

ViralZone 2024 provides higher-resolution images and advanced virus-specific resources

Edouard De Castro , Chantal Hulo, Patrick Masson, Andrea Auchincloss , Alan Bridge and Philippe Le Mercier *

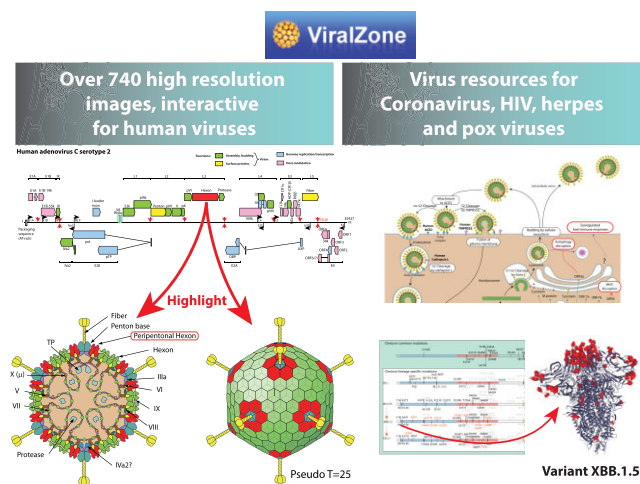
Swiss-Prot group, SIB Swiss Institute of Bioinformatics, CMU, 1 Michel Servet, 1211 Geneva 4, Switzerland

*To whom correspondence should be addressed. Tel: +41 22 379 58 70; Fax: +41 22 379 58 58; Email: Philippe.lemercier@sib.swiss

Abstract

ViralZone (<http://viralzone.expasy.org>) is a knowledge repository for viruses that links biological knowledge and databases. It contains data on virion structure, genome, proteome, replication cycle and host-virus interactions. The new update provides better access to the data through contextual popups and higher resolution images in Scalable Vector Graphics (SVG) format. These images are designed to be dynamic and interactive with human viruses to give users better access to the data. In addition, a new coronavirus-specific resource provides regularly updated data on variants and molecular biology of SARS-CoV-2. Other virus-specific resources have been added to the database, particularly for HIV, herpesviruses and poxviruses.

Graphical abstract



Introduction

ViralZone is an online resource and database that provides information about viruses, particularly their molecular biology, taxonomy and the diseases they cause. It is a valuable tool for researchers, students and anyone interested in learning more about viruses. ViralZone provides a comprehensive taxonomy of viruses, classifying them into different families, genera and species according to the International Committee on Taxonomy of Viruses (ICTV) (1). The database contains 879 virus description pages with detailed descriptions of viruses from different virus families or genera, including their genome structure, replication cycle, host range and associated diseases. It also contains information on the proteins produced by viruses, their functions and their role in the viral life cycle; all are linked to UniProt entries (2).

In addition, there are 352 pages describing viral molecular processes: from transcription, replication to icosahedral viral structure and host-virus interactions. All data come from publications, textbooks and feedback from experts to digitise the global knowledge of viruses. The ViralZone pages provide contextual links to many resources, the most important being: UniProt (2), National Center for Biotechnology Information (NCBI) (3), Virus Pathogen Resource (ViPR) (4), ICTV (1), Gene Ontology (GO) (5), RCSB Protein Data Bank (RCSB PDB) (6), Chemical Entities of Biological Interest (ChEBI) (7) and ChEMBL (8).

Overall, ViralZone provides educational materials and resources to help users learn more about virology and related topics. The database is maintained and regularly updated by the Swiss Institute of Bioinformatics (SIB) and is freely avail-

Received: September 18, 2023. Revised: October 9, 2023. Editorial Decision: October 10, 2023. Accepted: October 12, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

