

# Peer Community Journal

Section: Microbiology

Research article

Published  
2026-05-28

Cite as

Alexander G Lucaci, Hisham Shaikh, Li Chong Chong, Rachid Tahzima, Marco Forgia, Karima Mansour, Shoichi Sakaguchi, So Nakagawa, Xin Hou, Tatiana Demina, Philmar Raj Jayaraj Mallika, Anne Kupczok, Spyros Lytras, Humberto Debat, Justine Charon, Michael Urzo, Milica Raco, Rachel Kim, Ricardo Rivero, Dimitris Karapliafis, Leyla Sirkinti, Laura Luebbert, Luca Nishimura, Rayan Chikhi, Lander De Coninck, Florian Charriat, Emma Soufir, Vladimir Gajdov, Thomas Krannich, Gytis Dudas, Cédric Lood, Josue Rodriguez-Ramos, Anja Pecman, Uri Neri, Almut Werner, Mia Le, Bolaji Osundahunsi, Nils Petersen, François Maclot, Serafin Gutierrez, Sofia Paraskevopoulou, Luke Hillary and Ingrida Olendraite (2026) *Consensus statement from the second RdRp Summit: towards a unified framework for RNA virus biology*, Peer Community Journal, 6: e50.

Correspondence  
agl4001@med.cornell.edu  
ingridaolendraite@gmail.com

Peer-review  
Peer reviewed and  
recommended by  
PCI Microbiology,  
<https://doi.org/10.24072/pci.microbiol.100236>

  
This article is licensed  
under the Creative Commons  
Attribution 4.0 License.

## Consensus statement from the second RdRp Summit: towards a unified framework for RNA virus biology

Alexander G Lucaci<sup>1, #, 1</sup>, Hisham Shaikh<sup>2, #, 2</sup>, Li Chong Chong<sup>3, #, 3</sup>, Rachid Tahzima<sup>4, 4</sup>, Marco Forgia<sup>5, 5</sup>, Karima Mansour<sup>6, 6</sup>, Shoichi Sakaguchi<sup>7, 7</sup>, So Nakagawa<sup>8, 8</sup>, Xin Hou<sup>9, 9</sup>, Tatiana Demina<sup>10, 10</sup>, Philmar Raj Jayaraj Mallika<sup>11, 11</sup>, Anne Kupczok<sup>12, 12</sup>, Spyros Lytras<sup>13, 13</sup>, Humberto Debat<sup>14, 14</sup>, Justine Charon<sup>15, 15</sup>, Michael Urzo<sup>16, 16</sup>, Milica Raco<sup>17, 17</sup>, Rachel Kim<sup>18, 18</sup>, Ricardo Rivero<sup>19, 19</sup>, Dimitris Karapliafis<sup>12, 12</sup>, Leyla Sirkinti<sup>20, 20</sup>, Laura Luebbert<sup>21, 21</sup>, Luca Nishimura<sup>22, 22</sup>, Rayan Chikhi<sup>23, 23</sup>, Lander De Coninck<sup>24, 24</sup>, Florian Charriat<sup>25, 25</sup>, Emma Soufir<sup>25, 26</sup>, Vladimir Gajdov<sup>27, 27</sup>, Thomas Krannich<sup>28, 28</sup>, Gytis Dudas<sup>29, 29</sup>, Cédric Lood<sup>30, 30</sup>, Josue Rodriguez-Ramos<sup>31, 31</sup>, Anja Pecman<sup>32, 32</sup>, Uri Neri<sup>33, 33</sup>, Almut Werner<sup>34, 34</sup>, Mia Le<sup>35, 35</sup>, Bolaji Osundahunsi<sup>36, 36</sup>, Nils Petersen<sup>37, 37</sup>, François Maclot<sup>38, 38</sup>, Serafin Gutierrez<sup>25, 39</sup>, Sofia Paraskevopoulou<sup>28, 28</sup>, Luke Hillary<sup>40, 40</sup>, and Ingrida Olendraite<sup>41, 41</sup>

Volume 6 (2026), article e50

<https://doi.org/10.24072/pcjournal.727>

### Abstract

RNA-dependent RNA polymerase, or RdRp, remains the central molecular hallmark of RNA viruses. It serves as both a universal anchor for virus detection and a critical target for understanding the functional and evolutionary properties of RNA viruses. Since the inaugural RdRp summit in 2023, there have been significant advances in sequencing, structural prediction and artificial intelligence, all of which have accelerated the pace of RNA virus discovery and taxonomic annotation, revealing unprecedented levels of viral diversity, including novel phyla and unique genome architectures. Recent advances include the discovery of novel viral phyla such as Ambiviricota and the application of AI-driven models like LucaProt, highlighting both the rapid expansion of viral diversity and the growing role of machine learning in RNA virus research. The second RdRp summit, which was held in Lisbon in May 2025, gathered a group of research scientists from diverse subfields of virology to address emerging challenges in RNA virus biology. These challenges ranged from standardising annotation and data sharing to harnessing structure-guided phylogenetics and petabyte-scale computational tools. Here, our consensus statement outlines key progress, current and future challenges and community-driven initiatives, including benchmarking, virus-host inference, and ongoing knowledge exchange efforts - all of which are designed to unify the field. Importantly, this statement reflects a clear community consensus and provides concrete recommendations to prioritize standardized benchmarking, structure-informed evolutionary analysis, and reproducible virus-host inference as foundational pillars for advancing RNA virus research. By fostering an environment of sustained collaboration, our efforts aim to build a coherent framework for modern RNA virus biology and to accelerate the exploration of the hidden RNA virosphere.

<sup>1</sup>The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA, <sup>2</sup>Research Department, Flanders Marine Institute (VLIZ), InnovOcean Site, Ostend, Belgium, <sup>3</sup>Institute for Experimental Virology, TWINCORE Centre for Experimental and Clinical Infection Research, a Joint Venture Between the Hannover Medical School (MHH) and the Helmholtz Centre for Infection Research (HZI), Hannover, Germany, <sup>4</sup>University of Brussels (VUB) - Bio2Byte / Structural Biology and Artificial Intelligence Lab, Pleinlaan 2 1050 Brussels, Belgium, <sup>5</sup>IPSP-CNR, Via Amendola 122/D, 70126, Bari, Italy, <sup>6</sup>Department of Plant Protection, Faculty of Agrobiological, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic, <sup>7</sup>Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Box 190, Lomma 234 22, Sweden, <sup>8</sup>Department of Microbiology and Infection Control, Faculty of Medicine, Osaka Medical and Pharmaceutical University, Osaka, Japan, <sup>9</sup>Department of Molecular Life Science, Tokai University School of Medicine, Kanagawa, Japan, <sup>10</sup>Institut Pasteur, Université Paris Cité CNRS UMR2000, Evolutionary Genomics of RNA Viruses Unit, Paris, France, <sup>11</sup>University of Helsinki, Faculty of Agriculture and Forestry, Department of Microbiology, Helsinki, Finland, <sup>12</sup>Molecular Virology Laboratory, ICAR-National Research Centre for Banana, Tiruchirappalli, Tamil Nadu, India, <sup>13</sup>Bioinformatics Group, Wageningen University & Research, Wageningen, Netherlands, <sup>14</sup>Division of Systems Virology, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, <sup>15</sup>MRC-University of Glasgow Centre for Virus Research, Glasgow, UK, <sup>16</sup>National Institute of Agricultural Technology, Cordoba, Argentina, <sup>17</sup>Univ. Bordeaux, INRAE, BFP, UMR 1332, F-33140 Villenave d'Ornon, France, <sup>18</sup>Virology Laboratory, Microbiology Division, Institute of Biological Sciences, University of the Philippines Los Baños, Philippines, <sup>19</sup>Oregon State University, Corvallis, USA, <sup>20</sup>Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea / School of Biological Sciences, Seoul National University, Seoul, Republic of Korea, <sup>21</sup>Paul G. Allen School for Global Health, Washington State University, Pullman, Washington, USA, <sup>22</sup>Department of Molecular and Medical Virology, Ruhr University Bochum, Bochum, Germany, <sup>23</sup>Eric and Wendy Schmidt Center & Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, MA, USA, <sup>24</sup>Department of Organismic and Evolutionary Biology, Harvard University, MA, USA, <sup>25</sup>Division of Systems Virology, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, <sup>26</sup>Institut Pasteur, Université Paris Cité, CNRS UMR3525, Paris, France, <sup>27</sup>KU Leuven, Division of Clinical and Epidemiological Virology, Leuven, Belgium, <sup>28</sup>ASTRE, CIRAD, INRAE, Université de Montpellier, Montpellier, France, <sup>29</sup>PPCEI, INSERM, Université de Montpellier, Montpellier, France, <sup>30</sup>Scientific Veterinary Institute Novi Sad, Novi Sad, Serbia, <sup>31</sup>Genome Competence Center, Robert Koch Institute, 13353 Berlin, Germany, <sup>32</sup>Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania, <sup>33</sup>Department of Biology, University of Oxford, Oxford, United Kingdom, <sup>34</sup>Environmental and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA, <sup>35</sup>National Institute of Biology, Department of Biotechnology and Systems Biology, Ljubljana, Slovenia, <sup>36</sup>DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, <sup>37</sup>Institute for Microbiology, Christian-Albrechts-University, Kiel, Germany, <sup>38</sup>Bernhard Nocht Institute for Tropical Medicine (BNITM), Hamburg, Germany, German Center for Infection Research (DZIF), partner site Hamburg-Lübeck-Borstel-Riems, Hamburg, Germany, Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany, <sup>39</sup>Department of Entomology and Plant Pathology, University of Arkansas, USA, <sup>40</sup>Bernhard Nocht Institute for Tropical Medicine (BNITM), Hamburg, Germany, German Center for Infection Research (DZIF), partner site Hamburg-Lübeck-Borstel-Riems, Hamburg, Germany, <sup>41</sup>Fruit Biology and Pathology unit, French National Institute for Agricultural Research (INRAE) / University of Bordeaux, Villenave d'Ornon, France, <sup>42</sup>Laboratory of Virology, Montpellier University Hospital, Montpellier, France, <sup>43</sup>Department of Plant Pathology, University of California Davis, Davis, CA, 95616, USA, <sup>44</sup>Division of Virology, Department of Pathology, University of Cambridge, United Kingdom, <sup>45</sup>VUGENE, Lithuania, <sup>#</sup>Equal contribution

