 | D |
| | VirFinder [59] | Tool for finding sequences of viruses through assembled metagenomic data | 78.74 | D |
| | virMine [60] | For identification of viral sequences using complicated metagenome datasets | 9.35 | D |
| | VirusDetect[c] [51] | Tool for efficient virus discovery | 29.3 | W,D |
| | VirusSeeker [48] | Discovers virus and analyses virome composition | 18.97 | W,D |
| Control viral diseases | COVFlow [88] | Method for phylodynamic analysis of viruses from specified sequences of SARS-CoV-2 | 0.39 | D |
| | DbatVir [78] | Resource of bat-linked viruses | 17.58 | W,D |
| | MNPDenseNet[d] [85] | Framework for detecting monkeypox viruses from skin images | 7.2 | D |
| | PASC [81] | For classification of viruses | 18.9 | W |
| | SANTA-SIM [89] | Simulation of evolution of viral sequence under selection and recombination | 5.7 | D |
| | SDT [82] | For classifying viruses on the basis of pairwise sequence alignment | 160.3 | W,D |
| | TIPRA [87] | Tool for pandemic preparedness and the influenza risk assessment | 4.66 | W,D |
| | vConTACT [83] | Classification tool for dsDNA viruses that infects bacterial and archaeal communities | 35.2 | D |
| | VirAnnot [90] | Estimates viral diversity and assigns operational taxonomy units | 6.19 | W,D |
| | VirMAP [79] | Maximal viral information recovery tool | 9.16 | D |
| | ViromeScan [80] | For metagenomic viral community profiling | 15.43 | W,D |
| | VirSorter2 [91] | Detects diverse DNA and RNA viruses | 204.52 | D |
| | Virus-Host DB [92] | Integrating viral genomes together with host taxonomies | 44.34 | W |
| | ZikaVR[c] [93] | For genomics, proteomics, phylogenetic, and therapeutic analysis | 7.51 | W |

[a]CI, citation index (number of citations per year and as of January 2025). [b]W, web based. D, downloadable. W, D, both Web based and downloadable. All the tools mentioned are freely available. [c]Currently inaccessible. [d]CI could not be calculated (published in the past 4 months).

determining the course/severity of viral infection and hence human–virus PPIs are potential drug targets. Towards this end, various resources/tools have been developed: (i) HVIDB (PPI database), LSTM-PHV, and P-HIPSTer (both predict human–virus PPIs) [24–26], (ii) hepatitis C virus protein interaction database (HCVpro) [27], (iii) NCBI-HHPID (compiles all interactions between HIV-1 and human proteins) [28], (iv) PHIAF, PHISDetector (both predict phage–host interaction) and Microbe Versus Phage (MVP) (microbe–phage interaction database) [29–31], (v) DeepHost (phage host prediction tool based on Convolution Neural Networks) [32] (vi) iPHoP (a machine learning based tool for predicting phage–host interactions) [33], (vii) ViRBase v3.0 (repository for information on ncRNA SNPs, ncRNA-related drugs and virus–host ncRNA interactions) [34], (viii) VirHostNet 2.0 and Viruses.STRING (both are databases for exploration of virus–host PPIs) [35, 36], (ix) CoVex (explores SARS-CoV-2 host interactomes) [37], and (x) VirHostMatcher-Net (predicts virus–prokaryotes interactions) [38].

## Vaccine development

Vaccination provides adaptive immunity to a disease and is the most efficient method of preventing infectious diseases. However, conventional strategies to develop vaccines have severe limitations [39]. Hence, computational methods can be profoundly useful for developing novel and effective vaccines, and gear up the long/expensive procedure of vaccine development. Modern vaccine development involves the computational techniques that explore antigen/adjuvant selection as well as prediction of immune response, epitope, and toxicology/allergenicity. This requires utilization of diverse computational tools encompassing immune-informatics, comparative proteomics, gene expression analysis, *in silico* hierarchical identification of protective antigens, reverse and structural vaccinology, molecular docking, and modelling of protein structures (Table 2) [40].

The COVID-19 pandemic has led to the development of DBCOVP database that provides extensive information on the functional and immunological (T-cell and B-cell epitopes) properties of virulent glycoproteins from coronavirus genomes [41]. Furthermore, as exemplified by the recent lumpy skin disease epidemic in cattle, viral diseases are a major threat to the animal husbandry and cause considerable damage to humans and animals. Hence, it is imperative to prevent, diagnose and treat veterinary viral infections [for instance, influenza A virus (IAV) causes severe respiratory disease in birds and some mammals]. ISU *FLU*ture, is a clinical diagnostic database and interactive web interface that allows to monitor patterns in swine IAV positive cases. The objective of this resource is to aid veterinarians to manage and control IAV by exploring the genetic diversity of surface glycoproteins between circulating IAV and vaccine strains [42]. Influenza risk assessment tool (IRAT) is a tool that assesses the potential pandemic risk (based on the properties, ecology, and epidemiology of the virus) posed by IAV [43]. This helps in ascertaining a suitable candidate vaccine virus and developing a human vaccine before infection spreads. Influenza Research Database (IRD) is a freely available online

bioinformatic resource to search, analyse, visualize, store as well as share data for influenza research [44]. Likewise, MMV-db is a database dedicated to vaccinomics and therapeutics information on mammarenaviruses [45]. Generation of neutralizing antibodies is essential for an effective vaccine as they can inhibit viral infection of a target cell. Taking this into account, CATNAP tool was developed to analyse neutralizing antibody panels in conjunction with HIV sequence data [46].

## Diagnosis and novel virus discovery

Novel viruses are the causal agents of emerging pandemics/epidemics. Hence, identification and diagnosis of new/rare viruses are paramount to develop strategies against viral diseases and manipulate viruses for our own advantage. The COVID-19 pandemic has not only ignited the scientific interest in novel virus discovery but has also highlighted the importance of understanding the origin of an outbreak, implementing specific control measures (for instance, 'lockdowns' were unknown in the pre-COVID era) and early diagnosis (so as to initiate/modify the course of prevention/treatment). Various computational resources have been developed for discovering and diagnosing new viruses (Table 2).

High-throughput sequencing techniques play an important role in novel viral discovery and identification, monitoring phylogenetic evolution and clinical diagnosis (for instance, designing of fluorescent PCR probes for COVID-19 detection). Towards this end, Virus Identification Pipeline (VIP) that identifies and discovers the virus from metagenomics data was developed [47]. Likewise, a BLAST-based NGS data analysis pipeline, VirusSeeker was designed for eukaryotic viral discovery as well as virome composition analysis [48]. In similar vein, V-pipe orchestrates the assessment of viral genomic diversity from high-throughput data [49]. MetaviralSPAdes and VirusDetect provide opportunities to reveal novel viruses by identifying viral genomes in metagenomic assembly graphs and by analysing the large-scale small RNA datasets, respectively [50, 51]. PhaGCN2 predicts taxonomy at family level as well as identifies members of the new virus families which have not been defined in International Committee on Taxonomy of Viruses (ICTV) [52]. Virus-specific tools for diagnosis of distinct strains include (i) PuMA, an unbiased papillomavirus genome annotation tool that reports novel papillomaviruses [53], (ii) SimFlu, an analytical tool that simulates multi-level codon variation parameters over time in influenza virus genomic regions and hence predict probable future variants of influenza viruses [54], and (iii) Papillomavirus Episteme (PaVE), which provides a typing tool to assess the novelty of a newly sequenced virus [55]. On the other hand, VirClust annotates viral sequences based on homology and core protein detection [56].

Historically, diseases caused by viruses have been known prior to the discovery of their causal agents such as AIDS, cervical cancers, etc. Thus, discovery of viruses will characterize the diseases they cause, much before the pathogenicity of these agents is defined. Hence, it is important to detect and characterize viruses. Recent advancement in metagenomic sequencing helps in identification of unknown viruses or short viral sequences, e.g., DeepVirFinder [57], VIBRANT [58], VirFinder [59], virMine [60], and Cenote-Taker 2 [61], predict/identify viral sequences in metagenomic data and discovers viruses from unknown datasets. Similarly, ViraMiner, DisCVR, and Genome Detective both detect potential unknown/highly divergent viral genomes in human biospecimens and from high-throughput sequencing data, respectively [62–64]. CliqueSNV is an algorithm for early detection and identification of minor viral haplotypes [65] while coronaSPAdes, ends the limitation of other assemblers that are unable to assemble several viruses into a single contig and hence recover RNA virus species [66]. On the other hand, RVRD is a compilation of RNA virus genomes from NCBI, ViralZone, and ViPR to provide enhanced identification of RNA viruses in metagenomic data [67]. Likewise, for human endogenous retroviruses (HERV), the HERVd resource precisely defines HERV families and provides better classification and identification [68].

