Editorial: Advances in viromics: new tools, challenges, and data towards characterizing human and environmental viromes

Mária Džunková[1]*, Cristina Moraru[2] and Karthik Anantharaman[3]

[1]Institute for Integrative Systems Biology, University of Valencia and Consejo Superior de Investigaciones Científicas (CSIC), Valencia, Spain, [2]Environmental Metagenomics, Faculty of Chemistry, Research Center One Health Ruhr of the University Alliance Ruhr, University of Duisburg-Essen, Essen, Germany, [3]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, United States

Editorial on the Research Topic
Advances in viromics: new tools, challenges, and data towards characterizing human and environmental viromes

Bacteriophages (viruses of bacteria) are incredibly diverse. They can either reduce the population of their bacterial host, or carry on infection without causing the cell lysis, which can be even beneficial to their host (Chevallereau et al., 2022). Studying bacteriophages is important for understanding how they regulate bacterial populations, which can have practical applications in healthcare, agriculture, biotechnology, and environmental management (Domingo-Calap et al., 2016; Ye et al., 2019). While most of the knowledge on bacteriophage biology has been gathered from experiments with cultures of phages in laboratories, the largest diversity of bacteriophages has been revealed by DNA sequencing of environmental samples, in which most of the bacteria are yet-to-be-discovered and yet-to-be-cultured (Bodor et al., 2020). Bacteria harbor universal marker genes which help to estimate their diversity and target environments with a large proportion of novel bacterial taxa, however, this is not the case of the bacteriophages (Low et al., 2019). The lack of a universal marker gene has been hindering bacteriophage research in the beginning of the viromics era, nevertheless, fortunate advancements in this field in the recent years have been achieved by a multitude of emerging computational tools.

The expansion of the number of computational tools for analysis of viromes is triggered by an urgent need to improve detection of novel types of bacteriophages, to predict functions of phages containing large portion of novel proteins, and to adapt the taxonomic classification system to the large number of emerging phages (Andrade-Martínez et al., 2022). Thanks to these computational efforts, researchers can conduct comprehensive studies on how bacteriophages influence various biological aspects in diverse environments without the need to culture the phages. However, the effectiveness of the current computational tools is affected by various factors, including distinct training sets of samples utilized during their development, diverse reference sequence databases, and variations in how they handle bacterial sequence contamination. The objective of the Research Topic "*Advances in viromics: new tools, challenges, and data toward characterizing human and environmental viromes*" is to address the emerging challenges in the field of viromics.

The most challenging issue in the computational analysis of viromes is choosing the best tool for the detection and selection of phage contigs. The study of Schackart et al. evaluates the performance of five homology-based tools: VirSorter (Roux et al., 2015), MARVEL (Amgarten et al., 2018), viralVerify (Antipov et al., 2020), VIBRANT (Kieft et al., 2020), and VirSorter2 (Guo et al., 2021), and three sequence-based tools: VirFinder (Ren et al., 2017), DeepVirFinder (Ren et al., 2020), Seeker (Auslander et al., 2020). The initial study design involved 19 tools, however, 11 of them were discarded due to runtime exceptions, hard-coded paths, lack of clear documentation, scalability issues, and inability to run instances on different cores, which are the common reasons why many bioinformatic tools are not widely used by the scientific community. The eight remaining tools were assessed in terms of their precision, sensitivity and specificity. The effects of contig length, phage taxonomy, sequencing and assembly errors, eukaryotic contamination, low viral content, and sequencing error were examined. Their performance was evaluated using four benchmark datasets: (1) set of fragments from complete genomes, (2) phageome set created from simulated reads from phage genomes and then assembled and binned, (3) simulated metagenome set created the same way as the phageome, but with marine samples, (4) colorectal cancer and gut virome datasets coming from real metagenomes.

All tools, especially the homology-based tools, showed reduced sensitivity for non-caudovirales sequences. Similarly, the low viral abundance in simulated metagenomes decreased precision of all tools, although the homology-based tools performed slightly better and they were also more robust to eukaryotic contamination. In contrast, the sequence-based tools performed better with shorter contigs. It is important to note that in the case of real metagenomes (non-simulated data), there was only a little overlap between the eight tested tools. The results suggest that the choice of tools depends on the specific research question and the type of metagenomes being analyzed. Schackart et al. recommend using DeepVirFinder for analysis of purified viromes due to its high sensitivity. DeepVirFinder and Metaphinder were shown to be very efficient in identification of novel phages. For studying dominant phages with a high confidence, VirSorter2 or viralVerify are suggested. In summary, a multi-tool approach combining sequence-based and homology-based tools is recommended. It is also important to keep in mind that other aspects of performance, such as parameter tuning, database update, and ability to detect prophage and plasmid sequences might be considered in decision making when selecting the most suitable computational tool for virome analysis. We believe that conducting similar studies on a regular basis will be very useful for keeping up to date with the emerging computational tools, and constant updates of the phage sequence databases and phage taxonomy.