Peer Community Journal is a member of the  
Centre Mersenne for Open Scientific Publishing

<http://www.centre-mersenne.org/>

e-ISSN 2804-3871

## Introduction

RNA-dependent RNA polymerase (RdRp) is a key enzyme in the life cycle of RNA viruses, and serves as the central molecular hallmark of *Orthornavirae*, providing both a conserved anchor for their identification, and a critical target for functional assessment of genome replication, and evolutionary analysis (Koonin et al., 2015). Its functional core, the palm domain, consists of seven conserved motifs (G, F1–3, A, B, C, D, and E, from the N-terminus to the C-terminus) that fold into a palm-like structure accommodating the RNA template during the replication process (Bruenn, 2003). Among these seven motifs, motifs A, B, and C harbor the conserved residues required for synthesis of the new strand (Tahzima et al., 2025). This conservation is shared among all viral RdRps, with the exception of permuted RdRps, in which motif C is located upstream of motifs A and B without disrupting the overall folded structure of the protein.

Due to its inherent role in viral replication and its ubiquitous presence across major RNA virus lineages, RdRp therefore remains one of the most powerful markers for uncovering viral diversity, especially in the era of metagenomic and metatranscriptomic datasets (Charon et al., 2022; Wolf et al., 2018). The relatively conserved catalytic motifs enable detection even across highly divergent sets of viral clades, making it indispensable for the characterization of the immense “viral dark matter” that remains unculturable and unclassified. By recognizing the central role of RdRp in RNA virus biology, a global community of scientific researchers from diverse disciplines including virology, bioinformatics, structural biology, epidemiology, and evolutionary genomics convened for the inaugural RdRp summit in Valencia, Spain, on 22–23 May 2023 (Charon et al., 2024). The 2023 summit established the foundational need for scalable RNA virus discovery tools and catalyzed the development of community resources such as RdRpCATCH and RdRpScan, providing a clear foundation for the 2025 summit’s subsequent focus on standardization and structure-guided phylogenetics. A second RdRp summit took place in May 2025 in Lisbon, Portugal, with the purpose of discussing the latest advancements, and remaining challenges in the identification, annotation, and phylogenetic interpretation of RdRp sequences. Participants from across the globe gathered in order to share benchmarks, discuss emerging tools and define best practices for large-scale RNA virus discovery. This second summit succeeded in fostering new collaborative initiatives to identify and address the emerging challenges of the field.

### Advancements in RdRp Research Since the 2023 Summit

Since the first RdRp summit, several studies have been published in the field of RNA virus discovery and characterization (Figure 1). These studies focus on benchmarking RNA processing methods for High-Throughput Sequencing (HTS) (Schönegger et al., 2023), and on the conceptualization and development of new tools for RdRp identification (Liu et al., 2022) (Figure 2, Top). Technological advances in long-read sequencing have encouraged the publication of studies applying these methods to virome investigations through total RNA sequencing (Pichler et al., 2023). This approach could potentially open new avenues for describing viral diversity at the viral isolate level in the future. More broadly, recent work reflects several emerging trends, including the increasing integration of metatranscriptomic data with ecological and environmental metadata, a shift toward total RNA sequencing approaches to better capture segmented viral genomes, and the growing use of AI-driven homology detection methods to identify highly divergent RdRp sequences. Together, these developments point toward a more comprehensive and data-integrated framework for RNA virus discovery.

Most importantly, progress in the accessibility and scalability of protein structural prediction and modeling has provided the community with powerful tools to investigate the phylogenetic relationships between divergent viruses (Litvin et al., 2025). Widely adopted approaches, including structure prediction tools such as AlphaFold, ESMFold, and ProST5, together with structure-comparison frameworks like Foldseek, have enabled the detection of deep evolutionary relationships that are often inaccessible through sequence-based methods alone. In parallel,

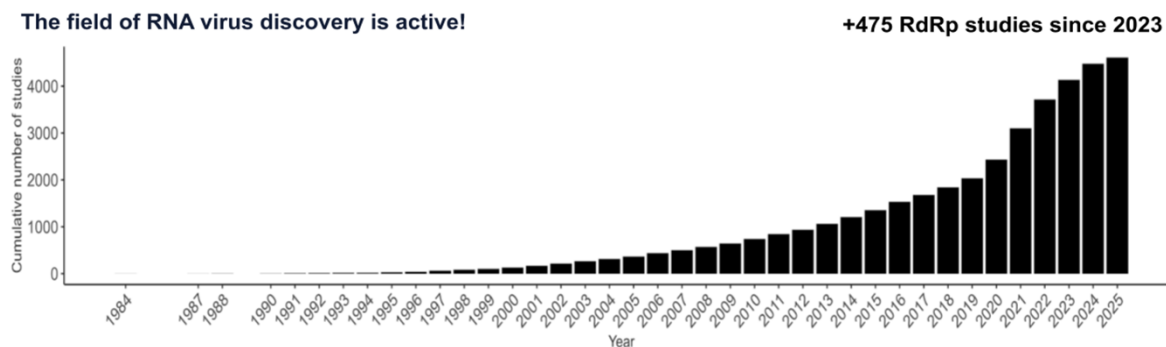
genome completion approaches such as SegFinder are beginning to address challenges associated with fragmented and segmented viral genomes (Heinzinger et al., 2024; Jumper et al., 2021; Lin et al., 2023). These tools also show promise in detecting new clades of RNA viruses with ORFs lacking any detectable homology to known open reading frames (Yin & Fischer, 2006) and completing viral genomes by identifying additional genomic fragments coding for distantly related proteins. These applications are made possible by continuous efforts in the analysis (Chikhi et al., 2025) and annotation (Hou et al., 2024) of global RNAseq databases, which are now available to the RNA virology community.

All these efforts have led to a remarkable expansion of viral taxonomy, both in terms of the number of known viral species and higher-level classifications. Notably, two new phyla have been added to the *Riboviria* realm: *Ambiviricota* (Kuhn et al., 2024) and *Artimaviricota* (Urayama et al., 2024). The discovery of viruses belonging to the phylum *Ambiviricota* also revealed, for the first time, that RNA viruses can utilize a genomic organization based on circular RNA sequences that replicate via a rolling circle mechanism (Forgia et al., 2023). Accelerated discovery and RdRp-based phylogeny of viruses known as bunyavirals led to the establishment of the class *Bunyaviricetes* under the *Negarnaviricota* phylum, accommodating negative-sense RNA viruses.