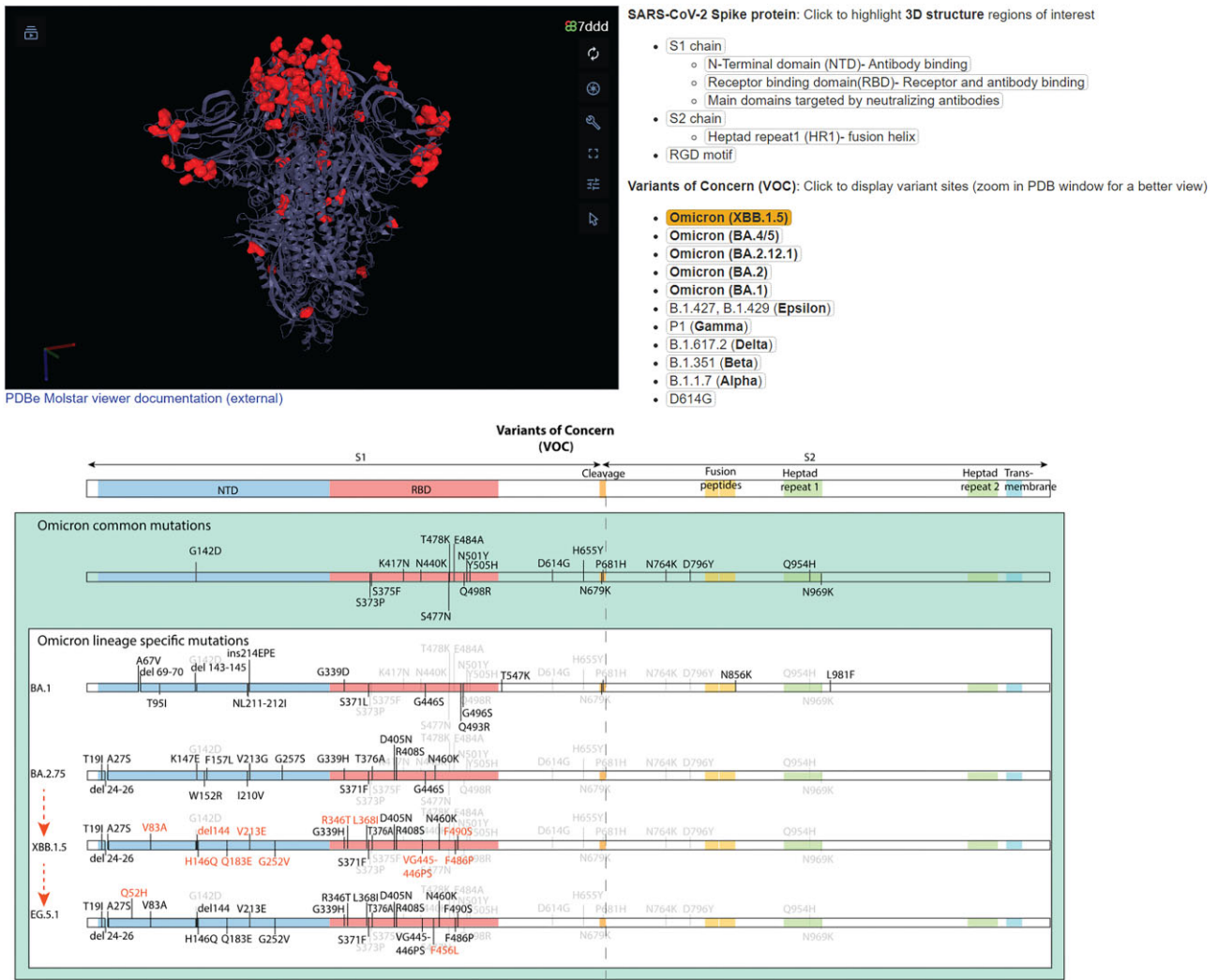


Figure 2. Page of SARS-CoV-2 variants. The page shows a dynamic 3D structure of the PDB:7DDD spike protein with Mol* Viewer (16). The mutations in each spike variant are highlighted in red after selecting the variant in the right list. At the bottom of the image is the graphical representation of the spike mutations present in the last major circulating variants.

and added javascripts to the associated pages to add interactivity. The parts of the virion can be highlighted by mouseover, making it easier to see the global localization of each protein in the viral structure. In addition, mouseover automatically highlights the coding gene in the genome map. When hovering over a gene in the genome map, a UniProt popup appears displaying abstracted data of the encoded protein and, if available, a structure image. If this gene encodes a structural product, the corresponding proteins are highlighted in the virion image (Figure 1).

Specific virus resources

The first specific virus resource was developed for hepatitis B virus (HBV) (9), but resources have since been developed for HIV (10), herpesviruses, coronaviruses and monkeypox virus. These resources provide detailed data focused on specific viruses of global importance. They provide further details on the molecular biology of viruses with emphasis on the following: Replication cycle, host-virus interactions, antiviral drugs, transcription and translation processes in detail.

The advent of COVID has highlighted the importance of rapid and accurate access to expertise on viruses. Research and medicine must have access to knowledge and data to develop accurate research, diagnostics, vaccines and therapeutics. To address this need, dedicated resources for SARS-CoV-2 viruses have been developed in ViralZone. The resource provides curated data on the biology of the virus: genome, transcriptome, proteome, enzymes and replication cycle; known antiviral drugs; vaccines; and links to epidemiological data (Nextstrain) (11). The variant page describes all major circulating variants that are or have been identified as of concern or of interest by World Health Organization (WHO) (12). The list is updated monthly as new variants of concern emerge (Figure 2). Each variant is linked several resources: CoVariants (<https://covariants.org>), Pango Lineages (<https://cov-lineages.org>), outbreak.info (13) and BV-BRC (14). Moreover, reference sequences are selected for each variant in the international nucleotide sequence database (INSDC) (15) in collaboration with Nextstrain (11) and ViPR (14) so that the major virus databases provide the same references.

