

## VirJenDB: the comprehensive virus database based in Jena

Recent advances in sequencing technologies have resulted in a deluge of virus sequences, lately consisting mostly of SARS-Cov-2, into the global repositories such as the INSDC[1]:ENA[2] and GISAID[3]. However, many other viruses are under sequenced, and metadata is often lacking which hampers the re-use of the data[4]. As a service and upcoming Use Case of the NFDI4Microbiota consortium[5], we have developed at FSU Jena[6] the VirJenDB database, a web-based platform for the re-use and analysis of publicly available metadata and sequences from all viruses following FAIR[7] and Open Science[8] principles.

Our goals are 1) to build a lasting infrastructure with useful features based on regular feedback from virus researchers, 2) to contribute to metadata standards through the curation of the VirJenDB dataset and 3) to support virus researchers in gaining and disseminating research data management skills.

We ingested publicly available virus sequences and metadata from BV-BRC[9]; NCBI Virus[10]; ICTV[11]; and ViralZone[12]. The source data was integrated into a virus data model organized in a MySQL database with source code and documentation available on GitHub[13]. VirJenDB was built as an OpenStack project on de.NBI[14] and uses the Aruna storage system[15]. The frontend was powered by the Java REACT[16] and node.js[17] frameworks. Key features of the current VirJenDB include semantic search, taxonomy browser and download of virus sequences, statistical figures and integrated metadata. The beta version web interface can be accessed at <https://virjendb.uni-jena.de>.

Upcoming developments and areas of improvement include automation of data ingestion and the addition of gene annotations, sequence alignments, metagenome (mg) sequences and mg-derived genomes. Further, we plan to integrate the following tools: sequence search and automatic and community curation, as well as provide subsets of virus sequences for external use in bioinformatic pipelines. We envision a secure workbench for users to upload and analyze their own virus data with additional tools such as phylogenetic and variation analyses. VirJenDB will help to advance global virus research by integrating virus research data, and connecting researchers with relevant tools and training.

### References

1. <https://www.insdc.org>
2. <https://www.ebi.ac.uk/ena/browser/home>
3. <https://www.gisaid.org>
4. <https://www.doi.org/10.3390/v15091834>
5. <https://www.nfdi4microbiota.de>
6. <https://www.uni-jena.de>
7. <https://www.doi.org/10.1038/sdata.2016.18>
8. <https://www.unesco.org/en/open-science?hub=686>
9. <https://www.bv-brc.org>
10. <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#>
11. <https://ictv.global>
12. <https://viralzone.expasy.org>
13. <https://www.github.com>
14. <https://www.denbi.de>
15. <https://aruna-storage.org>
16. <https://react.dev>
17. <https://nodejs.org/en>

Keywords: NFDI4Microbiota, RNA, DNA, Virus, Database, Genome, Phage, Bacteriophage, Next-Generation Sequencing, Genome, Metadata, FAIR, RDM

### Type of Poster

A solution

**Primary author:** CASSMAN, Noriko (Friedrich Schiller University)

**Co-authors:** ZIRAKSAZ, Hamed (FSU Jena); MARZ, Manuela (FMI, FSU Jena); SAGHAEI, Shahram (FSU Jena)

**Presenter:** CASSMAN, Noriko (Friedrich Schiller University)

**Session Classification:** Poster

**Track Classification:** Poster