



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The International Virus Bioinformatics Meeting 2023

Citation for published version:

Hufsky, F, Abecasis, AB, Babaian, A, Beck, S, Brierley, L, Dellicour, S, Eggeling, C, Elena, SF, Gieraths, U, Ha, AD, Harvey, W, Jones, TC, Lamkiewicz, K, Lovate, GL, Lücking, D, Machyna, M, Nishimura, L, Nocke, MK, Renard, BY, Sakaguchi, S, Sakellaridi, L, Spangenberg, J, Tarradas-Aleman, M, Triebel, S, Vakulenko, Y, Wijesekara, RY, González-Candelas, F, Krautwurst, S, Pérez-Cataluña, A, Randazzo, W, Sánchez, G & Marz, M 2023, 'The International Virus Bioinformatics Meeting 2023', The International Virus Bioinformatics Meeting 2023, Valencia, Spain, 24/05/23 - 26/05/23 pp. 1-27.
<https://doi.org/10.3390/v15102031>

Digital Object Identifier (DOI):

[10.3390/v15102031](https://doi.org/10.3390/v15102031)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

General rights































Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



The International Virus Bioinformatics Meeting 2023

Franziska Hufsky ^{1,2} , Ana B. Abecasis ^{1,3} , Artem Babaian ^{1,4,5} , Sebastian Beck ⁶ , Liam Brierley ^{1,7} , Simon Dellicour ^{1,8,9} , Christian Eggeling ^{1,10} , Santiago F. Elena ^{1,11} , Udo Gieraths ¹², Anh D. Ha ¹³ , Will Harvey ¹⁴ , Terry C. Jones ^{1,12,15} , Kevin Lamkiewicz ^{1,2} , Gabriel L. Lovate ^{1,2} , Dominik Lücking ¹⁶ , Martin Machyna ¹⁷ , Luca Nishimura ^{1,18,19} , Maximilian K. Nocke ^{1,20} , Bernard Y. Renard ^{1,21}, Shoichi Sakaguchi ²² , Lygeri Sakellaridi ²³ , Jannes Spangenberg ^{1,2} , Maria Tarradas-Aleman ²⁴ , Sandra Triebel ^{1,2} , Yulia Vakulenko ²⁵ , Rajitha Yasas Wijesekara ²⁶ , Fernando González-Candelas ^{1,11,27} , Sarah Krautwurst ² , Alba Pérez-Cataluña ^{1,28} , Walter Randazzo ^{1,28} , Gloria Sánchez ^{1,28}  and Manja Marz ^{1,2,29,30,31,32,*} 

- ¹ European Virus Bioinformatics Center, 07743 Jena, Germany; ana.abecasis@ihmt.unl.pt (A.B.A.); liam.brierley@liverpool.ac.uk (L.B.); simon.dellicour@ulb.be (S.D.); christian.eggeling@uni-jena.de (C.E.); santiago.elena@csic.es (S.F.E.); terence.jones@charite.de (T.C.J.); kevin.lamkiewicz@uni-jena.de (K.L.); gabriel.lencioni.lovate@uni-jena.de (G.L.L.); maximilian.nocke@ruhr-uni-bochum.de (M.K.N.); bernhard.renard@hpi.de (B.Y.R.); fernando.gonzalez@uv.es (F.G.-C.); alba.perez@iata.csic.es (A.P.-C.); wrandazzo@iata.csic.es (W.R.); gloriasanchez@iata.csic.es (G.S.)
- ² RNA Bioinformatics and High-Throughput Analysis, Friedrich Schiller University Jena, 07743 Jena, Germany; sarah.krautwurst@uni-jena.de
- ³ Global Health and Tropical Medicine, GHTM, Associate Laboratory in Translation and Innovation towards Global Health, LA-REAL, Instituto de Higiene e Medicina Tropical, IHMT, Universidade NOVA de Lisboa, UNL, Rua da Junqueira 100, 1349-008 Lisboa, Portugal
- ⁴ Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada
- ⁵ Donnelly Centre, University of Toronto, Toronto, ON M5S 1A1, Canada
- ⁶ Leibniz Institute of Virology, Department Viral Zoonoses—One Health, 20251 Hamburg, Germany; sebastian.beck@leibniz-liv.de
- ⁷ Department of Health Data Science, University of Liverpool, Liverpool L69 3GF, UK
- ⁸ Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, CP160/12, 50 av. FD Roosevelt, 1050 Bruxelles, Belgium
- ⁹ Laboratory for Clinical and Epidemiological Virology, Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, University of Leuven, 3000 Leuven, Belgium
- ¹⁰ Institute of Applied Optics and Biophysics, Friedrich Schiller University Jena, Max-Wien-Platz 1, 07743 Jena, Germany
- ¹¹ Institute for Integrative Systems Biology (I2SysBio), CSIC-Universitat de Valencia, Catedrático Agustín Escardino 9, 46980 Valencia, Spain
- ¹² Institute of Virology, Charité, Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany
- ¹³ Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA
- ¹⁴ The Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, UK
- ¹⁵ Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
- ¹⁶ Max-Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany
- ¹⁷ Paul-Ehrlich-Institut, Host-Pathogen-Interactions, 63225 Langen, Germany
- ¹⁸ Department of Genetics, School of Life Science, The Graduate University for Advanced Studies (SOKENDAI), Mishima 411-8540, Japan
- ¹⁹ Human Genetics Laboratory, National Institute of Genetics, Mishima 411-8540, Japan
- ²⁰ Department for Molecular & Medical Virology, Ruhr University Bochum, 44801 Bochum, Germany
- ²¹ Digital Engineering Faculty, Hasso Plattner Institute, University of Potsdam, 14482 Potsdam, Germany
- ²² Department of Microbiology and Infection Control, Faculty of Medicine, Osaka Medical and Pharmaceutical University, Osaka 569-8686, Japan; shoichi.sakaguchi@ompu.ac.jp
- ²³ Institute for Virology and Immunobiology, University of Würzburg, Versbacher Str. 7, 97078 Würzburg, Germany
- ²⁴ Computational Genomics Lab., Department of Genetics, Microbiology and Statistics, Institut de Biomedicina UB (IBUB), Universitat de Barcelona (UB), 08028 Barcelona, Spain
- ²⁵ Martsinovskiy Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov First Moscow State Medical University, 119991 Moscow, Russia
- ²⁶ Institute for Bioinformatics, University of Medicine Greifswald, Felix-Hausdorff-Str. 8, 17475 Greifswald, Germany
- ²⁷ Joint Research Unit “Infection and Public Health” FISABIO, University of Valencia, 46010 Valencia, Spain



Citation: Hufsky, F.; Abecasis, A.B.; Babaian, A.; Beck, S.; Brierley, L.; Dellicour, S.; Eggeling, C.; Elena, S.F.; Gieraths, U.; Ha, A.D.; et al. The International Virus Bioinformatics Meeting 2023. *Viruses* **2023**, *15*, 2031. <https://doi.org/10.3390/v15102031>

Academic Editor: Eric O. Freed

Received: 10 July 2023

Revised: 8 September 2023

Accepted: 14 September 2023

Published: 30 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

²⁸ VISAFFELab, Department of Preservation and Food Safety Technologies, Institute of Agrochemistry and Food Technology, IATA-CSIC, 46980 Valencia, Spain

²⁹ German Center for Integrative Biodiversity Research (iDiv), 04103 Halle-Jena-Leipzig, Germany

³⁰ Michael Stifel Center Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

³¹ Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena, 07745 Jena, Germany

³² Leibniz Institute for Age Research—Fritz Lippman Institute, 07745 Jena, Germany

* Correspondence: manja@uni-jena.de

Abstract: The 2023 International Virus Bioinformatics Meeting was held in Valencia, Spain, from 24–26 May 2023, attracting approximately 180 participants worldwide. The primary objective of the conference was to establish a dynamic scientific environment conducive to discussion, collaboration, and the generation of novel research ideas. As the first in-person event following the SARS-CoV-2 pandemic, the meeting facilitated highly interactive exchanges among attendees. It served as a pivotal gathering for gaining insights into the current status of virus bioinformatics research and engaging with leading researchers and emerging scientists. The event comprised eight invited talks, 19 contributed talks, and 74 poster presentations across eleven sessions spanning three days. Topics covered included machine learning, bacteriophages, virus discovery, virus classification, virus visualization, viral infection, viromics, molecular epidemiology, phylodynamic analysis, RNA viruses, viral sequence analysis, viral surveillance, and metagenomics. This report provides rewritten abstracts of the presentations, a summary of the key research findings, and highlights shared during the meeting.