## Modern challenges in RdRp Biology

### RNA Virus Categorization

Environmental sampling and sequencing has revolutionized RNA virus taxonomy by expanding the known diversity of RNA viruses (Simmonds et al., 2017), effectively doubling the recognized repertoire over the past five years (twofold increase) through discoveries across a wide range of ecosystems and hosts. Many of these newly identified viruses lack known pathogenicity or clear evolutionary ties to existing taxa (Neri et al., 2022; Wolf et al., 2020). In response to this surge in genomic data, the International Committee on Taxonomy of Viruses (ICTV) has adopted phylogenetic analysis as the main criterion for classifying newly identified viruses, even in the absence of biological information (International Committee on Taxonomy of Viruses Executive Committee, 2020). This shift acknowledges the practical challenges of characterising the rapidly growing number of viral sequences but has also sparked debate among virologists about the implications for species definitions and biological relevance (Gibbs, 2020; Neri et al., 2022; Simmonds et al., 2017; Wolf et al., 2020) (Figure 2, Bottom). This taxonomic framework relies on computational methods that can classify metagenomic sequences despite limited biological information. Reference-based methods are often used for taxonomic classification which compare sequences to curated viral databases, along with marker-based methods, which rely on conserved genes in viral clades. For RNA viruses, the RdRp gene is typically used because it is the most conserved genomic region and is widely used to infer evolutionary relationships (Tang et al., 2022). However, the extensive sequence divergence and structural variability of RdRp often complicate accurate reconstruction and limit taxonomic resolution (Holmes & Duchêne, 2019). To address these challenges, Tang et al. developed RdRpBin, a computational tool combining an alignment-based strategy and machine learning models to improve RdRp sequence detection and classification (Tang et al., 2022). Importantly, RdRpBin represents one example of a broader class of alignment-informed machine learning approaches, rather than a singular solution endorsed by the consortium. Complementary tools, including CHEER and VirHunter, further illustrate the diversity of classification frameworks available, while recent phylogeny-aware methods such as PhyloTUNE and ROADIES extend these capabilities by integrating evolutionary context into sequence analysis (Deng et al., 2025; Gupta et al., 2025).



**Figure 1** - Cumulative growth of studies focused on RdRp over the years. The rate of new studies and associated SRA submissions appears to be plateauing, whereas the average size of individual sequencing runs is steadily increasing (Chikhi et al., 2025). Since the last RdRp summit in 2023, 475 extra studies on RdRp have been conducted. This figure was generated by doing a keyword search of the SRA for “RdRp”, therefore it may miss some metatranscriptome studies.

### A unified RNA virus data landscape

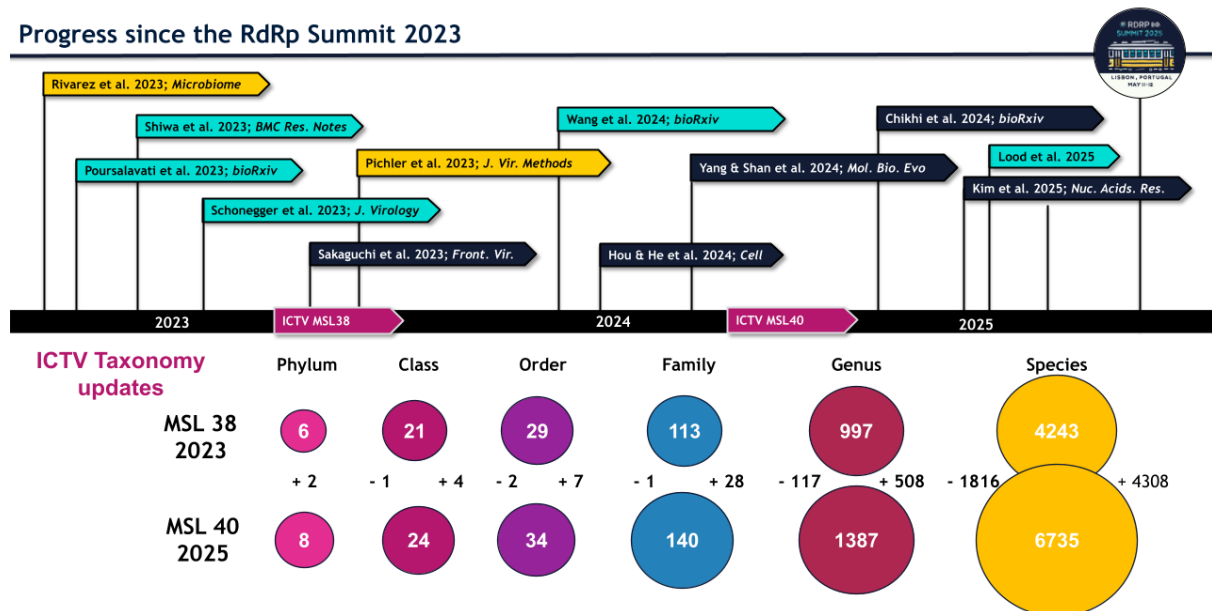
We define the “viral landscape” as the physical, genomic and ecological distribution of viruses in a natural environment, distinguishing it from the “RNA virus data landscape,” which refers to the digital repositories, databases, and bioinformatic frameworks used to store and analyze viral sequences. The analysis of samples using metagenomics methods often reveals diverse and largely unexplored RNA viral genomes in the environment. By advancing the discovery, annotation, and standardized sharing of these viral genomes, the RNA virus data landscape can be unified. Although efforts have been made in the past to standardize the reporting of viral genome data (MIUViG), these were mostly tailored to dsDNA bacteriophages. This framework incorporates RNA virus-specific reporting requirements, including genome strandedness, the presence of subgenomic RNAs, details of RNA extraction and library preparation methods (e.g., poly-A selection versus rRNA depletion), and approaches for recovering segmented genomes. Therefore, a growing need exists for an update to these standards to include the specific needs for RNA virus reporting, e.g. methods used for recovery of segmented genomes, along with user-friendly tools (e.g. SUTVK (Standardized Uncultivated Virus Taxonomy and Knowledge), <https://github.com/LanderDC/sutvk>) that ensures viral genome data are not only accurately represented but also FAIR-compliant. SUTVK is designed to be compatible with multiple major repositories (including NCBI, ENA, and DDBJ) to facilitate FAIR-compliant data sharing across the global bioinformatics infrastructure. Viral metagenomics is revealing the hidden diversity of viruses across diverse ecological gradients in both urban (e.g., global RNA viromes in cities) (Gao et al., 2024) and natural settings (e.g., rodent-associated viruses in Serbia, water-associated viromes in high-altitude lakes or bee-and pollen-associated viromes in Canadian tree fruit orchards) pointing towards their broader ecological and public health relevance (Vansia et al., 2024; Wu et al., 2025). An integrated vision for building a comprehensive and accessible global viral landscape exists, and is guided through the improvement of methods, tools, and collaborative standards.

### Expanding the RNA virus discovery toolkit

Advances in computational biology, sequencing technologies, and machine learning are transforming how RNA viruses are detected, classified, and studied. Traditional alignment-based methods have been supplemented by profile Hidden Markov Models (pHMMs) (<https://github.com/dimitris-karapliafis/RdRpCATCH>, <https://www.biorxiv.org/content/10.64898/2026.02.05.703936v1>), structure-aware homology detection, and, more recently, deep learning models that can be applied to identify viruses from genomic and metagenomic data, such as CHEER, VirHunter, Virtifier, and RNN-VirSeeker (Liu et

al., 2022; Miao et al., 2022; Shang & Sun, 2021; Sukhorukov et al., 2022). These tools employ convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Despite their versatility, both face limitations in processing biological sequences: CNNs may encounter challenges with inputs of varying lengths and capturing global correlations, while RNNs struggle with longer sequences due to vanishing or exploding gradients and difficulties in capturing long-term dependencies. To address this, Hou et al. employed a new AI model based on transformer architecture (i.e., LucaProt) for RNA virus discovery that utilizes both protein sequences and the structural characteristics of viral RdRp sequences, and shows its power for uncovering remote viral RdRp homologies and functional signatures, even in sequences with no detectable similarity to known references (Hou et al., 2024; Nakagawa & Sakaguchi, 2024).

These tools enable researchers to uncover RNA viruses that would have previously escaped detection due to their low sequence similarity to known taxa. At the same time, long-read and single-cell sequencing technologies are enhancing genome assembly and host linkage, especially in complex environmental samples. This growing arsenal of methods facilitates the discovery of the hidden RNA virosphere, annotation of functional viral elements, and inference of host–virus interactions, thereby enriching both taxonomic resolution and biological interpretation.