Genetic changes in viruses, that are responsible for creating novel viruses, are the result of recombination events (occur when coinfecting viruses exchange genetic information) and mutations (errors incorporated in viral genome). The recombination episodes impact the evolution of virus species, while several mutations accumulate in viral strains over time and change its fundamental properties. Tools that help in analysing or identifying mutations/recombinations are: (i) Virus Mutations, Integration sites and *cis*-effects (ViMIC), which provides information on viral mutation/integration sites as well as target genes [69], (ii) ViMRT, a search engine for automated virus mutation recognition [70], (iii) MicroGMT, identifies as well as characterizes indels and point mutations [71], (iv) PoSeiDon, a pipeline to identify recombination events and positively selected sites in protein-coding sequences [72], (v) RDP-5, an algorithm for detecting and visualizing recombination [73], (vi) Recombination Analysis Tool (RAT), a tool for finding viral recombinants [74], and (vii) Virus Recombination Mapper (ViReMa), specifically developed to accurately identify recombinational events in the longer reads [75]. Moreover, identifying the changes in viral genomes can also be used to diagnose diseases. Genome sequencing plays a crucial role in finding these differences and reveals viral phenotypic characteristics, such as antigenic determinants. To facilitate understanding of virus genome sequencing, sequence data should be processed within virus-specific computational resources. GLUE is a bioinformatics platform for development of virus sequence data resources [76]. Comparative genomics is also very important to find evolutionary changes in the viral genome and hence, might potentially discover a novel virus as well. VIRIDIC is one such tool that computes distances/similarities amongst viral genomes [77].

## Control of viral diseases

Getting control over viral diseases is not possible without understanding the viral pathogenesis, evolution, heterogeneity, and functionality. Viral metagenomic studies offer the opportunity to comprehend and expand this knowledge. Approaches like 'virus taxonomy classification' enable the clinical, molecular, and evolutionary traits of the virus to be placed in a framework that includes and connects all viruses. Taxonomic perspective depicts the origination/replication pattern of a specific virus, ascertains the possible host it emerged from, how it causes diseases and how the host responds when afflicted. All these parameters help us to device a possible way of controlling diseases (e.g., possible treatments and produce vaccines) and several tools have been developed towards this end (Table 2). For instance, taxonomy has helped to understand (i) the most likely natural host for viruses (i.e. bats for novel coronavirus) and (ii) evolution in viral genome by genetic recombination. DBatVir provides information about the genetic variability of the bat-associated viruses to prevent future epidemics [78]. VirMAP and ViromeScan (metagenomic viral community profiling tool) both are meant for viral identification and taxonomic classification from metagenomic data [79, 80]. Similarly, PASC (PAirwise Sequence Comparison) and SDT (Sequence Demarcation Tool) both use taxonomic

classification methods based on pairwise genetic identity measures and determine demarcations at distinct taxonomic levels such as strain, species, genera, and subfamily/family [81, 82]. VICTOR and vConTACT also do taxonomic classification, but specifically for prokaryotic viruses [83, 84]. In similar vein, MNPDenseNet is a machine learning architecture developed for detecting monkeypox viruses from skin images [85]. Similarly, for keeping a check on epidemic HIV infections and patients undergoing antiretroviral treatment, a statistical model Naomi was designed to control this disease at the district level in sub-Saharan Africa while TIPRA assesses risk associated with influenza virus [86, 87].

Phylogenetic analysis is a powerful means to evaluate interspecies differences to optimize public health responses to viral outbreaks. COVFlow is a pipeline for performing phylogenetic and phylodynamic analysis of SARS-CoV-2 genomes to check the spread and evolution of the variants [88]. Undoubtedly, simulation tools for replicating natural selection, recombination and complex demography patterns, are important for understanding the complex evolutionary scenarios. One such tool that imitates evolution dynamics of viruses under selection and recombination is SANTA-SIM [89]. Viral metagenomics analyses help to understand the genetic diversity and content of all the prevailing viral communities. VirAnnot (automated viral diversity estimation) is a resource for estimating viral diversity and identification of Operational Taxonomic Units (OTUs) in metavirome data [90]. Most viral signals remain concealed in the metagenomic datasets due to the absence of universal gene markers and inadequate advancement in the identification tools. Considering these challenges, VirSorter2 tool was designed to identify diverse DNA and RNA viruses [91]. The gap in analysis of organism interaction networks was filled by GenomeNet Virus-Host Database that provides the taxonomic links between viruses and their hosts [92]. Vaccine epitopes, distinct RNAs (siRNAs, miRNAs, and sgRNAs) as well as repurposed drug candidates are potential therapeutic components for controlling the viral diseases. ZikaVR is one such integrated repository developed for interventions against ZIKV [93].

## Computational approaches towards COVID-19

The SARS-CoV-2 pandemic had triggered worldwide efforts to understand the virus, to uncover its pathogenesis and develop diagnosis/treatment options. Due to the widespread calamity caused by COVID-19, several resources have been built specifically for the coronavirus (Table 3). Computational strategies were adopted explicitly as a rapid necessity for quick detection, vaccine development and treatment of SARS-CoV-2 infection. There are various bioinformatics tools for diagnosis of SARS-CoV-2 infection, conducting sequence analysis, monitoring the COVID-19 pandemic, evaluating containment strategies, studying coronavirus evolution, finding potential drug targets, and developing therapeutic measures [94]. Pangolin (Phylogenetic Assignment of Named Global Outbreak Lineages), ViTAL (Vision TrAnsformer based Low coverage), and HAVoC (Helsinki university Analyzer for Variants of Concern) assign lineage to a new SARS-CoV-2 genomic sequence while CovidPhy classifies SARS-CoV-2 genomes into a clade and provides geographic information of this clade [95–98]. Likewise, Covidex is a machine learning based subtyping tool whereas RASCL is useful for assessing the SARS-CoV-2 clades through molecular sequence analysis [99, 100]. MyCov is an R package for classification of coronaviruses [101]. Since, mutations can render the variants to evade existing treatment strategies, tracking genetic changes is very essential to provide control on the pandemic. CoV-GLUE, PipeCoV, CoV2Var, and Cov-Seq are tools for analysing genomic variations in the SARS-CoV-2 sequences [102–105]. Likewise, ViruClust also facilitates the identification of mutants and/or variants of potential concerns by comparing viral genomes whereas MicroGMT tracks mutations for SARS-CoV-2 and other microbial genomes [56, 71]. On the contrary, EpiSurf has been developed for assessing amino acid changes in SARS-CoV-2 epitopes and other viral strains, SCoV2-MD organizes atomistic simulations of the SARS-CoV-2 proteome and IDbSV is a database for keeping a check on the evolution and genomic variations of SARS-CoV-2 [106–108]. To understand the molecular mechanisms underlying SARS-CoV-2 infection and the changes in host cellular pathways, DeepIPs enables prediction of phosphorylation sites in the SARS-CoV-2 infected host cells [22]. In similar vein, COVID-AMD is a resource for ascertaining the animal models for coronaviruses [109]. Additionally, Shapify is useful for the prediction of pseudoknots RNA secondary structures in SARS-CoV-2 genomes [110]. Therapeutic databases/tools hold the potential to facilitate drug discovery; for instance, VDA-RWLRLS is a drug-prioritizing framework [5]. Furthermore, to expedite COVID-19 therapeutics development, CoVex and CORDITE resources were developed that integrate the SARS-CoV-2, host and drug interactions [4, 37]. ENNAVIA, iACVP, and CoronaPep are tools to identify/predict ACVPs (Anti-CoronaVirus Peptides) for the prevention and treatment of COVID-19 [11, 111, 112]. Maintaining and managing biological data (particularly molecular data) is a global concern. Towards this end, VirusLab was designed for searching, visualizing, analysing, and managing the SARS-CoV-2 data [113], whereas PAGER-CoV is a database that collates pathways, annotated genes, and gene signatures [114]. Some other useful tools that have been developed are: (i) AutoCoV, to track the early spread of COVID-19 [115], (ii) CoV2ID, for the detection of SARS-CoV-2 infection [116], (iii) CoVDB, for carrying out comparative analysis of coronavirus genes and genomes [117], (iv) COVIDium, a compendium of COVID-19 resources [118], and (v) gcCov, hosts data from publicly available genomes/publications [119].