Keywords: bioinformatics; tools; machine learning; bacteriophages; virus discovery; virus classification; virus visualization; viral infection; viromics; molecular epidemiology; phylodynamic analysis; RNA viruses; viral sequence analysis; viral surveillance; metagenomics

1. Introduction

The International Virus Bioinformatics Meeting (ViBioM) was the sixth edition of the virus bioinformatics meeting organized by the European Virus Bioinformatics Center (EVBC). The EVBC was founded in 2017 to bring together experts in virology and virus bioinformatics in Europe [1,2]. The EVBC is constantly growing, having currently 276 members (~12% increase since the last meeting in 2022 [3]) from 158 research institutes distributed over 41 countries worldwide. ViBioM 2023 took place in Valencia, Spain from 24–26 May 2023. From all registered participants, ~23% are EVBC members; thus, ViBioM is attracting scientists far beyond the EVBC community. In contrary to 2022, the participants had a highly interactive scientific environment by face-to-face interactions. In total, the meeting featured 8 invited and 19 selected talks in eleven sessions on three days, as well as 74 posters, which were presented during three poster sessions.

2. Scientific Program

A number of high-quality presentations were given by leading experts and junior scientists on several different topics in virus bioinformatics. From 61 submissions (a ~17% increase compared to 2022 [3]), we selected 21 talks (acceptance rate: ~34%; two of them have been merged, and one has been invited for a keynote). Due to the high amount of submissions on SARS-CoV-2-related research, we decided to add an additional conference day. On the first day, we were focusing on five topics in five sessions: Phages (see Section 2.1), Virus discovery and classification (see Section 2.2), Virus visualization (see Section 2.3), Viral infection (see Section 2.4), and Viromics (see Section 2.5). On day two, we aimed for three sessions, namely: Molecular epidemiology and phylodynamic analyses (see Section 2.6), RNA viruses: structure and evolution (see Section 2.7), and Viral sequence analysis (see Section 2.8). Finally on the third day, we explored the fields of: Machine learning in viral surveillance (see Section 2.9), Viral pathogenesis (see Section 2.10), and Metagenomics for Identifying and Tracking Potential Zoonotic Viruses (see Section 2.11). Lara Fuhrmann

(VILOCA: Sequencing quality-aware haplotype reconstruction and mutation calling for short- and long-read data) was selected for Best ECR Talk Award. During three poster sessions, 74 posters were presented. Two presenters were selected for the Best Poster Award: Muriel Ritsch (Non-retroviral RNA viruses integrated into the human genome) and Jordi Sevilla (Tracking intra host evolution of SARS-CoV-2).

2.1. Phages

Phages are relevant to all human beings due to their abundance, predation, disease control potential, influence on the gut microbiome, biotechnological applications, and contributions to genetic diversity and evolution. Their study and utilization hold great promise for addressing challenges in healthcare, agriculture, and environmental sustainability. Beside an awe-inspiring and fantastic presentation by Robert Edwards, we received also a brilliant presentation by Yasas Wijesekara, who is also the travel award winner for this year. This session was chaired by EVBC member Anca Segall (San Diego State University, San Diego, CA, USA).

Jaeger: A Deep Learning Approach for Predicting Bacteriophage Sequences in Metagenomic Data (by Rajitha Yasas Wijesekara)

Microbial communities are complex admixtures containing vastly different organisms representing the three domains of life and their viruses. Bacteriophages, the viruses that infect bacteria, are ubiquitous in almost every environment and play a crucial role in shaping the ecological and evolutionary processes of ecosystems by controlling bacterial abundances [4–7]. They also influence bacterial phenotype in the virocell state by altering bacterial metabolism and drive global nutrient flow [8,9]. Detecting phages in metagenomic datasets requires specialized bioinformatic tools. Wijesekara et al. presented Jaeger <https://github.com/Yasas1994/Jaeger>, a novel artificial intelligence (AI) tool for predicting bacteriophage sequences in metagenomic data that can be applied to individual reads (70% accuracy), assembled contigs (90% accuracy), and bins (93% accuracy). Additionally, Jaeger can detect prophages (74% accuracy), and identify other sequence categories such as eukaryotic, bacterial, and archaeal genomes. Jaeger utilizes a deep learning model with dilated convolution and residual connections that learn feature representations from nucleotide sequences, which are subsequently used for classification. The authors demonstrate that their novel neural architecture performs better than other available methods. Specifically, they compared the performance of Jaeger to PPRMeta, DeepVirFinder, Seeker, and VirSorter2 on the phages in the IMGVR (v4) database and real metagenomic datasets from three different biomes, showing 10–35% decrease in false positive rate without compromising on sensitivity [10–15]. Together, Jaeger adds a new AI-powered tool to the metagenomics toolbox that will help to understand the composition of complex communities from metagenomic sequencing.

2.2. Virus Discovery and Classification

Virus discovery and classification are fundamental aspects of virology. Through ongoing research and technological advancements, new viruses are continuously being identified. Classification involves categorizing viruses based on their genetic material, structure, and mode of replication. This knowledge aids in understanding viral evolution, transmission, and the development of diagnostic tools and antiviral strategies. This session was chaired by Justine Charon (University of Sydney, Australia).

2.2.1. Illuminating the RNA Virome through Ultra-Massive Sequence Analysis (by Artem Babaian)

Transcriptomic/metatranscriptomic sequencing is revolutionizing the exploration of Earth's virome. Yet analysis methods are inefficient and don't scale to the available data. The global biology community has freely shared >30 petabases ($3 \cdot 10^{16}$ nt) of sequence data from 10+ million biological samples [16,17]. Painstakingly collected over 15 years, public

data encompass all continents, oceans, thousands of animals, plant, and fungal species, and estimated [16,18] to be valued at \$3.6–14.9 billion dollars in direct sequencing cost.

To uncover the total diversity of RNA viruses, Babaian et al. developed a cloud-based sequence alignment platform called Serratus www.serratus.io, with which they analyzed 7.4 million public sequencing datasets for the RNA viral hallmark gene, RNA-dependent RNA polymerase [19]. They identified the equivalent of >300,000 novel RNA viruses which is over an order of magnitude increase of known virus diversity; including at least nine new species of Coronaviruses (CoV).

Planetary-scale informatics provide unprecedented depth of insights with which to describe virus evolution and ecology. Seven of the novel CoV identified are encoded on segmented genomes (of monophyletic origin). The majority of novel nidoviruses were identified in samples from aquatic vertebrates (axolotl, leopard frog, seahorse, fugu fish, etc.) supporting that there is a vast uncharacterized reservoir of marine nidoviruses. With a high-sensitivity search, Babaian et al. detect genome-fragments from >40 distinct uncharacterized nidoviruses. For both novel and known nidoviruses, the authors expand virus host-ranges, geographic distributions, ecological niches and identify potential hidden reservoirs in an unbiased manner which at times, challenge the preconceptions about nidovirales [19].

Explosive growth in data volumes and the breadth of Earth's biodiversity captured by global biological sequencing, see Figure 1, will change bioinformatics in the near future. The next decade of virology will be illuminated by computational advances and requires a shift in our conceptualization of virus discovery and its applications to pandemic surveillance. What can be learned today (with scalable methods) from Nidovirales lays the foundation for how we will explore the 100+ million virus species we project to discover by 2030.

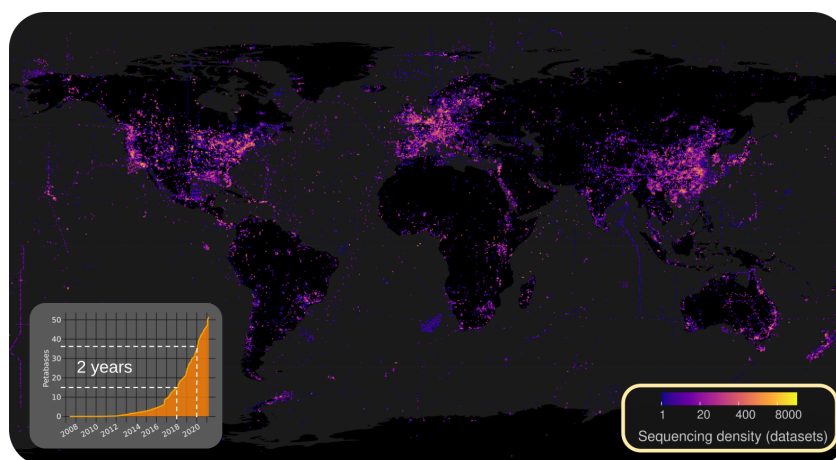


Figure 1. Unlocking Earth's Virome via public data. Petabytes of total (public and dbGAP) sequencing data in the Sequence Read Archive (SRA), collected over 15 years is growing exponentially (inlay) and spans the globe. Data geo-coded from SRA [16] and associated BioSamples [17] meta-data.