The Vaccines page describes the major vaccines developed during the pandemic (17), and details the vector used and

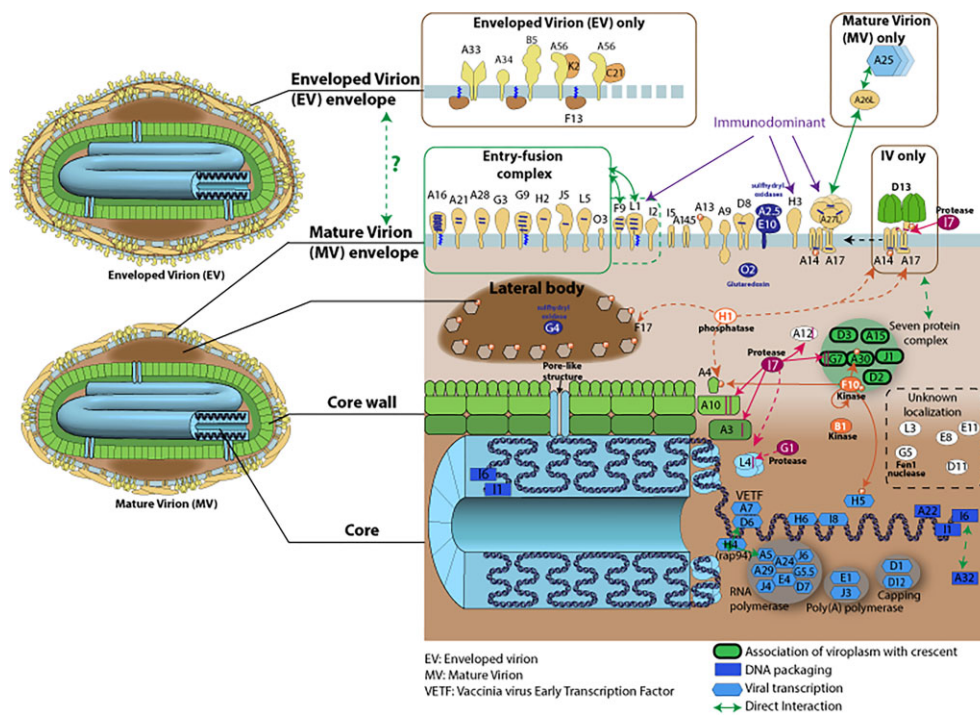


FIGURE 3. Structure of the poxvirus virion. The image shows the enveloped or mature virion with a detailed view of the internal structure with all viral proteins involved. The genes are named according to the vaccinia virus Western Reserve nomenclature.

the modifications added to the spike protein. The Interactome page describes the major functional interactions between host and virus, i.e. interactions whose function has been experimentally demonstrated to play a role in the viral cycle. Many other interactions have been suggested by large-scale or preliminary experiments and can be found in external sources.

The resources for HIV, herpesviruses, coronaviruses and monkeypox contain the same core data: Genome, proteome, host-virus interactions, replication cycle, and vaccines and antiviral drugs, as applicable. In addition, some more virus-specific information has been added: ‘HIV and Tuberculosis syndemic’ describes the mechanisms of coinfection with *Mycobacterium tuberculosis* (MTB) (18); the pages on virion organization of poxviruses and herpesviruses describe the complex internal composition of these virus particles (Figure 3) (19,20).

Additional data update

Several updates were made throughout the database, most notably improving the data in the Human Viruses table and adding a Vertebrate Host Receptor table. The Human Viruses and associated pathologies page (<https://viralzone.expasy.org/678>) was updated to link all viral diseases to their corresponding Wikipedia pages, which can be previewed using the new pop-up system. In addition, genomic links have been simplified using NCBI genome assembly, which provides a single stable access to genome assembly data for monopartite or segmented genomes (21).

The Virus Host Receptor table contains comprehensive data on cellular components that function as receptors for vertebrate virus entry (<https://viralzone.expasy.org/5356>). Only functionally demonstrated interactions were selected, and the table contains 270 virus-receptor interactions. The table in-

cludes data on the host involved, the interacting virus-cell components, a reference publication, and additional data such as the expression of organs of the cellular receptor that can be compared with the tropism of the virus.

Data availability

ViralZone resource can be accessed at <https://viralzone.expasy.org/>. Releases are published every month. All pictures and data are licensed under a Creative Commons Attribution 4.0 International License.

Acknowledgements

Many thanks to Urs Gruber, Rich Condit, Nissin Moussatche and Bernard Moss for feedback on adenovirus and poxvirus virions; to Richard Neher, Emma Hodcroft for SARS-CoV-2 variants; and to Jerven Boleman for references on svg.