## Other diverse tools

Other distinct areas in virology that need to be focused like host prediction, *de novo* assembly and annotation of viral sequences, prediction of viral complexity and recognition of core genes inside the viral ancestry. In this regard, several computational pipelines are available to understand the effect of viruses on host behaviour, diversity and function, and to estimate host contamination for viral contigs. Various tools available for virus–host interactions prediction are: (i) Random Forest Assignment of Hosts (RaFAH) and Who Is the Host (WIsH), both for predicting the host (bacteria and archaea) of viruses [120, 121] and (ii) Viral Protein Families-Class (VPF-Class) for predicting host of viral contigs [122]. PVPred-SCM has been developed for predicting the phage virion proteins [123]. Tools like IVA and VICUNA, perform *de novo* population consensus assembly of RNA virus genomes and highly diverse viral populations [124, 125]. On the other hand, VADR, VAPiD, and VIRULIGN annotate viral sequences, while Phigaro specifically annotates prophage sequences [126–130]. Recently, VIRify pipeline was developed that uses virus-specific protein profile HMMs for annotation and taxonomic classification [129].

There is dire need of advance research on viruses that are major global public health threat like influenza virus, dengue virus,

Table 3. SARS-CoV-2 centred viroinformatics tools

| Tools [reference] | Key feature(s) | CI[a] | Access[b] |
|---|---|---|---|
| AutoCoV [115] | For tracking early spread of COVID-19 | 1.87 | D |
| CORDITE [4] | Knowledgebase on possible drugs used for SARS-CoV-2 | 6.33 | W |
| CoronaPep[c] [112] | Tool for generation of anti-coronavirus peptide | 2.4 | W |
| CoronaSPAdes [66] | Assembler for RNA viral species recovery | 21 | W,D |
| CoVDB[c] [117] | Library of coronavirus genes and genomics | 2.7 | W |
| CoV-GLUE [102] | Tool for monitoring the genomic variations in SARS-CoV-2 | 40.66 | W |
| CoV-Seq [104] | For analysing and visualizing genomic sequences of SARS-CoV-2 | 7.2 | W,D |
| CoV2ID [116] | Oligo libraries for SARS-CoV-2 | 1.32 | W |
| COV2Var [105] | Function annotation database of SARS-CoV-2 genetic variation | 7.7 | W |
| CoVex [37] | Tool for exploration of SARS-CoV-2 host interactions and identification of the drugs | 40.52 | W |
| COVFlow [88] | For phylodynamic analysis of viruses from specified sequences of SARS-CoV-2 | 0.39 | D |
| Covidex[c] [99] | SARS-CoV-2 subtyping tool | 4.1 | W,D |
| COVID-AMD [109] | Platform for identification of animal models for coronaviruses | 0 | W |
| COVIDium [118] | COVID-19 resource compendium | 3.38 | D |
| CovidPhy [97] | SARS-CoV-2 variant phylogeographic analysis web tool | 3.99 | W |
| DBCOVP [41] | Database of coronavirus virulent glycoproteins | 5.87 | W |
| DeepIPs [22] | For finding phosphorylation sites in SARS-CoV-2-infected host cells | 26.59 | W,D |
| ENNAVIA [11] | Prediction of anti-coronavirus effects for therapeutic peptides | 21.4 | W, D |
| EpiSurf [106] | Tool for assessing amino acid changes in SARS-CoV-2 epitopes as well as other strains of viruses | 3.38 | W,D |
| gcCov [119] | Database for global coronavirus studies | 1.09 | D |
| HaVoC [96] | Platform for ancestry assignment for genomic sequences of SARS-CoV-2 | 12.8 | W,D |
| iACVP [111] | Identifies anti-coronavirus peptides | 15.3 | W,D |
| IDbSV [107] | For monitoring the evolution as well as variations of SARS-CoV-2 | 2.33 | D |
| MicroGMT [71] | Mutation tracking system for SARS-CoV-2 as well as for other microbial genotypes | 6.44 | D |
| MyCoV [101] | R package for taxonomic classification of coronaviruses | 3.76 | D |
| PAGER-CoV [114] | Database regarding pathways, annotated gene-lists, and gene signatures | 3.3 | W,D |
| Pangolin [95] | For assigning lineage to a particular SARS-CoV-2 genomic sequence | 298.8 | D |
| PipeCoV [103] | Tool to assemble, annotate, and identify variants in SARS-CoV-2 genomic sequences | 4.36 | D |
| RASCL [100] | Tool for rapid assessment of SARS-CoV-2 clades | 0.68 | W,D |
| SCoV2-MD[c] [108] | Database for atomistic simulations of the SARS-CoV-2 proteome | 8.75 | W |
| Shapify[c] [110] | Prediction of pseudoknots RNA secondary structures | 4.9 | D |
| VDA-RWLRLS[c] [5] | Platform for prioritizing SARS-CoV-2 drugs | 20.9 | D |
| VirusLab [113] | Program for analysing the SARS-CoV-2 data | 2.46 | W,D |
| ViTAL[d] [98] | Tool for lineage assignment to a new SARS-CoV-2 sequence | – | W |

[a]CI, citation index (number of citations per year and as of January 2025). [b]W, web based. D, downloadable. W, D, both Web based and downloadable. All the tools mentioned are freely available. [c]Currently inaccessible. [d]CI could not be calculated (published in the past 4 months).

HIV (human immunodeficiency virus), hepatitis virus, and Zika virus. Hence, several virus-specific stools have been developed to target these pathogens. IRD and GISAID are comprehensive compilations of influenza associated data and provide data sharing platforms, thus expediting the study and generation of drugs, diagnostics, and treatments against this virus [44, 131]. INSaFLU (INSide the FLU) is a framework to automate manipulation of NGS data for influenza whole-genome-based surveillance [132]. There is a need for appropriate selection of diagnostic primers for identification of flaviviruses because they cause various maladies such as extreme autoimmune diseases and neurological disorders. On that account, ZikaVR (Zika Virus Resource) was developed that embraces ZIKV genetic, proteomic as well as therapeutic knowledge for facilitating research on Zika virus [93]. Enveloped dsDNA viruses like hepadnaviruses (e.g., hepatitis virus) are of greatest concern because of their potentiality of their outbreaks and epidemic spread. Among the five kinds of hepatitis viruses (*A*, *B*, *C*, *D*, and *E*), forms *B* and *C* cause chronic disorders and together, are the most common cause of developing liver cancer and cirrhosis. Hepatitis C virus protein interaction database (HCVpro) has compiled information on protein–protein interactions and functional genomics data that supplements the search for diagnostic markers and therapeutics [27]. HBVdb is a specialized database to examine the genetic diversity and therapeutic resistance of HBV [133]. Likewise, HEVnet assists in sharing and analysing HEV sequence data accompanied by metadata [134]. Optimum natural immune response in HBV is mediated by epitope presentation and may help in the development of novel immunotherapeutic interventions. Hence, Hepitopes database that has the collection of these epitopes was developed [135]. To effectively track the spread of drug resistance, there is a need of cataloguing sensitive and specific mutations that can be adopted by all the surveillance studies. Towards this end, HIV Drug Resistance Database compiles common mutations of HIV [20]. The group of viruses called endogenous retrovirus (ERV) are integrated retroviral sequences that comprise 5%–8% of the human genome. To understand the biological relevance/impact of ERVs on human health, computational analysis is required. ERVmap reveals the transcription of human ERVs [136].

There are several other distinct/diverse virus databases that also facilitate viral research: (i) Virus taxonomy, database of universal virus taxonomy [137], (ii) CheckV, checks for quality and completeness of viral sequences in metagenomes [138], (iii) IMG/VR v3, largest collection of viral genomes extracted from metagenomes [139], (iv) OncoDB, for examining gene expression and viral infection in cancer [140], (v) Reference Viral Database (RVDB), compiles all viral, virus-like, and viral-associated nucleotide sequences [141], (vi) ViruSurf, database of viral sequences and associated metadata [142], (vii) VISDB, hosts viral integration site-related information in human genome

**A.**

Show 10 entries Search:

| Tools | Key features | CI | Access | Virus/Disease | Application |
|---|---|---|---|---|---|
| GISAID | Data-sharing platform for Influenza, Dengue, Mpox and hCoV-19 viruses | 381.99 | Web | Influenza, Dengue, Mpox and hCoV-19 viruses | Database |
| CheckV | Pipeline for assessing the quality and completeness of metagenome-assembled viral genomes | 349.6 | Install | All viruses | Metagenomics |
| Pangolin | For allocating the most possible lineage to a particular SARS-CoV-2 genomic sequence | 298.8 | Install | SARS-CoV-2, COVID, Coronavirus | Genome annotation |
| VirSorter2 | Detection method for distinct DNA as well as RNA viruses | 204.2 | Install | All viruses | Control of viral diseases, Diagnosis and novel virus discovery |
| Virus taxonomy | ICTV database | 170.65 | Web | All viruses | Genotyping |
| VIBRANT | Method for annotation of viral genomes and metagenomic assemblies | 165.99 | Web, Install | Human Immunodeficiency Virus (HIV), AIDS | Genome annotation |
| SDT | For classifying viruses on the basis of pairwise sequence alignment | 160.3 | Web, Install | All viruses | Control of viral diseases, Diagnosis and novel virus discovery |
| GVD | Identifying age-based patterns of virome diversity in the gut of humans | 131.5 | Web, Install | All viruses | Genome annotation |
| VIRIDIC | To compute the intergenomic resemblance of prokaryotes infecting viruses | 123.92 | Web, Install | All viruses | Control of viral diseases, Diagnosis and novel virus discovery |
| RDP5 | For studying the recombination in, and eliminating indications of recombination from nucleotide sequence datasets | 114 | Web, Install | All viruses | Diagnosis and novel virus discovery |