2.2.2. RNA Virus Discovery Using HMM of Large-Scale RNA-Dependent RNA Polymerase Sequence Data: NeoRdRp 2.0 (by Shoichi Sakaguchi)

RNA-dependent RNA polymerase (RdRp) is unique to RNA viruses and serves as a critical marker in searching RNA viruses from RNA sequencing data. Since virus detection methods rely on sequence similarity, it is difficult to identify viruses whose similar sequences are unavailable in the database. Therefore, Sakaguchi et al. developed an analysis pipeline that leverages a hidden Markov model (HMM) to enhance the detection of RdRp sequences with low similarity. The authors utilized three reported RdRp amino acid sequence datasets (hereafter called the 'seed datasets'): the 4620 sequences by Wolf et al. [20], the 14,680 sequences by Edgar et al. [19], and the 209,588 sequences by Zayed et al. [21]. Additionally, Sakaguchi et al. used 18,790 sequences from RNA viruses

registered in the NCBI Virus database as a ‘bulk virus dataset’ and 565,928 sequences registered in UniProtKB as a ‘test dataset’. The analysis pipeline began with clustering based on sequence similarity within the seed datasets, followed by a multiple sequence alignment for each cluster (see Figure 2). The authors defined a domain as a region where at least 75% of the entire sequence was aligned and created an HMM profile for each domain. They then applied the HMM profiles to the *hmmsearch* and detected RdRp sequences from the bulk virus dataset. Finally, they constructed the HMM profile NeoRdRp <https://github.com/shoichisakaguchi/NeoRdRp> from the RdRp domain dataset obtained from seed datasets and the RdRp domain sequences derived from the *hmmsearch* of the bulk virus dataset. The NeoRdRp yielded 13,038 HMM profiles, which encompassed known RdRp motifs. Sakaguchi et al. applied these HMM profiles to perform a *hmmsearch* of the test dataset and successfully detected 832 out of 836 RdRp sequences. The accuracy and specificity were 99.5% and 79.9%, respectively. It is planned to continually refine the system by updating and searching RNA sequencing data, aiming to utilize the system to detect novel RNA viruses. Version 1.1 has been reported in a paper [22].

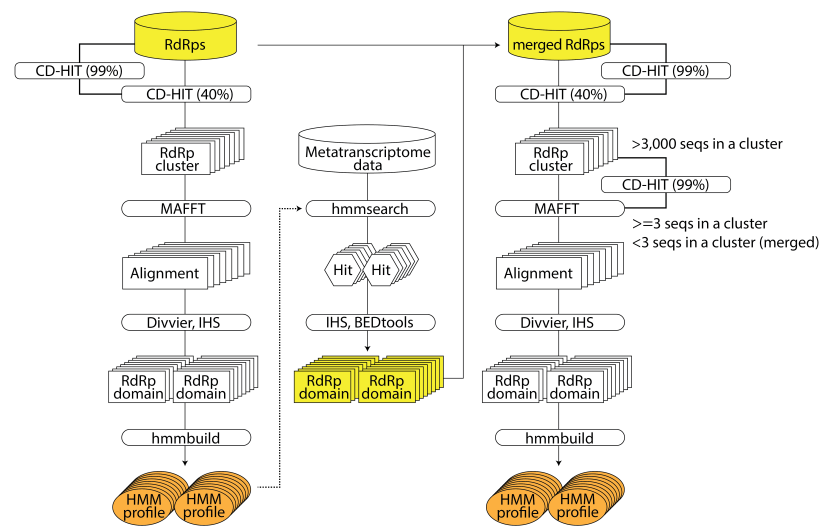


Figure 2. This is an overview of the pipeline utilized in NeoRdRp [22]. The input amino acid data and the resulting HMM profiles are shown in yellow and orange, respectively. First-round HMM profiles are created using curated RdRp seed datasets. These HMM profiles are then used to search for RdRp domain sequences using HMM search. Finally, second-round HMM profiles are created by the seed RdRp datasets with the obtained RdRp domain data.

2.2.3. Automated Classification of Giant Virus Genomes Using Protein Family Barcodes (by Anh Ha)

Large DNA viruses of the phylum Nucleocytoviricota, or ‘giant viruses’, are ubiquitous in the environments and play important roles in shaping the dynamics of global ecosystems. Documented members of the phylum are highly diverse and can be partitioned into six orders and up to 32 potential families [23]. Giant viruses are known to infect a broad range of eukaryotic hosts and appear to have undergone multiple gene exchanges with their hosts [24]. As a result, their genomes often carry diverse genes involved in various cellular processes, such as the TCA cycle, translation, light sensing, and cytoskeletal dynamics [25–27]. Due to the large phylogenetic breadth of this viral group and the highly complex, chimeric nature of their genomes, taxonomic classification of giant viruses, particularly incomplete metagenome-assembled genomes (MAGs) could be challenging. Here Ha et al. utilized a machine learning approach to predict the taxonomic classification of novel giant virus MAGs based on profiles of protein family content, which they refer to as protein family barcodes. They applied random forests to a training set of 1503 quality-checked, phylogenetically diverse Nucleocytoviricota genomes using a pre-selected set

of giant virus orthologous groups (GVOGs). The classification model was predictive of giant viruses' taxonomic group with a cross-validation accuracy of 99.5% to the Order level and 97% to the Family level. The authors observed that no individual GVOGs or genome features were critical to the algorithm's performance and the model's predictions, suggesting that classification predictions were based on a broad genomic signature, which lessened the necessity of a fixed set of marker genes for taxonomic assigning purposes. Their classification model was validated with an independent test set of 814 giant virus genomes with varied genomic completeness and predicted taxonomy with 97.5% and 95% accuracy to the Order and Family level, respectively. The results provide a fast and accurate method for the classification of giant viruses that could easily be adapted to other viral groups.

2.2.4. Using gb2seq to Work with Unannotated Viral Genomes Based on a GenBank Reference (by Terry Jones)

The talk by Jones et al. presented gb2seq <https://github.com/VirologyCharite/gb2seq>, a Python library and associated command-line scripts that derive information regarding unannotated viral genomes (e.g., a consensus called from a BAM file following alignment with bwa or bowtie), based on annotations in a GenBank reference. The library provides for: extraction of aligned features as nucleotide or amino acid sequences; retrieving information about what is at a site (features, nucleotide, amino acid, codon, frame, etc); translation of offsets between the reference and the unannotated genome with offsets that are absolute or relative to a feature; detailed alignment information for features; use of different aligners (MAFFT and edlib, currently); JSON annotations of the features of an unannotated genome; and convenience methods for checking genomes for sets of expected nucleotide or amino acid values. The command-line scripts provide a simple interface to the library functions, e.g., to extract a translated feature from a set of genomes and check for how many expected substitutions are present.

2.3. Virus Visualization

The integration of genomic information and imaging techniques offers valuable insights into viral biology. By combining approaches such as Fluorescence In Situ Hybridization (FISH), Genome-Wide Association Studies (GWAS), and multi-omics integration, researchers can gain a deeper understanding of viruses at a spatial and molecular level. Bioinformatics tools assist in automated analysis of imaging data, aiding in the identification and quantification of viral structures. This integration enhances our understanding of viral biology, mechanisms of viral pathogenesis, and facilitates the development of targeted interventions and therapeutics. Manja Marz (FSU, Jena, Germany) has invited Christian Eggeling to give a presentation on his research area with the aim to open a pathway for research combining virus omics data and morphological features.

Advanced Optical Microscopy of Virus-Cell Interactions: Challenges and Potentials (by Christian Eggeling)

Understanding virus infectivity also requires revealing molecular interaction details during the virus-host interaction. For example, an increased mobility and subsequent aggregation of envelope proteins during maturation of HIV-1 virions drives their potential to dock host-cell receptors and thus infectivity, as revealed by optical microscopy [28,29]. Unfortunately, the direct and non-invasive observation of the interactions in the living cell membrane is often impeded by principle limitations of conventional far-field optical microscopes, for example with respect to limited spatio-temporal resolution. However, the advent of super-resolution microscopy techniques has created unique opportunities of investigating the organization and mobility of viral and cellular molecules at the required small spatial scales [30,31]. Taking HIV-1 and SARS-CoV-2 as examples, Eggeling has presented how such advanced optical microscopy approaches help highlighting novel aspects of virus-membrane interactions (Figure 3) and what challenges one still faces. An important issue will be, how these microscopy-based information can be correlated

with bioinformatic data to boost information content. Presenting the potential of microscopy to bioinformatics at the ViBiom 2023 has done a first step.

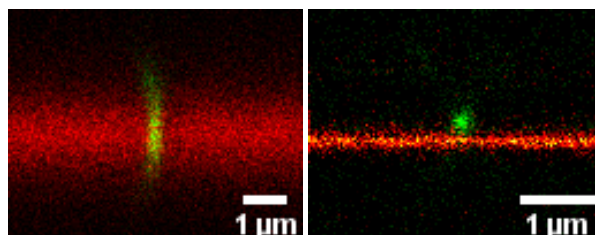


Figure 3. Conventional confocal (**left**) and super-resolution STED (**right**) microscopy images (side x-z view) of a SARS-CoV-2 virus like-particle (green, GFP-labelled) with a supported lipid bilayer (fire scale, DOPC lipids decorated with ACE-2, membranes labelled with fluorescent lipid analog), highlighting the improved spatial resolution to investigate molecular details in such interaction. Data taken by Ziliang Zhao (Jena).