The importance of automatic updates in the rapidly expanding field of viromics is underlined in the report of Albrycht et al.. This team developed Phage & Host Daily (PHD) web application (http://phdaily.info) that accesses reliable information on phage host specificity by publishing daily reports on phage-host interactions. The information is extracted from three large databases: International Committee on Taxonomy of Viruses (ICTV, Krupovic et al., 2021), National Center for Biotechnology Information (NCBI, Sayers et al., 2019) and the Genome Taxonomy Database (GTDB, Parks et al., 2022). The website allows users to search for specific taxa or browse taxonomic trees to access information on virus-host interactions. The results are presented as a table of pairwise interactions with details on source databases, taxonomic affiliations, genome composition, and assembly completeness. Additionally, PHD supports the construction of custom datasets for machine learning algorithms, including viruses from metagenomic sequences. The continuous updates of PHD align with taxonomic changes, contributing to a comprehensive classification scheme for phage research.

Viral taxonomy has undergone major transformations recently, with the Baltimore Classification being replaced by the "Megataxonomy of Viruses" (Koonin et al., 2020), in which viruses are grouped into 6 different realms, based on their evolutionary history. Within each realm, there are further 10 taxonomic levels comprising the scaffold for the hierarchical classification of viruses. In the case of phages, the most dramatic taxonomic changes affected the tailed dsDNA phages. These were all grouped until recently under the orphan *Caudovirales* order, into three families (*Podoviridae, Myoviridae,* and *Syphoviridae*). Under the new taxonomic system, the *Caudovirales* order and the three families were dissolved (Adriaenssens et al., 2021), because accumulated evidence suggested that they are not monophyletic. All tailed dsDNA phages are now classified in the *Duplodnaviria* realm, in the *Caudoviricetes* class. Their classification at the order and family level is currently in progress, and this has increased the need for dedicated bioinformatics tools. In their review, Zhu et al. compared the performance of different tools to assign viral contigs to existing taxons, from family to genus level. The four tools that passed the initial classification criteria (e.g., having updatable databases) were CAT, PhaGCN, MMseq2, and vConTACT 2.0. These were further evaluated using a variety of metrics and several test datasets. Each tool showed both strengths and weaknesses. For example, some tools showed higher accuracy but underperformed in the case of short contigs. Again, the choice of tools to use is in the hands of the researcher performing the study, depending on the scientific questions.

Large sequencing efforts in the last decades have revealed an enormous genetic diversity of bacteriophages and shed light on their ecological roles. Viruses in aquatic systems have received significant attention, but much of these efforts have been focused on marine systems (Coutinho et al., 2017; Breitbart et al., 2018; Ignacio-Espinoza et al., 2019). Consequently, freshwater viruses have remained significantly underexplored. To generate a comprehensive collection of freshwater viral genomes, Elbehery and Deng studied viruses from twelve representative freshwater environments (biomes) ranging from lakes, estuaries, wastewater, groundwater, and applied environments such as ballasts and fishponds. Publicly available metagenomic data from these environments was downloaded and quality control was conducted to ensure the removal of cellular contamination. This was followed by identification of viruses, establishment of viral clusters, prediction of hosts for viruses, and identification of viral auxiliary metabolic genes (AMGs). Overall, the study expands the known diversity of freshwater viruses and sheds light on their ecology in diverse freshwater biomes. Amongst the freshwater biomes