Showing 1 to 10 of 358 entries Previous 1 2 3 4 5 ... 36 Next

*Currently inaccessible*

**B.**

Show 10 entries Search: metagenomic

| Tools | Key features | CI | Access | Virus/Disease | Application |
|---|---|---|---|---|---|
| CheckV | Pipeline for assessing the quality and completeness of metagenome-assembled viral genomes | 349.6 | Install | All viruses | Metagenomics |
| VIBRANT | Method for annotation of viral genomes and metagenomic assemblies | 165.99 | Web, Install | Human Immunodeficiency Virus (HIV), AIDS | Genome annotation |
| DeepVirfinder | Viral identification using metagenomic data | 94.37 | Install | All viruses | Diagnosis and novel virus discovery |
| VirFinder | Tool for finding sequences of viruses through assembled metagenomic data | 78.74 | Install | All viruses | Diagnosis and novel virus discovery |
| MetaviralSPAdes | Assembly of viruses from metagenomic data | 51.4 | Install | All viruses | Diagnosis and novel virus discovery |
| RdRp-scan | Tool for the identification and annotation of different RNA viruses within metagenomic datasets | 26.22 | Install | All RNA viruses | Metagenomics |
| VIROME | Platform for classifying metagenomic sequences of viruses | 17.56 | Web | All viruses | Metagenomics |
| ViromeScan | Tool for metagenomic viral community profiling | 15.43 | Web, Install | All viruses | Control of viral diseases |
| VIP | Method for finding and identifying viruses using metagenomics | 14.86 | Install | All viruses | Control of viral diseases, Diagnosis and novel virus discovery, Metagenomics |
| RNN-VirSeeker | Tool for identifying brief viral sequences across metagenomic data | 12.77 | Install | All viruses | Metagenomics |

Showing 1 to 10 of 17 entries (filtered from 358 total entries) Previous 1 2 Next

*Currently inaccessible*

Figure 2. (A) Screenshot of VITALdb tools sorted by CI (reflect the popularity/usage of the tool). (B) Screenshot of VITALdb on search using desired keyword(s) (here, metagenomic).

[143], (viii) pVOGs, represents cluster of orthologous gene families in genomes of viruses that infect bacteria and archaea [144], (ix) VirusCircBase, contains data related to viral circular RNAs [145], (x) Stanford HIV Drug Resistance DB, includes genotypic-treatment and genotypic–phenotype correlations for HIV [146], and (xi) RdRp-Scan, a database to detect RNA-dependent RNA polymerases for virus genome annotation [147]. Various additional important tools such as FluTE, GATU, LANL HCV database, Metavir, NCBI Genotyping Tool, NCBI Viral Genomes Resource, NCBI-IVR, PHACCS, RotaC, SCUEAL, SeqMap,

SSE, VIPERdb, ViPR, ViralZone, ViroBLAST, VIROME, Virus Variation Resource, Virus-Genotyping Tools, Virus-Ploc, VirusFinder, VirusMINT, VirusSeq had already been covered/described in our earlier publication [148–170]. Amongst these, new versions of two tools Metavir and VIPERdb have been developed. The first version of Metavir provided a framework that allowed comprehensive virome analysis while the new release of Metavir (Metavir 2) was developed for comparative study of viromes to overcome challenges like rapid increase in the number of viromes and assembling the vast amount of metagenomic sequence data into large genomic fragments [151]. Initially, VIPERdb provided comprehensive structure-derived information on viruses but the improved version of VIPERdb (VIPERdb v3.0) [171], besides containing all the characterized viruses (~181 000) known till December 2020, also host tools to identify hot-spot residues in the desired set of structures and for anomaly detection in virus capsids. On the other hand, Gut Virome Database (GVD) compiles 33 242 distinct viral populations present in gut to understand the crucial role of viruses in human health and disease [172].

## Database architecture and web interface

For compiling the viroinformatics tools included in VITALdb, extensive search was carried out in literature using all possible diverse keywords. Subsequently, to facilitate effective access to this vital database (VITALdb), a web interface was developed using HTML, CSS, and JavaScript for its layout, style, design, and content structuring. The backbone of the database is an interactive table that collates all the viroinformatics resources along with their key features, citation indexes (CI) (reflecting the popularity/usage), accessibility, specificity, and application. To enhance the utility of VITALdb, we have implemented column sorting (Fig. 2A), instant search along with filtering function (Fig. 2B), and pagination using DataTables (a table enhancing JavaScript library). A user can search VITALdb using keywords mentioning specific virus, disease, feature, tool-name, or application.

## Conclusion

VITALdb is a comprehensive, insightful, and interactive resource of ~360 viroinformatics tools. It will enable users to get an overview of tools available for a desired virus/application and consequently select the best tool(s) to carry out their task. VITALdb is surely going to play a vital role in combating viruses and prevent future outbreaks/pandemics!

**Key Points**

- The recent pandemics/outbreaks of viral diseases have resulted in a boom of viroinformatics resources (~360 tools).
- VITALdb is a compendium of viroinformatics tools/algorithms for providing tremendous fillip to translate fundamental knowledge into applied research.
- Will facilitate early diagnosis of viral infections as well as development of new drugs/targets and vaccines.
- This database will enable scientists to get an overview of tools available for a desired virus/application and select the best tool(s) to carry out their task.

## Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: None declared.

## Funding

This work was supported by MHRD (BT/2014–15/Plan/P-955; BIO/-FIG/100700) and DHR (R.11013/51/2021-GIA/HR; R.11017/36/2023-GIA/HR) grants. S.K., K.S., and M.K. are thankful to MHRD for research fellowships.

## Data availability

The VITALdb resource is available for open access at https://compbio.iitr.ac.in/vitaldb.

## References

1. Sharma D, Priyadarshini P, Vrati S. Unraveling the web of viroinformatics: computational tools and databases in virus research. *J Virol* 2015;**89**:1489–501. https://doi.org/10.1128/JVI.02027-14
2. Pappas N, Roux S, Hölzer M. *et al.* Virus bioinformatics. In: Bamford DH, Zuckerman M (eds.), *Encyclopedia of Virology* 2021; 124–32. https://doi.org/10.1016/B978-0-12-814515-9.00034-5
3. Hao Y, Chen M, Othman Y. *et al.* Virus-CKB 2.0: viral-associated disease-specific chemogenomics knowledgebase. *ACS Omega* 2022;**7**:37476–84. https://doi.org/10.1021/acsomega.2c04258
4. Martin R, Löchel HF, Welzel M. *et al.* CORDITE: the curated CORona drug InTERactions database for SARS-CoV-2. *iScience* 2020;**23**:101297. https://doi.org/10.1016/j.isci.2020.101297
5. Shen L, Liu F, Huang L. *et al.* VDA-RWLRLS: an anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares. *Comput Biol Med* 2022;**140**:105119. https://doi.org/10.1016/j.compbiomed.2021.105119
6. Qureshi A, Kaur G, Kumar M. AVCpred: an integrated web server for prediction and design of antiviral compounds. *Chem Biol Drug Des* 2017;**89**:74–83. https://doi.org/10.1111/cbdd.12834
7. Qureshi A, Thakur N, Monga I. *et al.* VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. *Database* 2014;**2014**:bau103. https://doi.org/10.1093/database/bau103
8. Lin T-T, Sun Y-Y, Wang C-T. *et al.* AI4AVP: an antiviral peptides predictor in deep learning approach with generative adversarial network data augmentation. *Bioinform Adv* 2022;**2**:vbac080. https://doi.org/10.1093/bioadv/vbac080
9. Beltrán Lissabet JF, Belén LH, Farias JG. AntiVPP 1.0: a portable tool for prediction of antiviral peptides. *Comput Biol Med* 2019;**107**:127–30. https://doi.org/10.1016/j.compbiomed.2019.02.011
10. Thakur N, Qureshi A, Kumar M. AVPpred: collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res* 2012;**40**:W199–204. https://doi.org/10.1093/nar/gks450
11. Timmons PB, Hewage CM. ENNAVIA is a novel method which employs neural networks for antiviral and anti-coronavirus activity prediction for therapeutic peptides. *Brief Bioinform* 2021;**22**:bbab258. https://doi.org/10.1093/bib/bbab258