2.4. Viral Infection

Viral infections are a significant public health concern, causing a wide range of diseases in humans and animals. Understanding viral infections is crucial for developing effective prevention strategies, diagnostics, and therapeutics. Bioinformatics plays a vital role in the analysis of viral infections by integrating large-scale genomic, transcriptomic, and proteomic data to identify viral genomes, study their evolution, and predict their pathogenic potential. The session has been chaired by Anamarija Butkovic (Institut Pasteur, Paris, France).

2.4.1. SARS-CoV-2-Host Interactions at the Single-Cell Level: A Dynamical Complex Systems Approach (by Santiago F. Elena)

Elena et al. presented an analysis of the response of three different cell types during the progression of SARS-CoV-2 infection [32]. To do so, they have performed a meta-analysis of three publicly available single-cell RNA sequencing (scRNA-seq) datasets obtained from in vitro inoculation studies of human bronchial epithelial cells [33], and colon and ileum organoids [34]. ScRNA-seq has become a powerful technique to study the dynamic changes in the transcriptome of a population of cell subjected to an external stimulus. Indeed, different cells in the population are found at different stages of infection, as shown by the differences in viral reads that they contain. Therefore, a bulk analysis is clearly inappropriate to identify dynamical gene responses associated to infection. Instead, the authors have classified cells according to their infection status using virus accumulation as a proxy to time, see Figure 4. This approach has revealed that about 90% of genes exhibited a transcriptional response characterized by a triphasic pattern comprising an early down-regulatory phase at the beginning of infection, followed by a transient massive up-regulation and, finally, a final down-regulation as cells' viability declines, see Figure 4. Focusing in the 10% of genes that significantly deviate from this canonical triphasic behavior, Elena et al. found an enrichment in genes related to immune responses, translation and mitochondrial oxidative activities. Interestingly, their analyses have shown that the transcriptional profiles of genes encoding for the main intracellular sensors that recognize double-stranded RNAs and trigger the innate immune response upon infection, MDA-5 and RIG-I, do not deviate from the majoritarian response. This observation suggests that the transcriptional shutdown of the IFN response induced by MDA-5 and RIG-I is probably not specific, and results from a more general mechanism affecting most genes in the infected cell. Additionally, a correlation network analysis revealed a distinct correlated response of mitochondrially-encoded genes and genes involved in translation. Finally, the authors proposed a mechanistic model to explain this dynamic profiles. The key player in the model is viral protein Nsp1, that has been shown to block the export of specific mRNAs from the nucleus [35,36]. Expression of nsp1 acts like an off-switch for Nsp1-sensitive transcripts (e.g., interferon genes) that results in their accumulation in the nucleus while Nsp1-insensitive transcripts can transit normally

to the cytoplasm. The middle phase corresponds to the inactivation of the Nsp1-induced blockage, which releases the stalled transcripts to the cytoplasm that results in the increased transcription levels observed during the middle phase. Finally, the late phase corresponds to a shutdown of transcription that also results in a global decrease in transcripts levels.

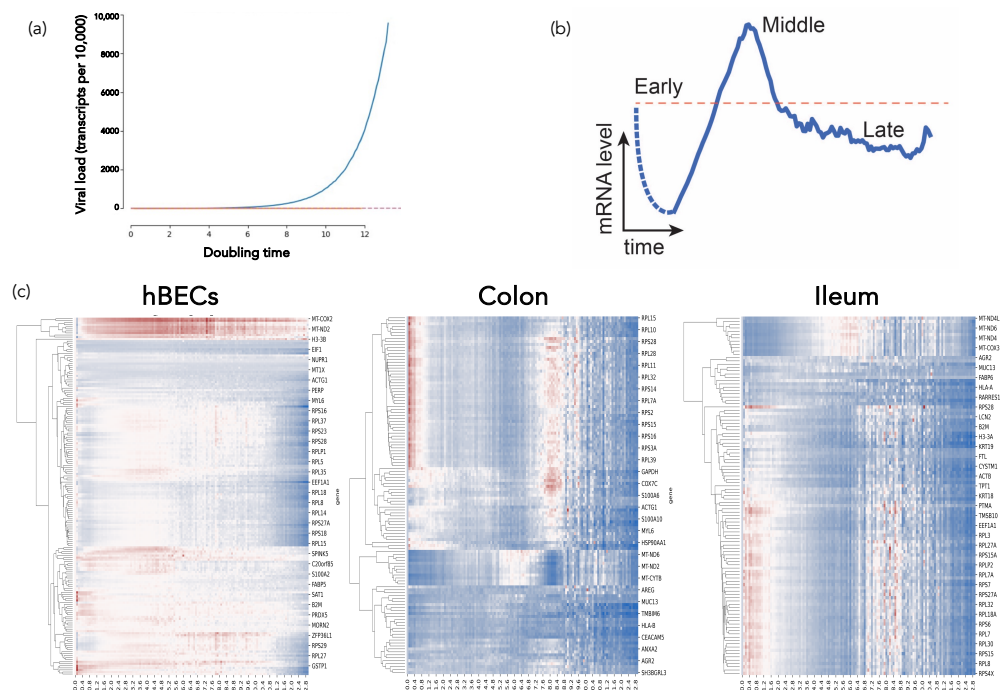


Figure 4. Dynamics of virus accumulation and cellular responses along infection. (a) Virus accumulation as a function of doubling times. Blue lines represent the number of viral transcripts in infected cells. Red lines the baseline established for non-infected cells. (b) Triphasic pattern of gene expression observed for 90% of cellular genes. (c) Clustering of gene expression profiles for each one of the three cell types included in the study.

2.4.2. Metabolic Labeling, Time Series, and Single Cells: A Multifaceted Approach to Studying Infection (by Lygeri Sakellaridi)

Viral infection is a dynamic biological process, facilitated by complex interactions between host and viral genes. On a cellular level, infection may result in different outcomes, e.g., lytic or latent. Elucidating the molecular mechanisms that determine the infection outcome is a challenging problem. In order to tackle this problem, Sakellaridi et al. proposed an approach that combines single cell RNA sequencing with metabolic RNA labeling. Metabolic labeling utilizes 4-thiol-uridine (4sU) which is introduced into cells and subsequently incorporated into newly transcribed RNA. Before sequencing, RNA is subjected to chemical treatment that results in the conversion of 4sU into cytosine or cytosine analogs [37–39]. This allows distinguishing between RNA that was synthesized before and after the start of labeling based on the presence of T-C mismatches in the newly transcribed reads. Accurate quantification of new and pre-existing reads is hindered by the low incorporation rate of 4sU in cells [37], as well as the presence of naturally occurring mismatches in reads. To address these challenges, the authors' lab developed GRAND-SLAM <https://github.com/erhard-lab> [40]: a statistical approach that uses a binomial distribution to provide unbiased estimates of the percentage of labeled RNA per gene, as well as its posterior distribution that represents uncertainty of quantification. This output can be incorporated in downstream analyses, such as estimation of RNA kinetics and changes in gene expression. To that end, Sakellaridi et al. developed grandR [41], a computational package that facilitates such analyses while taking advantage of the uncertainty estimates. Metabolic labeling of single cells makes it possible to acquire two snapshots of expression per cell in a population, corresponding to the transcriptomic state before (past state) and after (current

state) the end of labeling. Heterogeneity in the past state can be used to make potentially causal inferences in the current state: e.g., anti-correlation of expression of a given gene in the past state against the total viral gene expression in the current state would indicate an anti-viral gene. Extending this reasoning, time courses allow following the fate of cells over extended periods of time. Sakellaridi et al. introduce a novel trajectory inference approach for connecting cells from consecutive time points based on similarity between their past and current transcriptomic profiles. The authors envision the data as a network where cells are connected across time, and their changing expression trajectories 'flow' through the network. Mathematically, this is equivalent to the minimum cost-maximum flow problem and can be solved by a well-established optimization algorithm [42]. Solving the network results in expression trajectories that reflect the different paths a biological process may follow across time. By constructing such trajectories in a population of infected cells, it becomes possible to determine factors that influence the infection outcome.

2.5. Viromics

Viromics heavily relies on computational tools for the comprehensive analysis of viral genomes and metagenomic data. These tools encompass a wide range of bioinformatics algorithms and pipelines that aid in various stages of viromic analysis, including sequence quality control, read assembly, viral genome binning, taxonomic classification, and functional prediction. Computational approaches are essential for unraveling the complexity of viromic datasets, extracting valuable insights about viral diversity, evolution, and interactions within microbial communities. These tools enable the identification of novel viral species, characterization of viral genes and functional elements, exploration of viral-host interactions, and the potential discovery of new antiviral targets. Continuous advancements in computational tools and techniques are crucial for enhancing our understanding of the virosphere and its implications in diverse fields, from environmental microbiology to human health. Our viromics session has been hosted by Spyros Lytras (MRC, Glasgow, UK).