**Figure 2** - Advances in RNA virus biology and ICTV growth since 2023. (Top) Progress in RNA virus research following the inaugural RdRp Summit in 2023 includes long-read and single-cell sequencing, structure-based phylogenetics, large-scale protein modeling (AlphaFold, ESMfold, Foldseek, BFVD), and AI-driven discovery tools (LucaProt, CHEER, VirHunter), which together have revealed deeply divergent RNA virus lineages and illuminated the hidden virosphere. (Bottom) Each of the rings illustrates the rapid expansion of the ICTV taxonomy, including new phyla (*Ambiviricota* and *Artimaviricota*).

### Mining the planetary virosphere at scale and depth

Several global initiatives, such as VIRION (<https://www.viralemurgence.org/virion>), PREDICT (<https://ohi.vetmed.ucdavis.edu/programs-projects/predict-project>), and the NIH Human Virome Program (<https://commonfund.nih.gov/humanvirome>) have been instrumental in elucidating viromes across different environments, organisms, or ecosystems, each implementing distinct sampling and analytical strategies (Carlson et al., 2022; Wallace et al., 2025; Wang et al., 2025; Wu & Peng, 2024). Alongside these studies are the increasing number of research practices which introduce challenges in verifying the identity and taxonomy of uncultivated viruses. This

underscores the need for standardized methodologies that would ensure interoperability, reproducibility, and robust results that are both verifiable and reliable.

Leveraging large-scale computational and evolutionary approaches allows for the discovery of hidden dimensions of the planet's virosphere and enables the assessment of emerging viral threats at unprecedented scales (Chikhi et al., 2025; Edgar et al., 2022). For example, the Logan project (<https://logan-search.org>) provides a framework for the assembly and analysis of vast amounts of complex and fragmented genomic data. While recent breakthroughs have been driven largely by advances in computational methods, the standardized methodologies advocated by the consortium must extend across the entire pipeline, encompassing ecological sampling, RNA extraction, sequencing protocols, and downstream bioinformatic processing. This end-to-end standardization is essential to ensure that insights derived from large-scale analyses are comparable, reproducible, and biologically meaningful. The identification of endogenous viral elements embedded in public databases, using uncharacterized proteins as a window into ancient viral integrations, provides us with a deeper understanding of their evolutionary legacies in host genomes and their biological implications (Brown & Firth, 2024). The integration of bioinformatics, evolutionary biology, and big data illuminates the origins, trajectories, and zoonotic potential of RNA viruses on a global scale. Taken together, these tools provide insights into viral emergence, biological history, and future threats to human and animal health.

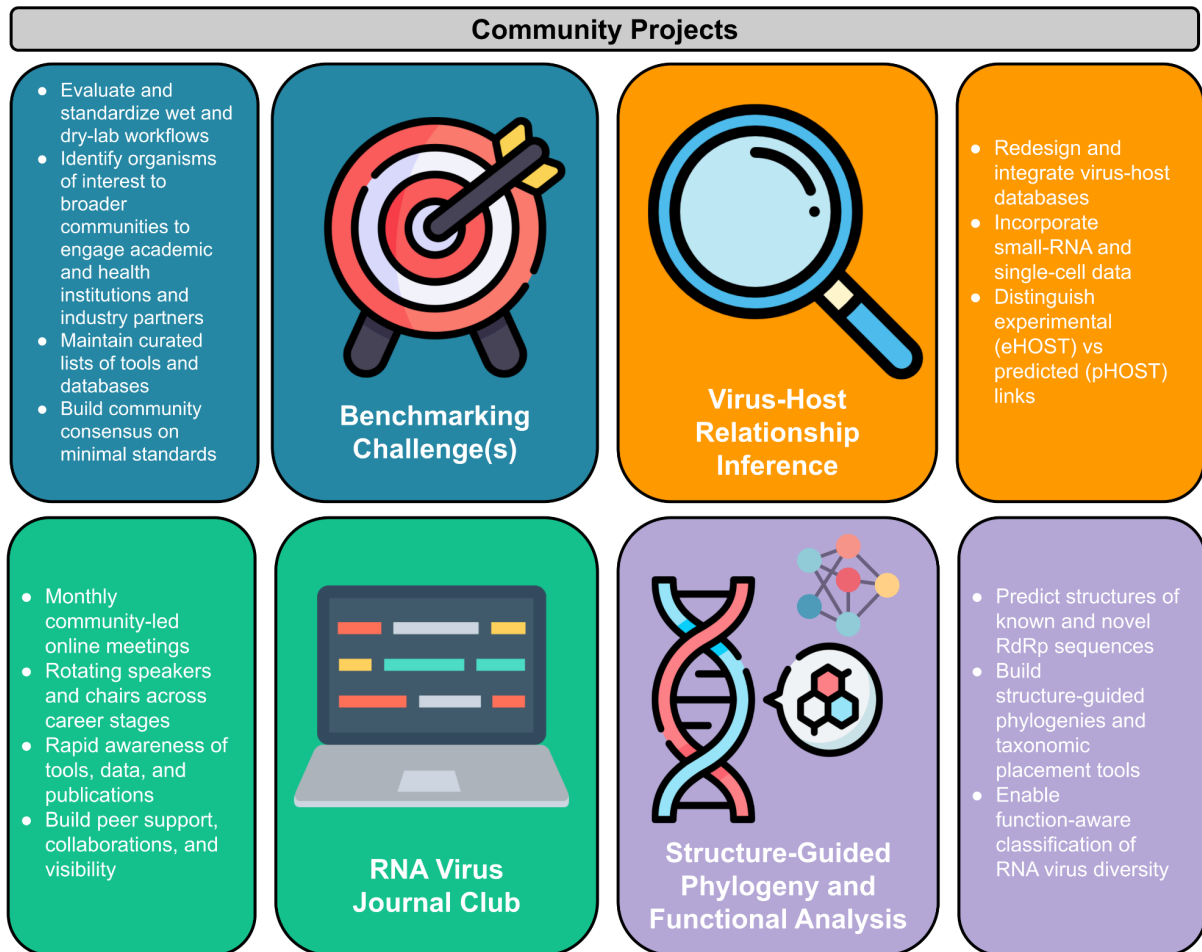
### **Illuminating the hidden RNA virosphere**

Approaches to uncovering the vast hidden diversity of RNA viruses, often referred to as “viral dark matter”, offer insights into one of the largest biological diversity reservoirs on the planet. By harnessing artificial intelligence (AI) to systematically document the hidden RNA virosphere, we have begun to understand and demonstrate how machine learning can be applied to detect and classify novel RNA viruses from increasingly complex datasets (Hou et al., 2024). Structural bioinformatics methods (BFVD-Foldseek) use protein-folding and alignment to identify deeply divergent viral sequences that elude current conventional sequence homology-based detection methods and further illuminate the structural and functional diversity (and their relationship) of viral dark matter (Kim et al., 2025). The extreme diversity presented by the marker gene RdRp can be examined using deep evolutionary analysis to demonstrate how this core viral enzyme varies significantly across lineages and diverse environments. The potentially transformative synergy between advances in AI methods, structural biology, and evolutionary sequence-based modeling is poised to fundamentally reshape our understanding of the breadth and depth of RNA viral diversity. The deliberate harmonization of these approaches constitutes a major strength of the field and has the potential to exert a lasting impact on future discoveries. Achieving this harmonization will require coordinated community efforts, including the establishment of benchmarking datasets for method evaluation, the adoption of standardized metadata reporting frameworks and the integration of structure-informed annotations and predictions into formal ICTV taxonomic workflows. Harmonization will be achieved through the establishment of community-agreed benchmarking datasets, the adoption of standardized metadata reporting, and the integration of structural predictions into official ICTV taxonomic pipelines. Together, these steps will enable more robust, reproducible, and scalable frameworks for RNA virus discovery and classification.

### **Towards community-driven solutions and future initiatives**

The RdRp Summit 2025 brought together researchers working on current challenges of RNA virus research, but also in actively addressing them through community-driven collaboration. As the field of RNA biology continues to expand at an unprecedented pace and is spurred on by advances in short- and long-read sequencing technologies, structural prediction, AI, and metagenomics - it has become increasingly important to align efforts around shared methodologies, standards, and infrastructures (Zielezinski et al., 2025).