12. Schaduangrat N, Nantasenamat C, Prachayasittikul V. *et al.* Meta-iAVP: a sequence-based meta-predictor for improving the prediction of antiviral peptides using effective feature representation. *Int J Mol Sci* 2019;**20**:5743. https://doi.org/10.3390/ijms20225743
13. Pang Y, Yao L, Jhong J-H. *et al.* AVPIden: a new scheme for identification and functional prediction of antiviral peptides based on machine learning approaches. *Brief Bioinform* 2021;**22**:bbab263. https://doi.org/10.1093/bib/bbab263
14. Huang KY, Kao HJ, Weng TH. *et al.* iDVIP: identification and characterization of viral integrase inhibitory peptides. *Brief Bioinform* 2022;**23**:bbac406. https://doi.org/10.1093/bib/bbac406
15. Sharma R, Shrivastava S, Singh SK. *et al.* Deep-AVPpred: artificial intelligence driven discovery of peptide drugs for viral infections. *IEEE J Biomed Health Inform* 2022;**26**:5067–74. https://doi.org/10.1109/JBHI.2021.3130825
16. Surana S, Arora P, Singh D. *et al.* PandoraGAN: generating antiviral peptides using generative adversarial network. *SN Comput Sci* 2023;**4**:607. https://doi.org/10.1007/s42979-023-02203-3
17. Qureshi A, Thakur N, Tandon H. *et al.* AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. *Nucleic Acids Res* 2014;**42**:D1147–53. https://doi.org/10.1093/nar/gkt1191
18. Bigot T, Temmam S, Pérot P. *et al.* RVDB-prot, a reference viral protein database and its HMM profiles. *F1000Res* 2019;**8**:530. https://doi.org/10.12688/f1000research.18776.1
19. Beerenwinkel N, Däumer M, Oette M. *et al.* Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res* 2003;**31**:3850–5. https://doi.org/10.1093/nar/gkg575
20. Shafer RW. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 2006;**194**:S51–8. https://doi.org/10.1086/505356
21. Huang K-Y, Lu C-T, Bretaña NA. *et al.* ViralPhos: incorporating a recursively statistical method to predict phosphorylation sites on virus proteins. *BMC Bioinformatics* 2013;**14**:S10. https://doi.org/10.1186/1471-2105-14-S16-S10
22. Lv H, Dao F-Y, Zulfiqar H. *et al.* DeepIPs: comprehensive assessment and computational identification of phosphorylation sites of SARS-CoV-2 infection using a deep learning-based approach. *Brief Bioinform* 2021;**22**:bbab244. https://doi.org/10.1093/bib/bbab244
23. Xiao X, Wu Z-C, Chou K-C. iLoc-virus: a multi-label learning classifier for identifying the subcellular localization of virus proteins with both single and multiple sites. *J Theor Biol* 2011;**284**:42–51. https://doi.org/10.1016/j.jtbi.2011.06.005
24. Tsukiyama S, Hasan MM, Fujii S. *et al.* LSTM-PHV: prediction of human-virus protein–protein interactions by LSTM with word2vec. *Brief Bioinform* 2021;**22**:bbab228. https://doi.org/10.1093/bib/bbab228
25. Lasso G, Mayer SV, Winkelmann ER. *et al.* A structure-informed atlas of human-virus interactions. *Cell* 2019;**178**:1526–1541.e16. https://doi.org/10.1016/j.cell.2019.08.005
26. Yang X, Lian X, Fu C. *et al.* HVIDB: a comprehensive database for human–virus protein–protein interactions. *Brief Bioinform* 2021;**22**:832–44. https://doi.org/10.1093/bib/bbaa425
27. Kwofie SK, Schaefer U, Sundararajan VS. *et al.* HCVpro: hepatitis C virus protein interaction database. *Infect Genet Evol* 2011;**11**:1971–7. https://doi.org/10.1016/j.meegid.2011.09.001
28. MacPherson JI, Dickerson JE, Pinney JW. *et al.* Patterns of HIV-1 protein interaction identify perturbed host-cellular subsystems. *PLoS Comput Biol* 2010;**6**:e1000863. https://doi.org/10.1371/journal.pcbi.1000863
29. Li M, Zhang W. PHIAF: prediction of phage-host interactions with GAN-based data augmentation and sequence-based feature fusion. *Brief Bioinform* 2022;**23**:bbab348. https://doi.org/10.1093/bib/bbab348
30. Gao NL, Zhang C, Zhang Z. *et al.* MVP: a microbe-phage interaction database. *Nucleic Acids Res* 2018;**46**:D700–7. https://doi.org/10.1093/nar/gkx1124
31. Zhou F, Gan R, Zhang F. *et al.* PHISDetector: a tool to detect Diverse In silico phage–host interaction signals for virome studies. *Genom Proteom Bioinform* 2022;**20**:508–23. https://doi.org/10.1016/j.gpb.2022.02.003
32. Ruohan W, Xianglilan Z, Jianping W. *et al.* DeepHost: phage host prediction with convolutional neural network. *Brief Bioinform* 2021;**23**:bbab385. https://doi.org/10.1093/bib/bbab385
33. Roux S, Camargo AP, Coutinho FH. *et al.* iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol* 2023;**21**:e3002083. https://doi.org/10.1371/journal.pbio.3002083
34. Cheng J, Lin Y, Xu L. *et al.* ViRBase v3.0: a virus and host ncRNA-associated interaction repository with increased coverage and annotation. *Nucleic Acids Res* 2022;**50**:D928–33. https://doi.org/10.1093/nar/gkab1029
35. 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. https://doi.org/10.1093/nar/gku1121
36. Cook HV, Doncheva NT, Szklarczyk D. *et al.* Viruses.STRING: a virus-host protein-protein interaction database. *Viruses* 2018;**10**:519.
37. 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. https://doi.org/10.1038/s41467-020-17189-2
38. Wang W, Ren J, Tang K. *et al.* A network-based integrated framework for predicting virus–prokaryote interactions. *NAR Genom Bioinform* 2020;**2**:lqaa044. https://doi.org/10.1093/nargab/lqaa044
39. Thomas S, Dilbarova R, Rappuoli R. Future challenges for vaccinologists. *Methods Mol Biol* 2016;**1403**:41–55.
40. Sunita, Sajid A, Singh Y. *et al.* Computational tools for modern vaccine development. *Hum Vaccin Immunother* 2020;**16**:723–35. https://doi.org/10.1080/21645515.2019.1670035
41. Sahoo S, Mahapatra SR, Parida BK. *et al.* DBCOVP: a database of coronavirus virulent glycoproteins. *Comput Biol Med* 2021;**129**:104131. https://doi.org/10.1016/j.compbiomed.2020.104131
42. Zeller MA, Anderson TK, Walia RW. *et al.* ISU FLUture: a veterinary diagnostic laboratory web-based platform to monitor the temporal genetic patterns of influenza a virus in swine. *BMC Bioinformatics* 2018;**19**:397. https://doi.org/10.1186/s12859-018-2408-7
43. Burke SA, Trock SC. Use of influenza risk assessment tool for prepandemic preparedness. *Emerg Infect Dis* 2018;**24**:471–7. https://doi.org/10.3201/eid2403.171852
44. Zhang Y, Aevermann BD, Anderson TK. *et al.* Influenza research database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res* 2017;**45**:D466–74. https://doi.org/10.1093/nar/gkw857
45. Khan T, Khan A, Wei D-Q. MMV-db: vaccinomics and RNA-based therapeutics database for infectious hemorrhagic