2.5.1. Ancient Virome Analyses Using Metagenomic Data from Ancient Individuals (by Luca Nishimura)

Ancient DNAs have been discovered from various kinds of archeological samples such as bones and mummified tissues. Those ancient samples contain ancient viral genomes which existed in ancient organisms' bodies [43]. Those ancient viral genomes are useful to elucidate past pandemic events and long-term viral evolution. For instance, human pathogenic viruses such as influenza A virus and hepatitis B virus have been analyzed since 1997. However, the number of ancient viruses thus far identified is small; most are human pathogenic viruses. Nishimura et al. analyzed ancient people's whole genomic sequencing (WGS) data to discover ancient viruses more comprehensively. They utilized genomic data from 36 ancient individuals who dwelled in the Japanese archipelago and also more than 300 publicly available data. Firstly, the authors conducted *de novo* assembly of non-human reads to obtain longer contigs. Those contigs were used for homology search against the modern viral reference genomes and other methods such as machine-learning methods and bacterial immunological memories to find viral genomes. As a result, more than 50,000 candidates of ancient viral contigs were detected, including about 200 high-quality ancient viral genomes. To characterize ancient virome, Nishimura et al. analyzed ORF components and phylogenetic relationships based on those contigs. For example, they obtained the nearly complete sequence of the Siphovirus contig89 (CT89) existing in the human oral environment [44]. The authors estimated the relationships between the modern CT89 and the most recent common ancestor by phylogenetic analyses. They then compared the viral components in each sample showing the differences between ancient and modern samples which might reflect different dietary characteristics. Those results suggest that the metagenomic data from ancient samples are useful for elucidating ancient viral characteristics and long-term viral evolution.

2.5.2. One's Trash Is Another's Treasure—Mining Viromics Datasets for Traces of EV Mediated Horizontal Gene Transfer (by Dominik Lücking)

Marine environmental viral metagenomes, commonly referred to as 'viromes', are typically generated by physically separating viral-like particles (VLPs) from the microbial fraction based on their size and mass. However, most of the methods used to enrich extracellular vesicles (EVs) and gene transfer agents (GTAs) simultaneously [45,46]. Consequently, the sequence space traditionally referred to as a 'virome' contains host-associated sequences, transported via EVs or GTAs. Lücking et al. therefore propose to call the genetic material, isolated from size-fractionated (0.22 µm) and DNase treated samples, protected environmental DNA (peDNA), Figure 5. This sequence space contains viral genomes, DNA transduced by proviruses and DNA transported in EVs and GTAs. Since there is currently no definitive genetic signature for EV-transported DNA, scientists rely on the successful removal of contaminating remaining cellular and free DNA, when analyzing peDNA. Using marine samples collected from the North Sea, the authors generated a thoroughly purified peDNA dataset and developed a bioinformatic pipeline to determine the potential origin of the purified DNA. This pipeline (<https://github.com/dluecking/mviest>) was applied to their dataset as well as existing global marine 'viromes', enabling the identification of known GTA and EV producers, and organisms with actively transducing proviruses as the source of the peDNA, thus confirming the reliability of their approach. Additionally, Lücking et al. identified novel and widespread EV producers and found quantitative evidence suggesting that EV-mediated gene transfer plays a significant role in driving horizontal gene transfer (HGT) in the world's oceans.

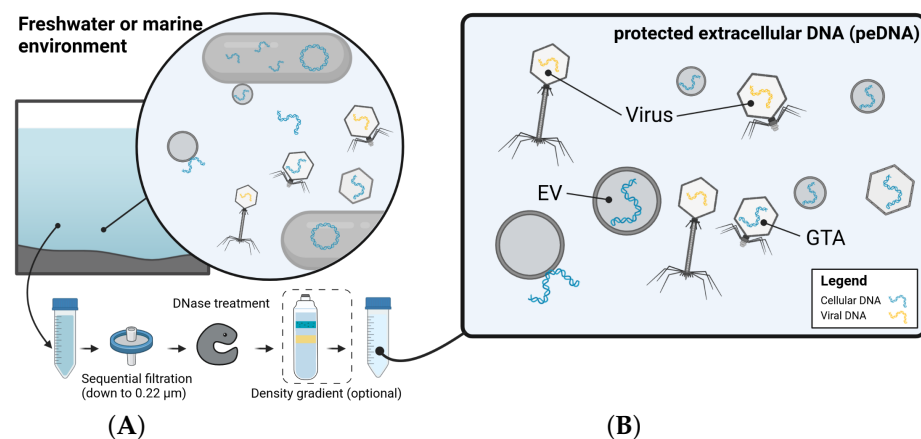


Figure 5. Conceptual composition of protected extracellular DNA. (A) Microbial entities present in a water body: microbial cells, viruses containing viral and microbial genetic material, gene transfer agents and extracellular vesicles containing host DNA. After size filtration (0.22 µm) and DNase treatment and, if applicable, purification via density gradients, microbial cells and free DNA are removed. (B) The remaining DNA makes up the sequence space of protected extracellular DNA, peDNA.

2.6. Molecular Epidemiology and Phylodynamic Analyses

Molecular epidemiology and phylodynamic analysis are two interconnected fields that utilize computational tools to study the spread and evolution of infectious diseases. Molecular epidemiology combines molecular biology and epidemiology to investigate the transmission dynamics, source attribution, and genetic characteristics of pathogens. Phylodynamic analysis involves the use of computational models and phylogenetic methods to infer the evolutionary history, population dynamics, and transmission patterns of pathogens based on genetic data. Computational tools play a crucial role in processing large-scale genomic datasets, reconstructing phylogenetic trees, estimating evolutionary rates, detecting transmission clusters, and identifying key determinants of disease emergence and spread. These tools enable researchers to gain insights into the factors driving disease

outbreaks, design effective control strategies, and contribute to global health surveillance efforts. This session was hosted by Francesca Young (MRC, Glasgow, UK).

2.6.1. HIV-1 Transmission Studies Using Phylogenetics: Can Evolution Help Guide Public Health Decisions? (by Ana Abecasis)

Since the 1990s, HIV-1 transmission chains reconstruction has been used for different purposes. It's first use for court cases of HIV-1 transmission has led to important conceptual and ethical discussions [47]. Later on, HIV-1 transmission chains reconstruction combined with socio-demographic and behavioral data has been used frequently in the context of public health epidemiological studies [48]. By combining socio-demographic, behavioral and clinical data, Abecasis et al. reconstructed transmission chains in different contexts to better understand the most important determinants of transmission of HIV-1 infection in each scenario. They addressed potential pitfalls and methodological constraints, and presented the main challenges, implications and applicability to future studies. The authors' results indicate the importance of its use to complement classical epidemiological approaches, indicating routes of transmission of HIV-1 infection in migrants living in Portugal [49] and transmission routes more associated with transmission of antiretroviral drug resistance [50].

2.6.2. Molecular Epidemiological Approaches to Investigate the Dispersal Dynamic of Viruses and the Environmental Factors Impacting It (by Simon Dellicour)

Recent advances in genomics, mathematical modelling and computational biology have enabled molecular approaches to become key methods to investigate the spread of viral infectious diseases. In the emerging field of molecular epidemiology, genetic analyses of pathogens are used to complement traditional epidemiological methods in various ways. For instance, genetic analyses offer the possibility to infer linkages between infections that are not evident without analyzing viral genomes. In particular, the development of phylogeographic methods has enabled to reconstruct dispersal history of epidemics in a discretised or on a continuous space, using only a relatively limited number of viral sequences sampled from known locations and times. At the Spatial Epidemiology Lab (SpELL, ULB) <https://github.com/sdellicour>, Dellicour et al. develop and apply new analytical approaches exploiting such phylogeographic reconstructions to test epidemiological hypotheses about the external and environmental factors impacting the dispersal history and dynamic of viral epidemics [51–53].

2.6.3. Phylodynamic Analysis of A(H5N1) Highly Pathogenic Avian Influenza Viruses Provides Insight into Movement Dynamics and Host Specificity (by Will Harvey)

Phylodynamic analyses, modelling approaches based on co-analysis of genetic sequences and associated metadata within a phylogenetic framework, can shed insight on the characteristics and drivers of outbreaks of infectious diseases. Since 2021, highly pathogenic avian influenza (HPAI) viruses of subtype A(H5N1) have caused a panzootic of unprecedented scale. This has affected both wild and domestic birds with high mortality outbreaks in atypical host species such as shorebirds. Spill-overs in a diverse range of mammalian species including die-offs involving thousands of pinnipeds have raised concerns over zoonotic potential. These viruses can readily exchange genomic segments with local low pathogenic avian influenza (LPAI) viruses via reassortment, a mechanism that can facilitate dramatic phenotypic change. Observational data suggest changes in the seasonality of recent HPAI A(H5N1) viruses and relaxation of host specificity, however virological or ecological explanations remain elusive.

Using phylodynamic approaches, it is possible to reconstruct spatio-temporally resolved phylogenetic trees and examine trends relative to pre-panzootic A(H5N8) viruses. Examination of patterns in the timing and locations of reassortment events shows variation in the evolutionary trajectories of reassortant lineages. Exploration of possible changes in host preference or specificity using tree-based metrics to estimate fitness in different host types shows, for example, lineage adapted to spread in birds of the avian orders Charadriiformes, that acquired gene segments LPAI H13 viruses. This demonstrates phenotypic

diversification among recent HPAI A(H5N1) viruses, see Figure 6. Phylogenetic discrete trait modelling indicates this modification in host specificity coincides with changes in the times and locations of virus movements such as those between Great Britain and continental Europe. Such diversification has consequences for spill-over opportunities at the wildlife-domestic poultry interface. Furthermore, a relaxation of host tropism increases the likelihood of further reassortment events increasing the potential for emergence of novel genotypes with changed characteristics.