A key goal of the summit is to foster a community that remains active between the biannual meetings. To support this objective, the summit serves as a hub to connect researchers facing similar challenges, enabling them to work collectively toward shared solutions. To this end, participants were invited to propose ideas for community-driven initiatives. A dedicated session was held to form collaborative groups around selected topics of mutual interest. In this report, we present the community projects that will be supported and developed by the RdRp Summit over the next two years (Figure 3). We propose three prioritized and interdependent initiatives: (1) the establishment of foundational Benchmarking Challenges to rigorously evaluate and standardize computational tools, (2) the application of these validated tools to Virus–Host Relationship Inference, and (3) the integration of these insights into Structure-Guided Phylogeny.



**Figure 3** - Community-driven initiatives launched at the second RdRp Summit. Four community projects were proposed to address critical gaps in RNA virus research identified by participants of the RdRp Summit 2025. (1) Benchmarking Challenges aim to develop gold-standard datasets, metrics, and curated tool lists to standardize both experimental and computational workflows. (2) Virus-Host Relationship Inference focuses on improving host assignment through integrated databases, multi-layered prediction approaches, and hackathon-driven curation. (3) RNA Virus Journal Club provides a platform for regular knowledge exchange, collaborative discussions, and community building. (4) Structure-Guided Phylogeny and Functional Analysis seeks to classify deeply divergent RNA viruses by integrating structural prediction, comparative analyses, and phylogenetic frameworks. Together, these initiatives establish the foundation for coordinated, sustained progress in RNA virus discovery.

## Community projects

To meet this need, the summit aims to establish collaborative initiatives that have been proposed and launched by the community. For example, as part of the 1<sup>st</sup> RdRp Summit, the community project of RdRpCATCH -RdRp Collaborative Analysis Tools with Collections of pHMMs- was introduced (<https://github.com/dimitris-karapliafis/RdRpCATCH>). Notably, the first RdRp Summit demonstrated that coordinated community action can substantially accelerate tool validation and comparative assessment, providing a scalable framework for benchmarking emerging methodologies across diverse datasets. This project consolidated multiple RdRp pHMM databases developed over the past five years into a single resource, addressing fragmentation in the field and reducing the technical barriers to the discovery of RNA viruses. The 2<sup>nd</sup> RdRp Summit led to the launch of multiple community projects, each designed to tackle key gaps in RNA virus discovery.

### *Benchmarking Challenge(s)*

The “Benchmarking Challenge” represents a coordinated community effort to establish standardized frameworks for the evaluation of RNA virus discovery and assembly tools. Central to this initiative is the development of shared benchmarking datasets spanning synthetic constructs, mock communities, and real-world multi-virome samples, enabling consistent and reproducible performance assessment across methods. By defining common evaluation metrics and testing conditions, this effort seeks to ensure comparability between tools, reduce fragmentation in methodological development, and prevent the proliferation of unvalidated approaches. Ultimately, the Benchmarking Challenge provides a foundation for rigorous, transparent, and scalable tool validation, supporting more reliable discovery and characterization of RNA viral diversity. The benchmarking challenges working group highlighted several core issues that are considered important. These include securing sustainable funding, defining the scope of the work, including which viruses to prioritize, and ensuring long-term oversight, since meaningful benchmarking is likely to be a slow and iterative process. In order to provide gold standard data sets, the objectives of this working group include identifying what can and should be benchmarked across both wet-lab experimental protocols and dry-lab computational analysis, including the generation of synthetic datasets, and determining how these benchmarks can be effectively implemented and compared in the field. To strengthen our approach, the group also emphasized learning from related initiatives to generate gold standard benchmarking data sets, such as existing certification and standards projects (i.e. CAMDA, <https://bipress.boku.ac.at/camda2025>) and CASP (Critical Assessment of Structure Prediction, <https://predictioncenter.org/index.cgi>), which could provide useful models and prevent duplication of effort. Additionally, the group will maintain a curated list of software, tools, databases, and resources for RNA virus analysis, prediction, annotation, phylogenetics, and related research, which is available at: <https://github.com/rdrp-summit/awesome-rna-virus-tools>.

### *Virus-Host Relationship Inference*

We suggest that databases being redesigned for virus–host inference should prioritize the integration of host–virus association metadata with genomic features (e.g., k-mer profiles, GC content), rather than duplicating the storage of raw sequencing data, which remains the responsibility of primary repositories such as the Sequence Read Archive. Participants identified key priorities including the redesign of database curation and integration, where current resources were deemed inadequate. Current collaborative efforts include building more accurate web-curated databases supported by both manual and semi-automated approaches. Databases will be extended by employing multiple dataset-specific statistics (e.g., k-mer analysis, GC-content, nucleotide frequencies), and expanding to underutilized data types such as small RNA sequencing and single-cell RNA data. We also emphasize distinguishing experimental “eHOSTs” from predicted host assignments “pHOSTs”, tracking co-occurrence in raw data and integrating evolutionary features like mutation and recombination rates. It is also important to distinguish

between hosts as the organism(s) within whose cells a virus replicates, and the organism from which virus might be collected from, as this can lead to spurious virus-host relationships without evidence that the virus successfully replicates within a particular organism.

The organization and importance of hackathons, community-driven curation, and benchmarking efforts will be central in order to make progress on the massive manual curation task. Broader considerations include the role of host taxonomic rank in classification and its importance, and the integration of viral-like elements and host-protein interactions, as well as the need for standardization of approaches. The group is currently working and will produce a comprehensive review of AI-based virus-host prediction tools, highlighting the necessity of multi-layered computational prediction methods and statistical frameworks, and community coordination.

#### *RNA Virus Journal Club*

To facilitate communication among RNA virus researchers, the RNA Virus Journal Club was established. Organized by the second RdRp Summit participants, the Journal Club serves as a platform to share the latest updates in the field of RNA virus research. Importantly, monthly meetings have rotating chairs and times to promote inclusivity. Within online meetings, the community could exchange research insights and challenges, get help and advice from peers, and find potential collaborators. Each session includes a presentation from one main speaker, followed by open discussion. Community members of all career stages are encouraged to nominate speakers and chairs and participate actively, ensuring the meetings reflect the needs and interests of the community. Short summaries of the meetings are available afterwards, also promoting offline discussion. The participation and speaker nomination forms are available at the RdRp summit website (<https://rdrp.io>). The Journal Club started in September 2025, and we hope this will be a useful tool for the community to keep active networking. Since its inception in September 2025, the journal club has maintained strong engagement, with an average attendance of 40-60 researchers per session, representing a diverse, global cross-section of the community.

#### *Structure-Guided Phylogeny*

The immense diversity of viral proteins results in a scarcity of sequence homologs in existing databases, posing major challenges for protein comparison and annotation (Kuchibhatla et al., 2014; Terzian et al., 2021). However, advances in highly accurate and scalable tools for protein structure prediction AlphaFold, ESMfold, ProST5 (Heinzinger et al., 2024; Jumper et al., 2021; Lin et al., 2023) and comparison (Foldseek) (van Kempen et al., 2024), combined with improved understanding of the statistical properties of unrelated structural folds, now enable the detection of homologs with near undetectable sequence similarity.

By comprehensively predicting the structures of the known RdRp diversity we can expand our search space for detecting novel distant homologs; not only relying on amino acid similarity or motif conservation but also utilising the deeper conservation present on the structural level of the proteins. The Big Fantastic Virus Database (BFVD, <https://bfvd.foldseek.com/>) (Kim et al., 2025) has already provided a great starting point for enabling structure-based homology searches against virus protein structures. However, there is a growing number of distant RdRp sequences whose structure has not been predicted yet. This unexplored diversity is exemplified by the findings of the LucaProt approach, describing 180 RNA virus supergroups based on deep RdRp homology, 23 of which were previously unknown and only 21 of which have been classified taxonomically (Hou et al., 2024).

The already known RdRp sequence diversity can largely help with confidently predicting their corresponding structures, since many of the best performing prediction methods rely on comprehensive multiple sequence alignments (MSA). Combining the LucaProt RdRp dataset with other expansive sequence databases (e.g., Logan) may improve MSA-based predictions. For cases where close known homologs remain sparse, protein Language Models (pLM)-based predictions can produce improved structures, having been shown to perform better than MSA-based methods when alignments are shallow (Mifsud et al., 2024).