fever-causing mammarenaviruses. *Database* 2021;**2021**: baab063. https://doi.org/10.1093/database/baab063
46. Yoon H, Macke J, West AP Jr. *et al.* CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res* 2015;**43**:W213–9. https://doi.org/10.1093/nar/gkv404
47. Li Y, Wang H, Nie K. *et al.* VIP: an integrated pipeline for metagenomics of virus identification and discovery. *Sci Rep* 2016;**6**:23774. https://doi.org/10.1038/srep23774
48. Zhao G, Wu G, Lim ES. *et al.* VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* 2017;**503**:21–30. https://doi.org/10.1016/j.virol.2017.01.005
49. Posada-Céspedes S, Seifert D, Topolsky I. *et al.* V-pipe: a computational pipeline for assessing viral genetic diversity from high-throughput data. *Bioinformatics* 2021;**37**:1673–80. https://doi.org/10.1093/bioinformatics/btab015
50. Antipov D, Raiko M, Lapidus A. *et al.* Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* 2020;**36**: 4126–9. https://doi.org/10.1093/bioinformatics/btaa490
51. Zheng Y, Gao S, Padmanabhan C. *et al.* VirusDetect: an automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology* 2017;**500**:130–8. https://doi.org/10.1016/j.virol.2016.10.017
52. Jiang JZ, Yuan WG, Shang J. *et al.* Virus classification for viral genomic fragments using PhaGCN2. *Brief Bioinform* 2023;**24**:bbac505. https://doi.org/10.1093/bib/bbac505
53. Pace J, Youens-Clark K, Freeman C. *et al.* PuMA: a papillomavirus genome annotation tool. *Virus Evol* 2020;**6**:veaa068. https://doi.org/10.1093/ve/veaa068
54. Ahn I, Kim H-Y, Jung S. *et al.* SimFlu: a simulation tool for predicting the variation pattern of influenza a virus. *Comput Biol Med* 2014;**52**:35–40. https://doi.org/10.1016/j.compbiomed.2014.06.003
55. Van Doorslaer K, Li Z, Xirasagar S. *et al.* The papillomavirus episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res* 2017;**45**:D499–506. https://doi.org/10.1093/nar/gkw879
56. Moraru C. VirClust-A tool for hierarchical clustering, core protein detection and annotation of (prokaryotic) viruses. *Viruses* 2023;**15**:e768869. https://doi.org/10.3390/v15041007
57. Ren J, Song K, Deng C. *et al.* Identifying viruses from metagenomic data using deep learning. *Quant Biol* 2020;**8**:64–77. https://doi.org/10.1007/s40484-019-0187-4
58. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 2020;**8**:1–23. https://doi.org/10.1186/s40168-020-00867-0
59. Ren J, Ahlgren NA, Lu YY. *et al.* VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 2017;**5**:69. https://doi.org/10.1186/s40168-017-0283-5
60. Garretto A, Hatzopoulos T, Putonti C. virMine: automated detection of viral sequences from complex metagenomic samples. *PeerJ* 2019;**7**:e6695. https://doi.org/10.7717/peerj.6695
61. Tisza MJ, Belford AK, Domínguez-Huerta G. *et al.* Cenote-taker 2 democratizes virus discovery and sequence annotation. *Virus Evol* 2021;**7**:veaa100. https://doi.org/10.1093/ve/veaa100
62. Vilsker M, Moosa Y, Nooij S. *et al.* Genome detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics* 2019;**35**:871–3. https://doi.org/10.1093/bioinformatics/bty695
63. Tampuu A, Bzhalava Z, Dillner J. *et al.* ViraMiner: deep learning on raw DNA sequences for identifying viral genomes in human samples. *PLoS One* 2019;**14**:e0222271. https://doi.org/10.1371/journal.pone.0222271
64. Maabar M, Davison AJ, Vučak M. *et al.* DisCVR: rapid viral diagnosis from high-throughput sequencing data. *Virus Evol* 2019;**5**:vez033. https://doi.org/10.1093/ve/vez033
65. Knyazev S, Tsyvina V, Shankar A. *et al.* Accurate assembly of minority viral haplotypes from next-generation sequencing through efficient noise reduction. *Nucleic Acids Res* 2021;**49**:e102–2. https://doi.org/10.1093/nar/gkab576
66. Meleshko D, Hajirasouliha I, Korobeynikov A. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. *Bioinformatics* 2021;**38**:1–8. https://doi.org/10.1093/bioinformatics/btab597
67. Lei S. A RNA virus reference database (RVRD) to enhance virus detection in metagenomic data. 2018;MSc thesis.
68. Paces J, Pavlícek A, Paces V. HERVd: database of human endogenous retroviruses. *Nucleic Acids Res* 2002;**30**:205–6. https://doi.org/10.1093/nar/30.1.205
69. Wang Y, Tong Y, Zhang Z. *et al.* ViMIC: a database of human disease-related virus mutations, integration sites and cis-effects. *Nucleic Acids Res* 2022;**50**:D918–27. https://doi.org/10.1093/nar/gkab779
70. Tong Y, Tan F, Huang H. *et al.* ViMRT: a text-mining tool and search engine for automated virus mutation recognition. *Bioinformatics* 2023;**39**:btac721. https://doi.org/10.1093/bioinformatics/btac721
71. Xing Y, Li X, Gao X. *et al.* MicroGMT: a mutation tracker for SARS-CoV-2 and other microbial genome sequences. *Front Microbiol* 2020;**11**:1502–2. https://doi.org/10.3389/fmicb.2020.01502
72. Hölzer M, Marz M. PoSeiDon: a Nextflow pipeline for the detection of evolutionary recombination events and positive selection. *Bioinformatics* 2021;**37**:1018–20. https://doi.org/10.1093/bioinformatics/btaa695
73. Martin DP, Varsani A, Roumagnac P. *et al.* RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol* 2020;**7**:veaa087. https://doi.org/10.1093/ve/veaa087
74. Etherington GJ, Dicks J, Roberts IN. Recombination analysis tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics* 2005;**21**:278–81. https://doi.org/10.1093/bioinformatics/bth500
75. Sotcheff S, Zhou Y, Yeung J. *et al.* ViReMa: a virus recombination mapper of next-generation sequencing data characterizes diverse recombinant viral nucleic acids. *GigaScience* 2023;**12**:giad009. https://doi.org/10.1093/gigascience/giad009
76. Singer JB, Thomson EC, McLauchlan J. *et al.* GLUE: a flexible software system for virus sequence data. *BMC Bioinformatics* 2018;**19**:532. https://doi.org/10.1186/s12859-018-2459-9
77. Moraru C, Varsani A, Kropinski AM. VIRIDIC—a novel tool to calculate the Intergenomic similarities of prokaryote-infecting viruses. *Viruses* 2020;**12**:1268.
78. Chen L, Liu B, Yang J. *et al.* DBatVir: the database of bat-associated viruses. *Database* 2014;**2014**:bau021.
79. Ajami NJ, Wong MC, Ross MC. *et al.* Maximal viral information recovery from sequence data using VirMAP. *Nat Commun* 2018;**9**:3205. https://doi.org/10.1038/s41467-018-05658-8
80. Rampelli S, Soverini M, Turroni S. *et al.* ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genomics* 2016;**17**:165. https://doi.org/10.1186/s12864-016-2446-3