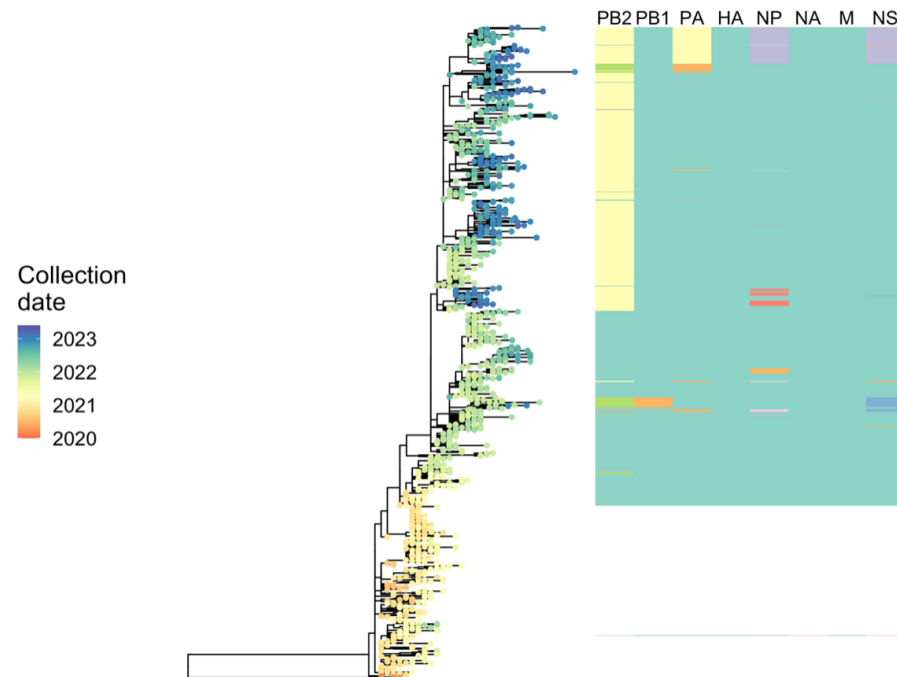


Figure 6. A phylogeny for HPAI H5N8 and descendant H5N1 viruses sampled in Europe since 2020 alongside schematic showing reassortment profile. Haemagglutinin phylogenetic tree generated from stratified sample of haemagglutinin nucleotide sequences sampled in Europe. Sampling date is indicated by tip node color according to the color key. To the right, a schematic shows clusters for each genomic segment for viruses of the H5N1 subtype, while the absence of such information indicates an H5N8 genome. For each segment/column, colors are assigned independently therefore the same color in two different columns should not be interpreted as indicating a shared evolutionary history between segments. Viruses with different combinations of colors across the 8 genomic segments are interpreted as arising from reassortment events.

2.7. RNA Viruses: Structure and Evolution

RNA secondary structures play a crucial role in the function and replication of RNA viruses, and computational tools are invaluable for their analysis. These tools employ algorithms and predictive models to identify and predict RNA secondary structures based on sequence information. They help in identifying conserved structural motifs, such as stem-loops and pseudoknots, that are important for viral replication, translation, and packaging. Computational tools also aid in studying RNA-RNA and RNA-protein interactions within the context of viral infections. By analyzing RNA secondary structures, researchers can gain insights into the mechanisms of viral pathogenesis, identify potential targets for antiviral interventions, and design novel therapeutics. These computational approaches greatly enhance our understanding of RNA virus biology and contribute to the development of effective strategies to combat viral diseases. The session was hosted by Dmitriy Frishman (TUM, Munich, Germany).

2.7.1. Viral RNA Secondary Structures: Canonical and Beyond (by Kevin Lamkiewicz & Sandra Triebel)

In recent years, RNA biology has evolved around identifying and annotating functional (non-)coding RNAs in organisms from all domains of life. Especially for RNA viruses, it is conceivable that genomic regions and transcripts serve additional functions essential for viral replication induced by their RNA secondary structures. In the canonical RNA structure model, base-pairing of complementary nucleotides will fold the RNA into local and global structural elements, such as the prominent stem-loop hairpin structure.

Critical assessment of RNA secondary structures within viral genomes has led to several molecular insights for viruses. For example, the internal ribosome entry site (IRES) of the Hepatitis C virus (HCV) is an RNA structure within the 5' UTR of the genome. It mediates the cap-independent translation initiation of viral proteins [54]. The 3' UTR in flaviviruses exposes, among others, exoribonuclease resistant RNA structures (xrRNAs), prolonging viral genomes within the hosts' cytoplasm [55,56]. Another well-described example is the UTRs of coronaviruses. Here, several stem-loops are seemingly crucial for viral transcription, translation, and replication [57–59]. Despite well-characterized RNA structures in untranslated regions of human-infecting RNA viruses, structures within coding sequences remain unclear. For example, it is hypothesized that RNA-RNA interactions play an essential role in the discontinuous transcription mechanism of coronaviruses. For Influenza A virus, one debated hypothesis is that the packaging of the eight segments is mediated via RNA-RNA interactions [60–62].

Therefore, the determination and functional analysis of structural RNA elements promises many novel insights into the viral life cycle and further new targets for therapeutic approaches and drugs. Bioinformatic models and tools are used to assess probable and kinetically favorable RNA foldings to facilitate and increase our understanding of RNA secondary structures in viral genomes. Multiple sequence alignments (MSAs) are commonly used to confirm the importance of functional regions via evolutionary conservation. Structural conservation indicated by compensatory mutations is considered with structure-guided MSAs proposed by Sankoff (implemented in, e.g., LocARNA [63]). To evaluate whether such tools directly apply to viral genomes and transcripts, Lamkiewicz and Triebel used LocARNA on a prominent region within the genome of *Filoviridae*. The transcription start site (TSS) is a conserved sequential motif upstream of each ORF in *Filoviridae*. Additionally, to sequence conservation, it is further embedded in an RNA structure which has been well-described in the literature [64,65].

The authors assessed the performance of an automated RNA secondary structure prediction method. The structure-guided MSA calculated by LocARNA is shown in Figure 7A. While structural conservation looks promising, the sequential motif is disrupted for the representative *Dianlovirus*. However, careful manual curation reveals that the TSS motif can be recovered, leading to less sequential diversity in alignment columns involved in base pairings, see Figure 7B. While software development is going in the right direction, the findings underscore the importance of critically evaluating the results obtained from automated tools and highlight the significance of expert intervention in refining and interpreting complex biological data.

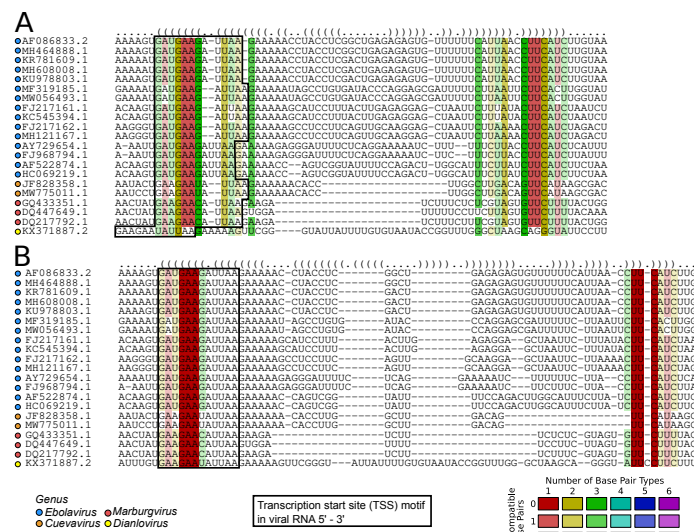


Figure 7. Alignment-based RNA secondary structure prediction upstream of the VP40 gene in *Filoviridae*. (A) The alignment calculated using LocARNA [63] reveals structural conservation with the presence of compensatory mutations. Up to three different base pair types are observed, as indicated by the color scheme (shown in the bottom right), within the non-coding region upstream of the VP40 gene. However, the alignment also shows disruption of the conserved transcription start site (TSS) sequence motif (illustrated by the black box). (B) Through manual curation, both the sequence and structure were preserved in the alignment, accurately representing the conserved elements. This refinement process led to fewer compensatory mutations, as illustrated by the base pair color scheme.