studied, the authors found significant variance in viral diversity between biomes, and also with samples from the same biome. While it is possible that some of these findings may be attributed to insufficient sequencing depth of viruses in these samples, the authors argue that viral diversity is dictated by changing environmental chemistry as posited previously (Adriaenssens et al., 2021) and also suggest a limited complexity of the freshwater core virome as observed in other environments such as the human gut and the oceans which are dominated by double stranded DNA viruses (now *Duplodnaviria* realm) in the class *Caudoviricetes* (Brum et al., 2015; Broecker et al., 2017). Finally, the study identifies a significant complement of AMGs in freshwater viruses. AMGs are host-derived genes on viral genomes that can provide increased fitness and infection efficiency to viruses. Freshwater AMGs identified include genes for degradation of aromatic and xenobiotic compounds and photosynthesis. Overall, Elbehery and Deng provide a collection of viral genomes that can serve as a foundation to investigate the ecology and evolution of freshwater viruses. Future studies can build upon this groundwork to focus on extraction of viromes (purified virus particles), deeper sequencing of freshwater viruses, and a holistic interpretation of freshwater viruses in the context of chemistry and their hosts to understand their roles, dynamics, and importance in freshwater systems.

The last study in this Research Topic highlights the differences between the traditional culture methods and different types of viromics datasets (dsDNA, ssDNA, dsRNA, and ssRNA) by testing wastewater treatment plant samples against strains of *Brevundimonas* and *Serratia.* In this study, Friedrich et al. associated the environmental viruses with these two bacterial host species by sequencing the phage particles harvested after the initial overlay and compared the resulting community to the collection of viruses obtained by traditional plaque assays. It was found that the majority of the dominant phages were successfully isolated, but a considerable diversity remained unavailable, including phages closely related to the phages isolated by the traditional plaque assay. Interestingly, this study provided experimental evidence for the first phage with a class-level host range, vB_SmaM-Otaku, which was found to infect both tested host strains belonging to two different classes of *Pseudomonadota*. Previous computational studies suggested existence of phages with a very broad host range (Paez-Espino et al., 2016), however, this is the first time it has been experimentally confirmed. The study also emphasizes the need for novel approaches to explore the unexplored viral diversity, especially ssDNA and RNA phages due to their unique characteristics and smaller genome sizes compared to DNA phages.

We can anticipate that in the very near future emerging direct RNA-Seq technologies, long-read sequencing, new whole genome amplification methods and novel experimental methods for linking

unculturable phages to their hosts will generate a large amount of knowledge and sequence data (Smith et al., 2022). It is expected that as our knowledge on bacteriophages expands, a corresponding increase in challenges is likely to arise. The recent research efforts have shown that combining DNA sequencing with experimental methods can provide evidence-based discoveries revealing how phages operate within a specific environment, e.g., by highlighting the importance of metabolic auxiliary genes in sulfur ecology (Kieft et al., 2021), or the role of phage-plasmids in spread of antibiotic resistances (Pfeifer et al., 2022). In light of these evolving technologies and innovative research approaches, it is clear that our understanding of phage biology will continue to expand.

Author contributions

MD: Writing—original draft, Writing—review and editing. CM: Writing—original draft, Writing—review and editing. KA: Writing—original draft, Writing—review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adriaenssens, E., Kropinski, A. M., Turner, D., Krupovic, M., Millard, A., Dutilh, B. E., et al. (2021). *ICTV Taxonomy Proposal 2021, 001B.* Abolish the order Caudovirales and the families Myoviridae, Siphoviridae and Podoviridae (Caudoviricetes). [WWW document]. Available online at: https://ictv.global/ictv/proposals/2021.001B.R.abolish_Caudovirales.zip

Amgarten, D., Braga, L., da Silva, and Setubal, A. (2018). MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Front. Genet.* 9, 304. doi: 10.3389/fgene.2018.00304

Andrade-Martínez, J. S., Camelo Valera, L. C., Chica Cárdenas, L. A., Forero-Junco, L., López-Leal, G., Moreno-Gallego, J. L., et al. (2022). Computational tools for

the analysis of uncultivated phage genomes. *Microbiol. Mol. Biol. Rev.* 86, e0000421. doi: 10.1128/mmbr.00004-21

Antipov, D., Raiko, M., Lapidus, A., and Pevzner, P. (2020). Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* 36, 4126–4129. doi: 10.1093/bioinformatics/btaa490

Auslander, N., Gussow, A., Benler, S., Wolf, Y., and Koonin, E. (2020). Seeker: alignment-free identification of bacteriophage genomes by deep learning. *Nuc.c Acids Res.* 48, e121. doi: 10.1093/nar/gkaa856

Bodor, A., Bounedjoum, N., Vincze, G. E., Kis, Á. E., Laczi, K., Bende, G., et al. (2020). Challenges of unculturable bacteria: environmental perspectives. *Rev. Environ. Sci. Biotechnol.* 19, 1–22. doi: 10.1007/s11157-020-09522-4