A database of RdRp structures (both predicted and experimentally determined) can aid in the quest to uncover more of the hidden RNA virosphere but also perform structure-guided phylogenetics of the RNA virosphere. Given the exponential increase in available (predicted) protein structures, new methods have been developed for structure-based sequence alignment (Gilchrist et al., 2024) and phylogenetics (Puente-Lelievre et al., 2024). We believe that these approaches can provide higher resolution in the more ancient parts of the RdRp evolutionary history that remain unresolved; for example, delineating the relatedness of ‘orphan’ taxonomic groups. We envision the creation of a comprehensive RdRp structure database, datasets of structure-guided phylogenies for all uncovered RdRp diversity, and a tool for placing novel RdRp candidates into this diversity based on both sequence and structural homology.

## Future initiatives

### *Quantifying evolutionary novelty*

The discovery of novel RdRps at scale has been facilitated by the growing availability of diverse and often underexplored sequencing datasets, including metatranscriptomes, viromes (enriched or rRNA-depleted), dsRNA sequencing approaches, and specialized protocols such as fragmented and loop primer ligated dsRNA sequencing (FLDS), many of which are publicly accessible through repositories such as the Sequence Read Archive. Within this landscape, total RNA sequencing and poly(A)-selected sequencing represent complementary methodological strategies: total RNA approaches enable the capture of segmented and non-polyadenylated viruses, while poly(A) selection enriches for eukaryotic host transcripts and specific viral clades. Coupled with increasingly sensitive homology detection methods, these advances have substantially expanded our ability to detect and characterize RNA viruses across diverse environments. Many (if not most) RNA virus discovery studies typically report numbers of newly discovered RdRp operational taxonomic units, OTUs (groups of homologous sequences clustering at an arbitrary identity threshold), or similar terms that communicate the scale but not depth of discovery. For example, a now classic study by (Li et al., 2015) reported the discovery of 112 novel negative sense single-stranded RNA viruses in arthropods but at the time did not have the quantitative means (e.g., branch length-based measures across large datasets). To communicate that a considerable number of these viruses comprised what is a very distinct family-level group now referred to as *chuviruses*.

At the RdRp Summit in Lisbon in May 2025, a metric was proposed that could help turn metatranscriptomic RNA virus discovery studies more quantitative by adapting what is called “phylogenetic diversity” in ecology. Branch lengths in phylogenetic trees typically represent independent amounts of character evolution (e.g. amino acids) and the sum of branch lengths added to trees by new sequences can be thought of as its evolutionary novelty. When sequences detected in a sample are analysed in this way, statements like “we detected 2 viruses (1 novel and 1 previously described)” can be refined further into “we detected 2 viruses contributing a total of 0.64 aa/subs/site and 0.0029 aa/s/s”, imparting to the reader that one sequence is substantially novel while the other is very closely related to known ones. While phylogenetic diversity metrics provide a powerful framework for quantifying viral discovery, their application is currently constrained by alignment quality and phylogenetic depth. Highly divergent viruses that cannot be reliably aligned remain difficult to incorporate into such analyses. Future iterations of this approach may therefore benefit from integrating alignment-free measures of evolutionary distance, including structure-based comparisons derived from tools such as Foldseek.

RNA viruses lend themselves naturally to such studies on account of the RdRp being a conserved gene shared by all RNA viruses. Furthermore, phylogenetic diversity can be quantified on phylogenetic trees encompassing theoretically, as few as two most closely related sequences, minimising phylogenetic uncertainty arising from poor alignment quality. One implementation of phylogenetic diversity quantification was used on orthomyxoviruses in (Batson et al., 2021), showcasing how phylogenetic diversity can: i) track the success of discovery methods to date, ii) visually track where novelty is found across a phylogenetic tree (using a method adapted from

(Obbard, 2018) forecast saturation (or not) of novelty, and iv) identify virus sub-groups contributing most novelty. Going forward, the quantitative tracking of RNA virus discovery efforts could be done at the initiative of study authors themselves or by other groups on publicly available data, with additional considerations such as focusing on novelty contributed by specific virus sub-groups (e.g. influenza vs quaranjaviruses) or host groups (e.g. arthropods vs vertebrates).

#### *Recovering complete segmented genomes*

For many unsegmented RNA viruses, modern metatranscriptomic methods suffice: a long open reading frame found on the same contig as known viral genes can usually be accepted as viral. In contrast, segmented RNA viruses - especially those with numerous short segments such as reoviruses or orthomyxoviruses - pose greater challenges. Their smaller segments often lose detectable amino acid homology faster than others (e.g. RdRp-encoding segments), and without physical linkage to known viral genes, they can vanish into metagenomic “dark matter,” leaving genomes incomplete.

Previously, small interfering RNA (siRNA) sequencing in arthropods helped address this problem by identifying dsRNA-derived sequences, enabling co-occurrence-based reconstruction of segmented genomes (Webster et al., 2015). This approach led to the discovery of Galbūt virus, a common *Drosophila melanogaster* partitivirus. Co-occurrence-based reconstruction has also proven effective in metatranscriptomic studies of individual hosts, revealing, for instance, that four mosquito-associated quaranjaviruses (*Orthomyxoviridae*) possess eight-segment genomes, with the smallest three lacking any recognisable homology (Batson et al., 2021). Falling sequencing costs and tools such as SegFinder (Liu et al., 2025) are likely to make complete genome reconstruction more routine. Ideally, though, researchers should take the initiative to explore and report RNA virus diversity in full (i.e. genome-resolved characterization of RNA viruses that includes complete genome reconstruction, linkage of all segments in multipartite viruses, and comprehensive analysis of sequence, structure, function, and evolutionary relationships beyond single-marker genes such as RdRp).

## Conclusion

By building on progress since the inaugural meeting in 2023, participants demonstrated breakthroughs in sequencing, structural modeling, and advances in AI-driven methods for characterization, which are quickly expanding viral taxonomy and illuminating the vast hidden dark matter of the RNA virus world. The summit has identified a number of pressing challenges in our field: this includes the need for standardized annotation, data sharing and advancing structure-guided phylogenetics. The summit also launched a number of community-driven initiatives: such as benchmarking challenges, virus-host inference projects and an RNA virus journal club, which can be attended by anyone. Through coordinating efforts across multiple disciplines, the summit has laid the groundwork for a shared research initiative that will accelerate the discovery, improve the classification of, and provide deeper evolutionary insights into the global RNA viral ecosystem. Join our initiative through <https://rdrp.io/>, which includes contact information and invitations to our communication channels.

## Acknowledgements

The authors gratefully acknowledge the contributions of all participants whose collaboration, presentations, and shared expertise made this consensus statement possible. Their collective efforts and commitment to the RNA virus field were essential. We acknowledge Valentyn Bezshapkin (<https://orcid.org/0000-0002-0912-4371>) for their insightful feedback and assistance. Preprint version 5 of this article has been peer-reviewed and recommended by Peer Community In Microbiology, (<https://doi.org/10.24072/pci.microbiol.100236>; Massart, 2026).

## Data, scripts, code, and supplementary information availability

Not applicable.

## Conflicts of interest disclosure

The authors declare they comply with the PCI rule of having no financial conflicts of interest. The 2025 RdRp summit was supported by the Peer Community In (PCI) open-science initiative

## Funding

The 2025 RdRp summit was supported by the Peer Community In (PCI) open-science initiative (<https://peercommunityin.org>) and the International Society for Microbial Ecology (ISME). RR was supported by funding provided by the US National Science Foundation (NSF DBI 2515340) to the Viral Emergence Research Institute (<https://www.viralemurgence.org>). GD is supported by EMBO installation grant EMBO-IG-5305-2023. TD was supported by grants from the Research Council of Finland (330977) and the Kone Foundation. LDC was supported by the Research Foundation Flanders (11L1325N).