2.7.2. Recombination and Modular Evolution of Positive-Strand RNA Viruses: Similar, but Not the Same (by Yulia Vakulenko)

Recombination is very common in positive-strand RNA viruses. Along with a high mutation rate, it is one of the major forces generating genetic diversity [66]. Vakulenko et al. systematically analyzed patterns of natural recombination in four (+)RNA virus families—*Astroviridae*, *Caliciviridae*, *Picornaviridae* and *Coronaviridae*, using both classical recombination detection methods [67,68] and by comparing correspondence of genetic distances in different genome regions [69]. The authors developed an R package, recDplot <https://github.com/v-julia/recDplot>, for visualizing recombination in viral sequences. A common (and generally known) feature of these virus families was frequent recombination between genome regions encoding nonstructural and structural proteins [67,70–72]. However, the recombination profiles within these genome regions were contrasting. In picornaviruses, there was frequent recombination within the nonstructural genome region with no prominent hotspots and almost absent recombination within the structural genome region. Caliciviruses routinely exchanged full structural and nonstructural blocks of the genome, but had few, if any, recombination events within these regions [73]. In astroviruses, moderate recombination was observed within both structural and nonstructural genomic regions. In coronaviruses, the spike gene, but not other structural proteins genes (E, M, N), was most commonly exchanged between coronaviruses. Recombination within the spike gene occurred more frequently than within the nonstructural region, and more commonly involved the entire domains of the spike protein [69]. Therefore, these (+)RNA viruses with very different genome organization and realization had a common general recombination pattern, which could effectively provide independent evolutionary trajectories for structural and non-structural proteins. Protein(s) function was the major factor defining their relative mobility by recombination. The authors speculate that this recombination profile may reflect the suggestive distinct evolutionary origin of these virus components billions of years ago. On the other hand, viruses of close families could have contrasting recombination patterns within these major genome blocks.

2.7.3. RNAswarm: A Modular Pipeline for Differential RRI Analysis in Influenza A Virus (by Gabriel Lencioni Lovate)

RNA proximity ligation methodologies such as PARIS, SPLASH, and 2CIMPL have offered experimental ways to detect RNA-RNA interactions (RRIs) on a large-scale [74–76]. However, an established bioinformatics pipeline for statistically comparing the frequency of RRIs across different strains or experimental conditions with high throughput has yet to be realized. To fill this gap, Lencioni Lovate et al. have developed RNAswarm <https://github.com/gabriellovate/RNAswarm>, a modular, reproducible Nextflow pipeline specifically designed for high-throughput differential RRI analysis. RNAswarm is openly accessible and is presently under active development. This tool has been successfully utilized in the analysis of SPLASH datasets of influenza A virus (IAV), a substantial global health threat due to its severe morbidity and mortality impacts. Through this application, RNAswarm has revealed differentially structured regions within IAV's segmented RNA genome, as described in a recent publication [62].

With RNAswarm, the authors have been able to identify strain-specific RRI sites in various IAV strains, thus validating previous findings [60] and quantifying variations in prevalence of RRIs across strains [62]. The pipeline, see Figure 8, initiates by processing raw reads from RNA proximity ligation experiments, then identifies RRIs, and finally utilizes DEseq2 to summarize differential RRI representation across distinct IAV strains or experimental conditions [77]. Additionally, a module for *de novo* annotation of discrete interactions—by generating pairwise matrices of chimeric reads and fitting Gaussian Mixture Models to identify normally distributed potential interactions is currently under development.

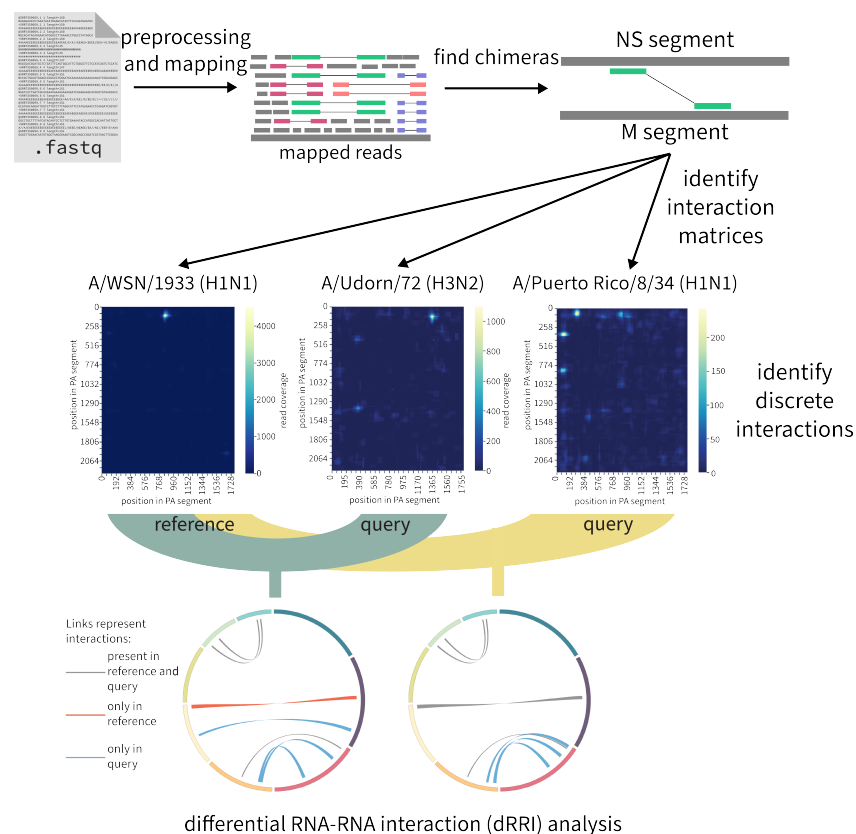


Figure 8. RNAswarm is a reproducible nextflow pipeline that analyzes datasets from RNA proximity ligation experiments. RNAswarm first pre-processes raw reads from RNA proximity ligation experiments, maps them to reference genomes, then identifies RRIs, and finally analyzes differential RRI representation across different strains or experimental conditions.

RNAswarm has proven successful in identifying key RRI across replicates and in detecting highly conserved or flexible interaction sites. It offers a reliable and automated approach for prioritizing and comparing RRI across diverse conditions or organisms. With the potential to uncover new RNA-RNA interactions in viruses and other organisms, the modularity and reproducibility of RNAswarm render it an invaluable tool for researchers investigating RNA-RNA interactions in various contexts.

2.8. Viral Sequence Analysis

Viral sequence analysis plays a pivotal role in understanding viral evolution, diversity, and functional characteristics. Computational tools are essential for analyzing viral sequences and extracting valuable insights. These tools encompass a range of bioinformatics algorithms and pipelines that aid in tasks such as sequence alignment, variant calling, phylogenetic reconstruction, and functional annotation. Computational approaches enable researchers to identify conserved regions, detect genetic variations, predict protein structures and functions, and explore viral-host interactions. Additionally, these tools facilitate the development of diagnostic assays, antiviral drugs, and vaccines by identifying viral targets and epitopes. Viral sequence analysis, powered by computational tools, accelerates our understanding of viral biology and informs strategies for disease surveillance, outbreak control, and therapeutic interventions. This session was hosted by Anne Kupczok (Wageningen University, The Netherlands).

2.8.1. Embedding Segmented Viral Genomes for Visualisation, Search, and Clustering (by Udo Gieraths)

Important human and animal pathogens like influenza and rotaviruses have segmented genomes. Such viruses can reassort during co-infection of a cell with different viral strains, in which case a new viral genome is created via the exchange of segments. In such cases, the evolutionary history, and therefore the phylogenetic trees of each segment, may differ considerably. This property hinders classical phylogenetic analysis and simple searches for similar genomes. Gieraths et al. presented an approach to embed segmented viral genomes in a mathematical space that allows efficient search and clustering. In the context of clustering, outliers that do not fit into any cluster are easily identified. These outliers represent rare reassortment events of particular interest, as reassortant viral genomes can give rise to highly pathogenic variants. Using the example of the influenza A virus, the authors show various applications of our developed segmented viral genome embedding in the context of search, clustering, and outlier detection.

2.8.2. Hyper-EINS: A Tool for Automated Identification of Insertions in the Hepatitis E Virus Hypervariable Region (by Maximilian Nocke)

Hepatitis E virus (HEV) infections are usually asymptomatic and self-limiting, while in immunocompromised or other risk group patients may develop chronic courses [78]. The hypervariable region (HVR) within HEV's first open reading frame is known to integrate sequence snippets of human and viral origin. Some insertions are associated with replication fitness *in vitro* and chronicity *in vivo* [79–82]. The off-label drug ribavirin is commonly used to treat chronically infected patients, but presents a high rate of treatment failures [83].

With their tool Hyper-EINS, Nocke et al. aim to provide a time efficient bioinformatics pipeline tool that automates identification and validation of insertions in high-throughput sequencing (HTS) data, while offering an easy to use graphical user interface (GUI) to simplify accessibility. Their tool will allow early identification of insertions possibly linked to treatment failure.

During Hyper-EINS development, the authors used, implemented and tested a broad variety of informatics tools and languages to increase efficiency of run time and storage usage.