81. Bao Y, Chetvernin V, Tatusova T. Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. *Arch Virol* 2014;**159**:3293–304. https://doi.org/10.1007/s00705-014-2197-x
82. Muhire BM, Varsani A, Martin DP. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 2014;**9**:e108277. https://doi.org/10.1371/journal.pone.0108277
83. Bolduc B, Jang HB, Doulcier G. *et al.* vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect archaea and bacteria. *PeerJ* 2017;**5**:e3243. https://doi.org/10.7717/peerj.3243
84. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* 2017;**33**:3396–404. https://doi.org/10.1093/bioinformatics/btx440
85. Demir FB, Baygin M, Tuncer I. *et al.* MNPDenseNet: automated monkeypox detection using multiple nested patch division and pretrained DenseNet201. *Multimed Tools Appl* 2024;**83**:75061–83. https://doi.org/10.1007/s11042-024-18416-4
86. Eaton JW, Dwyer-Lindgren L, Gutreuter S. *et al.* Naomi: a new modelling tool for estimating HIV epidemic indicators at the district level in sub-Saharan Africa. *J Int AIDS Soc* 2021;**24**:e25788. https://doi.org/10.1002/jia2.25788
87. World Health Organization. Tool for Influenza Pandemic Risk Assessment (TIPRA), Geneva, 2016.
88. Danesh G, Boennec C, Verdurme L. *et al.* COVFlow: phylodynamics analyses of viruses from selected SARS-CoV-2 genome sequences. *Peer Commun J* 2023;**3**:e101. https://doi.org/10.24072/pcjournal.333
89. Jariani A, Warth C, Deforche K. *et al.* SANTA-SIM: simulating viral sequence evolution dynamics under selection and recombination. *Virus Evol* 2019;**5**:vez003. https://doi.org/10.1093/ve/vez003
90. Lefebvre M, Theil S, Ma Y. *et al.* The VirAnnot pipeline: a resource for automated viral diversity estimation and operational taxonomy units (OTU) assignation for virome sequencing data. *Phytobiomes J* 2019;**3**:256–9. https://doi.org/10.1094/PBIOMES-07-19-0037-A
91. Guo J, Bolduc B, Zayed AA. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 2021;**9**:37. https://doi.org/10.1186/s40168-020-00990-y
92. Mihara T, Nishimura Y, Shimizu Y. *et al.* Linking virus genomes with host taxonomy. *Viruses* 2016;**8**:66. https://doi.org/10.3390/v8030066
93. Gupta AK, Kaur K, Rajput A. *et al.* ZikaVR: an integrated Zika virus resource for genomics, proteomics, phylogenetic and therapeutic analysis. *Sci Rep* 2016;**6**:32713. https://doi.org/10.1038/srep32713
94. Hufsky F, Lamkiewicz K, Almeida A. *et al.* Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Brief Bioinform* 2021;**22**:642–63. https://doi.org/10.1093/bib/bbaa232
95. O'Toole Á, Scher E, Underwood A. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* 2021;**7**:veab064. https://doi.org/10.1093/ve/veab064
96. Truong Nguyen PT, Plyusnin I, Sironen T. *et al.* HAVoC, a bioinformatic pipeline for reference-based consensus assembly and lineage assignment for SARS-CoV-2 sequences. *BMC Bioinformatics* 2021;**22**:373. https://doi.org/10.1186/s12859-021-04294-2
97. Bello X, Pardo-Seco J, Gómez-Carballa A. *et al.* CovidPhy: a tool for phylogeographic analysis of SARS-CoV-2 variation. *Environ Res* 2022;**204**:111909–9. https://doi.org/10.1016/j.envres.2021.111909
98. Jahshan Z, Yavits L. ViTAL: vision TrAnsformer based low coverage SARS-CoV-2 lineage assignment. *Bioinformatics* 2024;**40**:btae093. https://doi.org/10.1093/bioinformatics/btae093
99. Cacciabue M, Aguilera P, Gismondi MI. *et al.* Covidex: an ultrafast and accurate tool for SARS-CoV-2 subtyping. *Infect Genet Evol* 2022;**99**:105261. https://doi.org/10.1016/j.meegid.2022.105261
100. Lucaci AG, Zehr JD, Shank SD. *et al.* RASCL: rapid assessment of SARS-CoV-2 clades through molecular sequence analysis. *bioRxiv* 2022;2022.01.15.476448. https://doi.org/10.1101/2022.01.15.476448
101. Wilkinson DA, Joffrin L, Lebarbenchon C. *et al.* Analysis of partial sequences of the RNA-dependent RNA polymerase gene as a tool for genus and subgenus classification of coronaviruses. *J Gen Virol* 2020;**101**:1261–9. https://doi.org/10.1099/jgv.0.001494
102. Singer J, Gifford R, Cotten M. *et al.* CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. Preprints.org, Basel, Switzerland, 2020.
103. Oliveira RRM, Costa Negri T, Nunes G. *et al.* PipeCoV: a pipeline for SARS-CoV-2 genome assembly, annotation and variant identification. *PeerJ* 2022;**10**:e13300. https://doi.org/10.7717/peerj.13300
104. Liu B, Liu K, Zhang H. *et al.* CoV-Seq, a new tool for SARS-CoV-2 genome analysis and visualization: Development and usability study. *J Med Internet Res* 2020;**22**:e22299. https://doi.org/10.2196/22299
105. Feng Y, Yi J, Yang L. *et al.* COV2Var, a function annotation database of SARS-CoV-2 genetic variation. *Nucleic Acids Res* 2023;**52**:D701–13. https://doi.org/10.1093/nar/gkad958
106. Bernasconi A, Cilibrasi L, Al Khalaf R. *et al.* EpiSurf: metadata-driven search server for analyzing amino acid changes within epitopes of SARS-CoV-2 and other viral species. *Database* 2021;**2021**:122–32. https://doi.org/10.1093/database/baab059
107. Essabbar A, Kartti S, Alouane T. *et al.* IDbSV: an open-access repository for monitoring SARS-CoV-2 variations and evolution. *Front Med* 2021;**8**:765249–9. https://doi.org/10.3389/fmed.2021.765249
108. Torrens-Fontanals M, Peralta-García A, Talarico C. *et al.* SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions. *Nucleic Acids Res* 2021;**50**:D858–66. https://doi.org/10.1093/nar/gkab977
109. Wu Y, Li L, Wang K. *et al.* COVID-AMD database for coronavirus-infected animal models with comparative analysis tools. *Sci Rep* 2024;**14**:29567. https://doi.org/10.1038/s41598-024-80474-3
110. Trinity L, Wark I, Lansing L. *et al.* Shapify: paths to SARS-CoV-2 frameshifting pseudoknot. *PLoS Comput Biol* 2023;**19**:e1010922. https://doi.org/10.1371/journal.pcbi.1010922
111. Kurata H, Tsukiyama S, Manavalan B. iACVP: markedly enhanced identification of anti-coronavirus peptides using a dataset-specific word2vec model. *Brief Bioinform* 2022;**23**:bbac265. https://doi.org/10.1093/bib/bbac265
112. Kaushik AC, Mehmood A, Selvaraj G. *et al.* CoronaPep: an anti-coronavirus peptide generation tool. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**18**:1299–304. https://doi.org/10.1109/TCBB.2021.3064630
113. Pinoli P, Bernasconi A, Sandionigi A. *et al.* VirusLab: a tool for customized SARS-CoV-2 data analysis. *BioTech* 2021;**10**:27. https://doi.org/10.3390/biotech10040027
114. Yue Z, Zhang E, Xu C. *et al.* PAGER-CoV: a comprehensive collection of pathways, annotated gene-lists and gene signatures for

coronavirus disease studies. *Nucleic Acids Res* 2021;**49**:D589–99. https://doi.org/10.1093/nar/gkaa1094

115. Sung I, Lee S, Pak M. *et al.* AutoCoV: tracking the early spread of COVID-19 in terms of the spatial and temporal patterns from embedding space by K-mer based deep learning. *BMC Bioinformatics* 2022;**23**:149. https://doi.org/10.1186/s12859-022-04679-x
116. Carneiro J, Gomes C, Couto C. *et al.* CoV2ID: detection and therapeutics oligo database for SARS-CoV-2. *bioRxiv* 2020; 2020.04.19.048991.
117. Huang Y, Lau SK, Woo PC. *et al.* CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. *Nucleic Acids Res* 2008;**36**:D504–11. https://doi.org/10.1093/nar/gkm754
118. Satyam R, Yousef M, Qazi S. *et al.* COVIDium: A COVID-19 resource compendium. *Database* 2021;**2021**:baab057. https://doi.org/10.1093/database/baab057
119. Shi W, Fan G, Shen Z. *et al.* gcCov: linked open data for global coronavirus studies. *mLife* 2022;**1**:92–5. https://doi.org/10.1002/mlf2.12008
120. Coutinho FH, Zaragoza-Solas A, López-Pérez M. *et al.* RaFAH: host prediction for viruses of bacteria and archaea based on protein content. *Patterns* 2021;**2**:100274. https://doi.org/10.1016/j.patter.2021.100274
121. Galiez C, Siebert M, Enault F. *et al.* WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* 2017;**33**:3113–4. https://doi.org/10.1093/bioinformatics/btx383
122. Pons JC, Paez-Espino D, Riera G. *et al.* VPF-class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. *Bioinformatics* 2021;**37**:1805–13. https://doi.org/10.1093/bioinformatics/btab026
123. Charoenkwan P, Kanthawong S, Schaduangrat N. *et al.* PVPred-SCM: improved prediction and analysis of phage Virion proteins using a scoring card method. *Cells* 2020;**9**:353. https://doi.org/10.3390/cells9020353
124. Hunt M, Gall A, Ong SH. *et al.* IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* 2015;**31**:2374–6. https://doi.org/10.1093/bioinformatics/btv120
125. Yang X, Charlebois P, Gnerre S. *et al.* De novo assembly of highly diverse viral populations. *BMC Genomics* 2012;**13**:475. https://doi.org/10.1186/1471-2164-13-475
126. Libin PJK, Deforche K, Abecasis AB. *et al.* VIRULIGN: fast codon-correct alignment and annotation of viral genomes. *Bioinformatics* 2019;**35**:1763–5. https://doi.org/10.1093/bioinformatics/bty851
127. Schäffer AA, Hatcher EL, Yankie L. *et al.* VADR: validation and annotation of virus sequence submissions to GenBank. *BMC Bioinformatics* 2020;**21**:211. https://doi.org/10.1186/s12859-020-3537-3
128. Starikova EV, Tikhonova PO, Prianichnikov NA. *et al.* Phigaro: high-throughput prophage sequence annotation. *Bioinformatics* 2020;**36**:3882–4. https://doi.org/10.1093/bioinformatics/btaa250
129. Rangel-Pineros G, Almeida A, Beracochea M. *et al.* VIRify: an integrated detection, annotation and taxonomic classification pipeline using virus-specific protein profile hidden Markov models. *PLoS Comput Biol* 2023;**19**:e1011422. https://doi.org/10.1371/journal.pcbi.1011422
130. Shean RC, Makhsous N, Stoddard GD. *et al.* VAPiD: a lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. *BMC Bioinformatics* 2019;**20**:48. https://doi.org/10.1186/s12859-019-2606-y
131. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* 2017;**22**:30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494
132. Borges V, Pinheiro M, Pechirra P. *et al.* INSaFLU: an automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance. *Genome Med* 2018;**10**:46. https://doi.org/10.1186/s13073-018-0555-0
133. Hayer J, Jadeau F, Deléage G. *et al.* HBVdb: a knowledge database for hepatitis B virus. *Nucleic Acids Res* 2013;**41**:D566–70. https://doi.org/10.1093/nar/gks1022
134. Mulder AC, Kroneman A, Franz E. *et al.* HEVnet: a one health, collaborative, interdisciplinary network and sequence data repository for enhanced hepatitis E virus molecular typing, characterisation and epidemiological investigations. *Eurosurveillance* 2019;**24**:1800407. https://doi.org/10.2807/1560-7917.ES.2019.24.10.1800407
135. Lumley S, Noble H, Hadley MJ. *et al.* Hepitopes: a live interactive database of HLA class I epitopes in hepatitis B virus. *Wellcome Open Res* 2016;**1**:9–9. https://doi.org/10.12688/wellcomeopenres.9952.1
136. Tokuyama M, Kong Y, Song E. *et al.* ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc Natl Acad Sci* 2018;**115**:12565–72. https://doi.org/10.1073/pnas.1814589115
137. Lefkowitz EJ, Dempsey DM, Hendrickson RC. *et al.* Virus taxonomy: the database of the international committee on taxonomy of viruses (ICTV). *Nucleic Acids Res* 2018;**46**:D708–17. https://doi.org/10.1093/nar/gkx932
138. Nayfach S, Camargo AP, Schulz F. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 2021;**39**:578–85. https://doi.org/10.1038/s41587-020-00774-7
139. Roux S, Páez-Espino D, Chen IMA. *et al.* IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res* 2021;**49**:D764–75. https://doi.org/10.1093/nar/gkaa946
140. Tang G, Cho M, Wang X. OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. *Nucleic Acids Res* 2022;**50**:D1334–9. https://doi.org/10.1093/nar/gkab970
141. Goodacre N, Aljanahi A, Nandakumar S. *et al.* A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *mSphere* 2018;**3**:e00069–18. https://doi.org/10.1128/mSphereDirect.00069-18
142. Canakoglu A, Pinoli P, Bernasconi A. *et al.* ViruSurf: an integrated database to investigate viral sequences. *Nucleic Acids Res* 2021;**49**:D817–24. https://doi.org/10.1093/nar/gkaa846
143. Tang D, Li B, Xu T. *et al.* VISDB: a manually curated database of viral integration sites in the human genome. *Nucleic Acids Res* 2020;**48**:D633–41. https://doi.org/10.1093/nar/gkz867
144. Grazziotin AL, Koonin EV, Kristensen DM. Prokaryotic virus orthologous groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res* 2017;**45**:D491–8. https://doi.org/10.1093/nar/gkw975
145. Cai Z, Fan Y, Zhang Z. *et al.* VirusCircBase: a database of virus circular RNAs. *Brief Bioinform* 2021;**22**:2182–90. https://doi.org/10.1093/bib/bbaa052