## References

- Batson, J., Dudas, G., Haas-Stapleton, E., Kistler, A. L., Li, L. M., Logan, P., Ratnasiri, K., & Retallack, H. (2021). Single mosquito metatranscriptomics identifies vectors, emerging pathogens and reservoirs in one assay. *eLife*, 10, e68353. <https://doi.org/10.7554/eLife.68353>
- Brown, K., & Firth, A. E. (2024). Uncovering hundreds of exogenous and endogenous RNA viral RdRp sequences amongst uncharacterised sequences in public protein databases. *bioRxiv*, 2024.09.25.614983. <https://doi.org/10.1101/2024.09.25.614983>
- Bruenn, J. A. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Research*, 31(7), 1821–1829. <https://doi.org/10.1093/nar/gkg277>
- Carlson, C. J., Gibb, R. J., Albery, G. F., Brierley, L., Connor, R. P., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Frank, H. K., Muylaert, R. L., Poisot, T., Rasmussen, A. L., Ryan, S. J., & Seifert, S. N. (2022). The Global Virome in One Network (VIRION): An Atlas of Vertebrate-Virus Associations. *mBio*, 13(2), e0298521. <https://doi.org/10.1128/mbio.02985-21>
- Charon, J., Buchmann, J. P., Sadiq, S., & Holmes, E. C. (2022). RdRp-scan: A bioinformatic resource to identify and annotate divergent RNA viruses in metagenomic sequence data. *Virus Evolution*, 8(2), veac082. <https://doi.org/10.1093/ve/veac082>
- Charon, J., Olendraite, I., Forgia, M., Chong, L. C., Hillary, L. S., Roux, S., Kupczok, A., Debat, H., Sakaguchi, S., Tahzima, R., Nakagawa, S., Babaian, A., Abroi, A., Bejerman, N., Ben Mansour, K., Brown, K., Butkovic, A., Cervera, A., Charriat, F., ... Neri, U. (2024). Consensus statement from the first RdRp Summit: Advancing RNA virus discovery at scale across communities. *Frontiers in Virology*, 4, 1371958. <https://doi.org/10.3389/fviro.2024.1371958>
- Chikhi, R., Lemane, T., Loll-Krippleber, R., Montoliu-Nerin, M., Raffestin, B., Camargo, A. P., Miller, C. J., Fiamenghi, M. B., Agostinho, D. P., Majidian, S., Autric, G., Hugues, M., Lee, J., Faure, R., Curry, K. D., Sousa, J. A. M. de, Rocha, E. P. C., Koslicki, D., Medvedev, P., ... Babaian, A. (2025). Logan: Planetary-Scale Genome Assembly Surveys Life's Diversity. *bioRxiv*, 2024.07.30.605881. <https://doi.org/10.1101/2024.07.30.605881>
- Deng, D., Xu, W., Wu, B., Comes, H. P., Feng, Y., Li, P., Zheng, J., Chen, G., & Heng, P.-A. (2025). PhyloTune: An efficient method to accelerate phylogenetic updates using a pretrained DNA language model. *Nature Communications*, 16(1), 6905. <https://doi.org/10.1038/s41467-025-61684-3>
- Edgar, R. C., Taylor, B., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M., Korobeynikov, A., Chikhi, R., &

- Babaian, A. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature*, 602(7895), 142–147. <https://doi.org/10.1038/s41586-021-04332-2>
- Forgia, M., Navarro, B., Daghino, S., Cervera, A., Gisel, A., Perotto, S., Aghayeva, D. N., Akinyuwa, M. F., Gobbi, E., Zheludev, I. N., Edgar, R. C., Chikhi, R., Turina, M., Babaian, A., Di Serio, F., & de la Peña, M. (2023). Hybrids of RNA viruses and viroid-like elements replicate in fungi. *Nature Communications*, 14(1), 2591. <https://doi.org/10.1038/s41467-023-38301-2>
- Gao, Z., Wu, J., Lucaci, A. G., Ouyang, J., Wang, L., Ryon, K., Elhaik, E., Probst, A. J., Rodó, X., Velavan, T., Chasapi, A., Ouzounis, C. A., Oliveira, M., Dias-Neto, E., Osuolale, O. O., Poulsen, M., Meleshko, D., Bhattacharyya, M., Ugalde, J. A., ... Shi, T. (2024). Diversity and Distinctive Traits of the Global RNA Virome in Urban Environments (SSRN Scholarly Paper No. 4871972). Social Science Research Network. <https://doi.org/10.2139/ssrn.4871972>
- Gibbs, A. (2020). Binomial nomenclature for virus species: A long view. *Archives of Virology*, 165(12), 3079–3083. <https://doi.org/10.1007/s00705-020-04828-8>
- Gilchrist, C. L. M., Mirdita, M., & Steinegger, M. (2024). Multiple Protein Structure Alignment at Scale with FoldMason bioRxiv, 2024.08.01.606130. <https://doi.org/10.1101/2024.08.01.606130>
- Gupta, A., Mirarab, S., & Turakhia, Y. (2025). Accurate, scalable, and fully automated inference of species trees from raw genome assemblies using ROADIES. *Proceedings of the National Academy of Sciences*, 122(19), e2500553122. <https://doi.org/10.1073/pnas.2500553122>
- Heinzinger, M., Weissenow, K., Sanchez, J. G., Henkel, A., Mirdita, M., Steinegger, M., & Rost, B. (2024). Bilingual language model for protein sequence and structure. *NAR Genomics and Bioinformatics*, 6(4), lqae150. <https://doi.org/10.1093/nargab/lqae150>
- Holmes, E. C., & Duchêne, S. (2019). Can Sequence Phylogenies Safely Infer the Origin of the Global Virome? *mBio*, 10(2), e00289-19. <https://doi.org/10.1128/mBio.00289-19>
- Hou, X., He, Y., Fang, P., Mei, S.-Q., Xu, Z., Wu, W.-C., Tian, J.-H., Zhang, S., Zeng, Z.-Y., Gou, Q.-Y., Xin, G.-Y., Le, S.-J., Xia, Y.-Y., Zhou, Y.-L., Hui, F.-M., Pan, Y.-F., Eden, J.-S., Yang, Z.-H., Han, C., ... Shi, M. (2024). Using artificial intelligence to document the hidden RNA virosphere. *Cell*, 187(24), 6929-6942.e16. <https://doi.org/10.1016/j.cell.2024.09.027>
- International Committee on Taxonomy of Viruses Executive Committee. (2020). The new scope of virus taxonomy: Partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology*, 5(5), 668–674. <https://doi.org/10.1038/s41564-020-0709-x>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kim, R. S., Levy Karin, E., Mirdita, M., Chikhi, R., & Steinegger, M. (2025). BFVD—a large repository of predicted viral protein structures. *Nucleic Acids Research*, 53(D1), D340–D347. <https://doi.org/10.1093/nar/gkae1119>
- Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*, 479–480, 2–25. <https://doi.org/10.1016/j.virol.2015.02.039>
- Kuchibhatla, D. B., Sherman, W. A., Chung, B. Y. W., Cook, S., Schneider, G., Eisenhaber, B., & Karlin, D. G. (2014). Powerful Sequence Similarity Search Methods and In-Depth Manual Analyses Can Identify Remote Homologs in Many Apparently “Orphan” Viral Proteins. *Journal of Virology*, 88(1), 10–20. <https://doi.org/10.1128/jvi.02595-13>
- Kuhn, J. H., Botella, L., de la Peña, M., Vainio, E. J., Krupovic, M., Lee, B. D., Navarro, B., Sabanadzovic, S., Simmonds, P., & Turina, M. (2024). Ambiviricota, a novel ribovirian phylum for viruses with viroid-like properties. *Journal of Virology*, 98(7), e00831-24. <https://doi.org/10.1128/jvi.00831-24>
- Li, C.-X., Shi, M., Tian, J.-H., Lin, X.-D., Kang, Y.-J., Chen, L.-J., Qin, X.-C., Xu, J., Holmes, E. C., & Zhang, Y.-Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife*, 4, e05378. <https://doi.org/10.7554/eLife.05378>

- Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), 1123–1130. <https://doi.org/10.1126/science.ade2574>
- Litvin, U., Lytras, S., Jack, A., Robertson, D. L., Hughes, J., & Grove, J. (2025). Viro3D: A comprehensive database of virus protein structure predictions. *Molecular Systems Biology*, 21(11), 1599–1617. <https://doi.org/10.1038/s44320-025-00147-9>
- Liu, F., Miao, Y., Liu, Y., & Hou, T. (2022). RNN-VirSeeker: A Deep Learning Method for Identification of Short Viral Sequences From Metagenomes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3), 1840–1849. <https://doi.org/10.1109/TCBB.2020.3044575>
- Liu, X., Kong, J., Shan, Y., Yang, Z., Miao, J., Pan, Y., Luo, T., Shi, Z., Wang, Y., Gou, Q., Yang, C., Li, H., Li, C., Li, S., Zhang, X., Sun, Y., Holmes, E. C., Guo, D., & Shi, M. (2025). SegFinder: An automated tool for identifying complete RNA virus genome segments through co-occurrence in multiple sequenced samples. *Briefings in Bioinformatics*, 26(4), bbaf358. <https://doi.org/10.1093/bib/bbaf358>
- Massart, S. (2026). A growing community to navigate the deluge of sequencing data and provide a RdRp-based framework for data-oriented RNA virus biology. *Peer Community in Microbiology*, 100236. <https://doi.org/10.24072/pci.microbiol.100236>
- Miao, Y., Liu, F., Hou, T., & Liu, Y. (2022). Virtifier: A deep learning-based identifier for viral sequences from metagenomes. *Bioinformatics (Oxford, England)*, 38(5), 1216–1222. <https://doi.org/10.1093/bioinformatics/btab845>
- Mifsud, J. C. O., Lytras, S., Oliver, M. R., Toon, K., Costa, V. A., Holmes, E. C., & Grove, J. (2024). Mapping glycoprotein structure reveals Flaviviridae evolutionary history. *Nature*, 633(8030), 695–703. <https://doi.org/10.1038/s41586-024-07899-8>
- Nakagawa, S., & Sakaguchi, S. (2024). Exploring the hidden world of RNA viruses with a transformer-based tool. *Patterns (New York, N.Y.)*, 5(11), 101095. <https://doi.org/10.1016/j.patter.2024.101095>
- Neri, U., Wolf, Y. I., Roux, S., Camargo, A. P., Lee, B., Kazlauskas, D., Chen, I. M., Ivanova, N., Zeigler Allen, L., Paez-Espino, D., Bryant, D. A., Bhaya, D., RNA Virus Discovery Consortium, Krupovic, M., Dolja, V. V., Kyrpides, N. C., Koonin, E. V., & Gophna, U. (2022). Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell*, 185(21), 4023–4037.e18. <https://doi.org/10.1016/j.cell.2022.08.023>
- Obbard, D. J. (2018). Expansion of the metazoan virosphere: Progress, pitfalls, and prospects. *Current Opinion in Virology, Virus Structure and Expression • Viral Evolution*, 31, 17–23. <https://doi.org/10.1016/j.coviro.2018.08.008>
- Pichler, I., Schmutz, S., Ziltener, G., Zaheri, M., Kufner, V., Trkola, A., & Huber, M. (2023). Rapid and sensitive single-sample viral metagenomics using Nanopore Flongle sequencing. *Journal of Virological Methods*, 320, 114784. <https://doi.org/10.1016/j.jviromet.2023.114784>
- Puente-Lelievre, C., Malik, A. J., Douglas, J., Ascher, D., Baker, M., Allison, J., Poole, A., Lundin, D., Fullmer, M., Bouckert, R., Kim, H., Steinegger, M., & Matzke, N. (2024). Tertiary-interaction characters enable fast, model-based structural phylogenetics beyond the twilight zone. *bioRxiv*, 2023.12.12.571181. <https://doi.org/10.1101/2023.12.12.571181>
- Schönegger, D., Moubset, O., Margaria, P., Menzel, W., Winter, S., Roumagnac, P., Marais, A., & Candresse, T. (2023). Benchmarking of virome metagenomic analysis approaches using a large, 60+ members, viral synthetic community. *Journal of Virology*, 97(11), e01300-23. <https://doi.org/10.1128/jvi.01300-23>
- Shang, J., & Sun, Y. (2021). CHEER: Hierarchical taxonomic classification for viral metagenomic data via deep learning. *Methods (San Diego, Calif.)*, 189, 95–103. <https://doi.org/10.1016/j.ymeth.2020.05.018>
- Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M. Q., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., ... Zerbini, F. M.

- (2017). Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews. Microbiology*, 15(3), 161–168. <https://doi.org/10.1038/nrmicro.2016.177>
- Sukhorukov, G., Khalili, M., Gascuel, O., Candresse, T., Marais-Colombel, A., & Nikolski, M. (2022). VirHunter: A Deep Learning-Based Method for Detection of Novel RNA Viruses in Plant Sequencing Data. *Frontiers in Bioinformatics*, 2, 867111. <https://doi.org/10.3389/fbinf.2022.867111>
- Tahzima, R., Charon, J., Diaz, A., De Jonghe, K., Massart, S., Michon, T., & Vranken, W. (2025). Viral replication modulated by hallmark conformational ensembles: How AlphaFold-predicted features of RdRp folding dynamics combined with intrinsic disorder-mediated function enable RNA virus discovery. *Frontiers in Virology*, 5. <https://doi.org/10.3389/fviro.2025.1501616>
- Tang, X., Shang, J., & Sun, Y. (2022). RdRp-based sensitive taxonomic classification of RNA viruses for metagenomic data. *Briefings in Bioinformatics*, 23(2), bbac011. <https://doi.org/10.1093/bib/bbac011>
- Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R. E., Mom, R., Toussaint, A., Petit, M.-A., & Enault, F. (2021). PHROG: Families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics*, 3(3), lqab067. <https://doi.org/10.1093/nargab/lqab067>
- Urayama, S., Fukudome, A., Hirai, M., Okumura, T., Nishimura, Y., Takaki, Y., Kurosawa, N., Koonin, E. V., Krupovic, M., & Nunoura, T. (2024). Double-stranded RNA sequencing reveals distinct riboviruses associated with thermoacidophilic bacteria from hot springs in Japan. *Nature Microbiology*, 9(2), 514–523. <https://doi.org/10.1038/s41564-023-01579-5>
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 42(2), 243–246. <https://doi.org/10.1038/s41587-023-01773-0>
- Vansia, R., Smadi, M., Phelan, J., Wang, A., Bilodeau, G. J., Pernal, S. F., Guarna, M. M., Rott, M., & Griffiths, J. S. (2024). Viral Diversity in Mixed Tree Fruit Production Systems Determined through Bee-Mediated Pollen Collection. *Viruses*, 16(10), 1614. <https://doi.org/10.3390/v16101614>
- Wallace, M. A., Wille, M., Geoghegan, J., Imrie, R. M., Holmes, E. C., Harrison, X. A., & Longdon, B. (2025). Making sense of the virome in light of evolution and ecology. *Proceedings of the Royal Society B: Biological Sciences*, 292(2044), 20250389. <https://doi.org/10.1098/rspb.2025.0389>
- Wang, C., Zheng, R., & Sun, C. (2025). Characterization of deep-sea viruses reveals their unexpected diversity and role in facilitating host metabolism of complex organic matter. *bioRxiv*, 2025.03.21.644689. <https://doi.org/10.1101/2025.03.21.644689>
- Webster, C. L., Waldron, F. M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J. F., Brouqui, J.-M., Bayne, E. H., Longdon, B., Buck, A. H., Lazzaro, B. P., Akorli, J., Haddrill, P. R., & Obbard, D. J. (2015). The Discovery, Distribution, and Evolution of Viruses Associated with *Drosophila melanogaster*. *PLOS Biology*, 13(7), e1002210. <https://doi.org/10.1371/journal.pbio.1002210>
- Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., & Koonin, E. V. (2018). Origins and Evolution of the Global RNA Virome. *mBio*, 9(6), 10.1128/mbio.02329-18. <https://doi.org/10.1128/mbio.02329-18>
- Wolf, Y. I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., Krupovic, M., Fire, A., Dolja, V. V., & Koonin, E. V. (2020). Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nature Microbiology*, 5(10), 1262–1270. <https://doi.org/10.1038/s41564-020-0755-4>
- Wu, L., Liu, Y., Shi, W., Chang, T., Liu, P., Liu, K., He, Y., Li, Z., Shi, M., Jiao, N., Lang, A. S., Dong, X., & Zheng, Q. (2025). Uncovering the hidden RNA virus diversity in Lake Nam Co: Evolutionary insights from an extreme high-altitude environment. *Proceedings of the National Academy of Sciences*, 122(6), e2420162122. <https://doi.org/10.1073/pnas.2420162122>
- Wu, Y., & Peng, Y. (2024). Ten computational challenges in human virome studies. *Virologica Sinica*, 39(6), 845–850. <https://doi.org/10.1016/j.virs.2024.04.008>

- Yin, Y., & Fischer, D. (2006). On the origin of microbial ORFans: Quantifying the strength of the evidence for viral lateral transfer. *BMC Evolutionary Biology*, 6(1), 63. <https://doi.org/10.1186/1471-2148-6-63>
- Zielezinski, A., Gudyś, A., Barylski, J., Siminski, K., Rozwalak, P., Dutilh, B. E., & Deorowicz, S. (2025). Ultrafast and accurate sequence alignment and clustering of viral genomes. *Nature Methods*, 22(6), 1191–1194. <https://doi.org/10.1038/s41592-025-02701-7>