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3, 754–766. doi: 10.1038/s41564-018-0166-y

Broecker, F., Russo, G., Klumpp, J., and Moelling, K. (2017). Stable core virome despite variable microbiome after fecal transfer. *Gut Microb.* 8, 214–220. doi: 10.1080/19490976.2016.1265196

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., et al. (2015). Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* 348, 1261498. doi: 10.1126/science.1261498

Chevallereau, A., Pons, B. J., van Houte, S., and Westra, E. R. (2022). Interactions between bacterial and phage communities in natural environments. *Nat. Rev. Microbiol.* 20, 49–62. doi: 10.1038/s41579-021-00602-y

Coutinho, F. H., Silveira, C. B., Gregoracci, G. B., Thompson, C. C., Edwards, R. A., Brussaard, C. P. D., et al. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* 8, 1–12. doi: 10.1038/ncomms15955

Domingo-Calap, P., Georgel, P., and Bahram, S. (2016). Back to the future: bacteriophages as promising therapeutic tools. *HLA.* 87, 133–40. doi: 10.1111/tan.12742

Guo, J., Bolduc, B., Zayed, A., Varsani, A., Dominguez-Huerta, G., Delmont, T., et al. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9, 37. doi: 10.1186/s40168-020-00990-y

Ignacio-Espinoza, J. C., Ahlgren, N. A., and Fuhrman, J. A. (2019). Long-term stability and Red Queen-like strain dynamics in marine viruses. *Nat. Microbiol.* 4, 1–7. doi: 10.1038/s41564-019-0628-x

Kieft, K., Zhou, Z., and Anantharaman, K. (2020). VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8, 90. doi: 10.1186/s40168-020-00867-0

Kieft, K., Zhou, Z., Anderson, R. E., Buchan, A., Campbell, B. J., Hallam, S. J., et al. (2021). Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun.* 12, 3503. doi: 10.1038/s41467-021-23698-5

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 4, 19 doi: 10.1128./MMBR.00061-19

Krupovic, M., Turner, D., Morozova, V., Dyall-Smith, M., Oksanen, H. M., Edwards, R., et al. (2021). Bacterial Viruses Subcommittee and Archaeal Viruses Subcommittee of the ICTV: update of taxonomy changes in 2021. *Arch. Virol.* 166, 3239–3244. doi: 10.1007/s00705-021-05205-9

Low, S. J., DŽunková, M., Chaumeil, P. A., Parks, D. H., and Hugenholtz, P. (2019). Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nat. Microbiol.* 4, 1306–1315. doi: 10.1038/s41564-019-0448-z

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., et al. (2016). Uncovering Earth's virome. *Nature* 536, 425–430. doi: 10.1038/nature19094

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P-. A., Hugenholtz, P., et al. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nuc. Acids Res.* 50, D785–D794. doi: 10.1093/nar/gkab776

Pfeifer, E., Bonnin, R. A., and Rocha, E. P. C. (2022). Phage-plasmids spread antibiotic resistance genes through infection and lysogenic conversion. *mBio* 13, e0185122. doi: 10.1128/mbio.01851-22

Ren, J., Ahlgren, N., Lu, Y., Fuhrman, J., and Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5, 69. doi: 10.1186/s40168-017-0283-5

Ren, J., Song, K., Deng, C., Ahlgren, N., Fuhrman, J., Li, Y., et al. (2020). Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* 8, 64–77. doi: 10.1007/s40484-019-0187-4

Roux, S., Enault, F., Hurwitz, B., and Sullivan, M. (2015). VirSorter: Mining viral signal from microbial genomic data. *PeerJ* 3, e985. doi: 10.7717/peerj.985

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., Karsch-Mizrachi, I., et al. (2019). GenBank. *Nucleic Acids Res.* 48, D84–D86. doi: 10.1093/nar/gkz956

Smith, S. E., Huang, W., Tiamani, K., Unterer, M., Khan Mirzaei, M., Deng, L., et al. (2022). Emerging technologies in the study of the virome. *Curr. Opin. Virol.* 54, 101231. doi: 10.1016/j.coviro.2022.101231

Ye, M., Sun, M., Huang, D., Zhang, Z., Zhang, H., Zhang, S., et al. (2019). A review of bacteriophage therapy for pathogenic bacteria inactivation in the soil environment. *Environ. Int.* 129, 488–496. doi: 10.1016/j.envint.05062