146. Tang MW, Liu TF, Shafer RW. The HIVdb system for HIV-1 genotypic resistance interpretation. *Intervirology* 2012;**55**:98–101. https://doi.org/10.1159/000331998
147. Charon J, Buchmann JP, Sadiq S. *et al.* RdRp-scan: a bioinformatic resource to identify and annotate divergent RNA viruses in metagenomic sequence data. *Virus Evol* 2022;**8**:veac082. https://doi.org/10.1093/ve/veac082
148. Chao DL, Halloran ME, Obenchain VJ. *et al.* FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol* 2010;**6**:e1000656. https://doi.org/10.1371/journal.pcbi.1000656
149. Tcherepanov V, Ehlers A, Upton C. Genome annotation transfer utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics* 2006;**7**:150. https://doi.org/10.1186/1471-2164-7-150
150. Yusim K, Richardson R, Tao N. *et al.* Los alamos hepatitis C immunology database. *Appl Bioinformatics* 2005;**4**:217–25. https://doi.org/10.2165/00822942-200504040-00002
151. Roux S, Tournayre J, Mahul A. *et al.* Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* 2014;**15**:76. https://doi.org/10.1186/1471-2105-15-76
152. Rozanov M, Plikat U, Chappey C. *et al.* A web-based genotyping resource for viral sequences. *Nucleic Acids Res* 2004;**32**:W654–9. https://doi.org/10.1093/nar/gkh419
153. Brister JR, Ako-Adjei D, Bào Y. *et al.* NCBI viral genomes resource. *Nucleic Acids Res* 2014;**43**:D571–7. https://doi.org/10.1093/nar/gku1207
154. Bao Y, Bolotov P, Dernovoy D. *et al.* The influenza virus resource at the National Center for biotechnology information. *J Virol* 2008;**82**:596–601. https://doi.org/10.1128/JVI.02005-07
155. Angly F, Rodriguez-Brito B, Bangor D. *et al.* PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* 2005;**6**:41. https://doi.org/10.1186/1471-2105-6-41
156. Maes P, Matthijnssens J, Rahman M. *et al.* RotaC: a web-based tool for the complete genome classification of group a rotaviruses. *BMC Microbiol* 2009;**9**:238. https://doi.org/10.1186/1471-2180-9-238
157. Kosakovsky Pond SL, Posada D, Stawiski E. *et al.* An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS Comput Biol* 2009;**5**:e1000581. https://doi.org/10.1371/journal.pcbi.1000581
158. Simmonds P. SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 2012;**5**:50. https://doi.org/10.1186/1756-0500-5-50
159. Shepherd CM, Borelli IA, Lander G. *et al.* VIPERdb: a relational database for structural virology. *Nucleic Acids Res* 2006;**34**:D386–9. https://doi.org/10.1093/nar/gkj032
160. Pickett BE, Sadat EL, Zhang Y. *et al.* ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res* 2012;**40**:D593–8. https://doi.org/10.1093/nar/gkr859
161. Hulo C, de Castro E, Masson P. *et al.* ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res* 2011;**39**:D576–82. https://doi.org/10.1093/nar/gkq901
162. Deng W, Nickle DC, Learn GH. *et al.* ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics* 2007;**23**:2334–6. https://doi.org/10.1093/bioinformatics/btm331
163. Wommack KE, Bhavsar J, Polson SW. *et al.* VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand Genomic Sci* 2012;**6**:427–39. https://doi.org/10.4056/sigs.2945050
164. Hatcher EL, Zhdanov SA, Bao Y. *et al.* Virus variation resource – improved response to emergent viral outbreaks. *Nucleic Acids Res* 2017;**45**:D482–90. https://doi.org/10.1093/nar/gkw1065
165. Alcantara LCJ, Cassol S, Libin P. *et al.* A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res* 2009;**37**:W634–42. https://doi.org/10.1093/nar/gkp455
166. Shen H, Chou K-C. Virus-PLoc: a fusion classifier for predicting the subcellular localization of viral proteins within host and virus-infected cells. *Biopolymers* 2007;**85**:233–40. https://doi.org/10.1002/bip.20640
167. Wang Q, Jia P, Zhao Z. VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS One* 2013;**8**:e64465. https://doi.org/10.1371/journal.pone.0064465
168. Chatr-aryamontri A, Ceol A, Peluso D. *et al.* VirusMINT: a viral protein interaction database. *Nucleic Acids Res* 2009;**37**:D669–73. https://doi.org/10.1093/nar/gkn739
169. Chen Y, Yao H, Thompson EJ. *et al.* VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics* 2013;**29**:266–7. https://doi.org/10.1093/bioinformatics/bts665
170. Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 2008;**24**:2395–6. https://doi.org/10.1093/bioinformatics/btn429
171. Montiel-Garcia D, Santoyo-Rivera N, Ho P. *et al.* VIPERdb v3.0: a structure-based data analytics platform for viral capsids. *Nucleic Acids Res* 2021;**49**:D809–16. https://doi.org/10.1093/nar/gkaa1096
172. Gregory AC, Zablocki O, Zayed AA. *et al.* The gut Virome database reveals age-dependent patterns of Virome diversity in the human gut. *Cell Host Microbe* 2020;**28**:724–740.e8. https://doi.org/10.1016/j.chom.2020.08.003
173. Puschnik AS, Majzoub K, Ooi YS. *et al.* A CRISPR toolbox to study virus–host interactions. *Nat Rev Microbiol* 2017;**15**:351–64. https://doi.org/10.1038/nrmicro.2017.29
174. Fernandez-Pozo N, Rosli Hernan G, Martin Gregory B. *et al.* The SGN VIGS tool: user-friendly software to design virus-induced gene silencing (VIGS) constructs for functional genomics. *Mol Plant* 2015;**8**:486–8. https://doi.org/10.1016/j.molp.2014.11.024



*Briefings in Bioinformatics*, 2025, **26(2)**, bbaf084
https://doi.org/10.1093/bib/bbaf084
Review