Funding

This activity of the Swiss-Prot group is supported by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI; the HIV resource has been funded by s supported by the Swiss South African Joint Research Programme (SSJRP) research grant entitled ‘Swiss Prot/South Africa: Protein Bioinformatics Resource Development for Important Health-related Pathogens’. Funding for open access charge: SIB Swiss Institute of Bioinformatics.

Conflict of interest statement

None declared.

References

- Lefkowitz,E.J., Dempsey,D.M., Hendrickson,R.C., Orton,R.J., Siddell,S.G. and Smith,D.B. (2018) Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.*, **46**, D708–D717.
- UniProt Consortium (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
- Sayers,E.W., Bolton,E.E., Brister,J.R., Canese,K., Chan,J., Comeau,D.C., Connor,R., Funk,K., Kelly,C., Kim,S., *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **50**, D20–D26.
- Pickett,B.E., Sadat,E.L., Zhang,Y., Noronha,J.M., Squires,R.B., Hunt,V., Liu,M., Kumar,S., Zaremba,S., Gu,Z., *et al.* (2012) ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.*, **40**, D593–D598.
- Ontology Consortium,G., Aleksander,S.A., Balhoff,J., Carbon,S., Cherry,J.M., Drabkin,H.J., Ebert,D., Feuermann,M., Gaudet,P., Harris,N.L., *et al.* (2023) The Gene Ontology knowledgebase in 2023. *Genetics*, **224**, iyad031.
- Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Hastings,J., Owen,G., Dekker,A., Ennis,M., Kale,N., Muthukrishnan,V., Turner,S., Swainston,N., Mendes,P. and Steinbeck,C. (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
- Mendez,D., Gaulton,A., Bento,A.P., Chambers,J., De Veij,M., Félix,E., Magariños,M.P., Mosquera,J.F., Mutowo,P., Nowotka,M., *et al.* (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
- Masson,P., Hulo,C., De Castro,E., Bitter,H., Gruenbaum,L., Essioux,L., Bougueleret,L., Xenarios,I. and Le Mercier,P. (2012) ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Res.*, **41**, D579–D583.
- Druce,M., Hulo,C., Masson,P., Sommer,P., Xenarios,I., Le Mercier,P. and De Oliveira,T. (2016) Improving HIV proteome annotation: new features of BioAfrica HIV Proteomics Resource. *Database (Oxford)*, **2016**, 1–13.
- Hadfield,J., Megill,C., Bell,S.M., Huddleston,J., Potter,B., Callender,C., Sagulenko,P., Bedford,T. and Neher,R.A. (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, **34**, 4121–4123.
- WHO (2021) Classification of Omicron (B.1.1.529): sARS-CoV-2 Variant of Concern.
- Gangavarapu,K., Latif,A.A., Mullen,J.L., Alkuzweny,M., Hufbauer,E., Tsueng,G., Haag,E., Zeller,M., Aceves,C.M., Zaiets,K., *et al.* (2023) Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nat. Methods*, **20**, 512–522.
- Olson,R.D., Assaf,R., Brettin,T., Conrad,N., Cucinell,C., Davis,J.J., Dempsey,D.M., Dickerman,A., Dietrich,E.M., Kenyon,R.W., *et al.* (2023) Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.*, **51**, D678–D689.
- Arita,M., Karsch-Mizrachi,I. and Cochrane,G. (2021) The international nucleotide sequence database collaboration. *Nucleic Acids Res.*, **49**, D121–D124.
- Sehna,D., Bittrich,S., Deshpande,M., Svobodová,R., Berka,K., Bazgier,V., Velankar,S., Burley,S.K., Koča,J. and Rose,A.S. (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.*, **49**, W431–W437.
- Heinz,F.X. and Stiasny,K. (2021) Distinguishing features of current COVID-19 vaccines: knowns and unknowns of antigen presentation and modes of action. *NPJ Vaccines*, **6**, 104.
- Kwan,C.K. and Ernst,J.D. (2011) HIV and tuberculosis: a deadly human syndemic. *Clin. Microbiol. Rev.*, **24**, 351–376.
- Condit,R.C., Moussatche,N. and Traktman,P. (2006) In a nutshell: structure and assembly of the vaccinia virion. *Adv. Virus Res.*, **66**, 31–124.
- Heming,J.D., Conway,J.F. and Homa,F.L. (2017) Herpesvirus capsid assembly and DNA packaging. *Adv. Anat. Embryol. Cell Biol.*, **223**, 119–142.
- Kitts,P.A., Church,D.M., Thibaud-Nissen,F., Choi,J., Hem,V., Sapojnikov,V., Smith,R.G., Tatusova,T., Xiang,C., Zherikov,A., *et al.* (2016) Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.*, **44**, D73–D80.