A first command line only version of the analysis tool was written in Python 3 Python 3.10.12 and established the general workflow. Combined use of Cutadapt [84] and Trimmomatic [85] was applied to trim reads in an automated manner. To make read assign-

ment sensitive to unexpected insertions, *blastn* was integrated, accessing NCBI nucleotide database remotely [86,87]. Identified insertions were exclusively of humane origin. Due to this, *blastn* was replaced with *MMSeqs2* [88]. Furthermore, introduction of *MMSeqs2* to Hyper-EINS reduces false positive discovery rate by evading poorly annotated data from the NCBI nucleotide database. Computation of full insertion sequences is achieved by calling the *de novo* assembler Trinity [89].

To speed up analysis processes, Hyper-EINS was reimplemented in Julia. Furthermore, exchanging *blastn* with *MMSeqs2* reduced run time drastically and eliminated the need to access an online database, by providing a local database containing HEV and human reference genome. The tool's Julia version features a Gtk based prototype GUI.

Hyper-EINS was validated by monitoring dynamic rearrangements in the HVR of a chronically HEV infected patient. Nocke et al. identified novel insertions of critical impact for viral fitness as proven in a subgenomic replicon system *in vitro*. Interestingly, content and distribution of insertions in the viral population were very dynamic in this patient, underlining an important role in HEV chronicity and treatment failure.

In conclusion, Hyper-EINS has been designed as a user-friendly tool for detecting insertions of human or viral origin in the HVR of HEV from HTS data using computers with limited RAM and processing power. The graphical user interface visualizes the pipeline's output to make data interpretation and validation easier for the user. Early identification of HVR rearrangements in HEV infected patients can guide treatment decisions in a personalized medicine approach, based on both amplicon sequencing data specifically of the HVR or more spanned genomic regions covering the HVR. Hyper-EINS is written in Julia and will be publicly available in the future.

2.8.3. Magnipore: Predicting Differential Single Nucleotide Changes in Oxford Nanopore Technologies Sequencing Signal in SARS-CoV-2 (by Jannes Spangenberg)

Oxford Nanopore Technologies (ONT) revolutionizes the field of RNA analysis by enabling direct sequencing of ribonucleic acids (RNA) and facilitating the detection of RNA modifications. RNA modifications play a crucial role in cellular processes, including gene expression regulation, RNA stability, and protein synthesis [90]. They also have a significant impact on viral infection, replication, and even the host antiviral innate immunity [91–95].

However, existing methods and basecallers have limitations in directly detecting and characterizing RNA modifications comprehensively. As an alternative approach, Spangenberg et al. introduced Magnipore <https://github.com/JannesSP/Magnipore>, see Figure 9 a novel tool designed to identify significant signal shifts in ONT data obtained from samples of closely related or similar species. By comparing the ONT signals of two samples, Magnipore aims to uncover potential differential RNA modifications, providing insights into their occurrence and distribution.

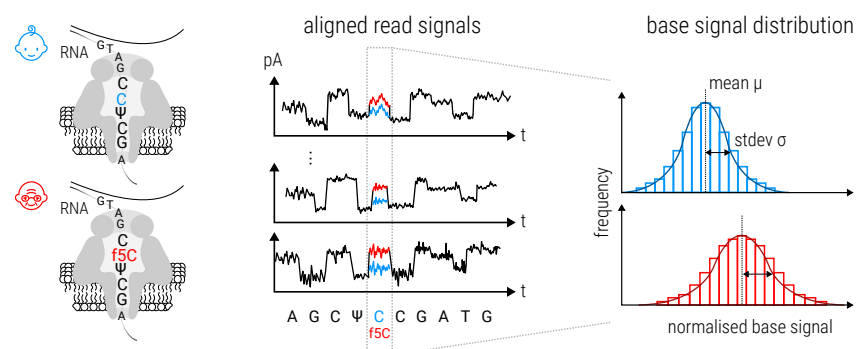


Figure 9. By analysing the raw signals of samples signal differences can be detected. Magnipore can read these differences and discriminate between mutations differences and possible differential RNA modifications.

In their study, the authors applied Magnipore to analyze 16 SARS-CoV-2 samples derived from human patients and cultivated in vero cells. These samples represent a diverse range of lineages, including the early 2020s Pango lineages (n = 6), B.1.1.7 (n = 2, Alpha), B.1.617.2 (n = 1, Delta), and B.1.1.529 (n = 7, Omicron). Through the utilization of position-wise Gaussian distribution models and employing a significance threshold, Magnipore effectively identifies differential ONT signals associated with mutations between the samples and potential RNA modifications.

Notably, when investigating the Alpha and Delta variants, Magnipore detects 55 mutations and identifies 15 sites that suggest the presence of differential modifications. Across all comparisons, Magnipore achieved a mutation detection rate of 89.1%. Furthermore, Magnipore unveiled promising potential differential modifications specific to virus variants and variant groups.

In summary, Magnipore represents an advancement in the analysis of RNA modifications within the context of viruses and viral variants. By utilizing Magnipore, researchers can uncover differential RNA modifications, providing critical insights into their role in cellular processes and their potential involvement in disease development.

2.9. Machine Learning in Viral Surveillance

Machine learning has emerged as a valuable tool in viral surveillance, enabling the analysis of large-scale genomic and epidemiological data. Computational tools utilizing machine learning algorithms can classify and predict viral strains, identify potential outbreaks, and assess the risk of viral transmission. These tools aid in the early detection and monitoring of viral diseases, enhancing public health surveillance efforts. Machine learning also helps uncover hidden patterns and relationships in viral data, providing insights into viral evolution, host tropism, and drug resistance. Leveraging machine learning in viral surveillance empowers researchers and public health agencies to make informed decisions for effective prevention and control strategies. The session was guided by Ingrida Olendraite (Cambridge, UK).

2.9.1. From High-Throughput Testing to Genomic Surveillance and Public Health Data Integration (by Bernhard Renard)

With its open-view approach and integration with high-throughput automation, genomic sequencing plays an increasing role in infectious disease diagnostics as well as in public health surveillance programs. Facilitated by algorithmic and machine learning approaches for signal processing and information aggregation, rapid sequencing procedures are arriving in clinical settings. Focus of the work of Renard et al. has been to facilitate data analysis already during run time of second and third generation sequencers in order to speed up diagnostics and decision making as well as enrichment of target sequences [96,97].

At the same time, the authors see an increase in genomic surveillance, which allows early detection of outbreaks and prediction of spreading patterns. Thereby, it can complement more traditional epidemiological approaches [98]. Renard et al. have introduced platforms for genomic surveillance to learn and predict movement and spreading patterns in population as well as for predicting genomic risk patterns [99–101]. In order to further leverage learning, Graph Neural Network allow integration and prediction across heterogeneous sources.

2.9.2. BLOODVIR: Virus Surveillance System for Plasma Pools Based on High-Throughput Sequencing and Machine Learning (by Martin Machyna)

Novel and re-emerging viruses pose a threat to the general public [102]. Many of these viruses are bloodborne and present an immediate risk to receivers of blood donations or blood-derived products. It is therefore vital to establish surveillance systems that are capable of detecting viruses in an unbiased manner before they can have a chance to spread in human population. Machyna et al. reported their efforts on development of BLOODVIR—a virus detection system for continuous monitoring of infection risks in blood plasma using high-throughput sequencing and machine learning. BLOODVIR uses a

combination of host depletion (nuclease treatment [103]) and viral enrichment (VirCapSeq-VERT [104]) to extract viral genomes and to create high-quality sequencing libraries from human blood plasma pools. Sequencing reads are classified with MiCoP [105] against a custom index created from a filtered Reference Viral Database (RVDB) [106] where genomes with very similar sequences (Mash [107] score < 0.15) were removed to reduce redundancy. The authors' evaluation of metagenomic classification tools for virus detection identified alignment-based classifiers as more suitable than k-mer-based. While k-mer approaches tend to be faster and have higher sensitivity under normal conditions, alignment-based methods such as MiCoP performed better under conditions when investigated sequences differed significantly from reference genome due to naturally occurring mutations. In order to improve their predictions of known and novel viruses from high-throughput sequencing data, Machyna et al. performed hyperparameter tuning of DeepMicrobes [108] deep neural network (DNN) model designed for metagenomic classification. They trained the model with HIV-1 B-subtype genome fragments that were in silico generated using a range of mutation rates. Evaluation with sequences generated from other HIV variants (A, C-L, N, O, P and U) revealed a dramatic improvement in the ability to detect these 'unseen' HIV variants by models trained with sequences created with $\geq 10\%$ mutation rate compared to model trained with non-mutated sequences. Machyna et al. tested BLOODVIR on plasma samples containing decreasing concentration of various viral standards and observed almost perfect linear relation between number of detected viral reads and virus concentration (average correlation coefficient = 0.958). In addition, their results indicate that BLOODVIR is capable of detecting as low as 100 viral genomes per mL of human blood plasma thus proving its usability for plasma pools with low viral titers. In conclusion, BLOODVIR surveillance system shows a great potential for monitoring of known and emerging viruses in human population.

2.9.3. Modelling the Zoonotic Capabilities of Avian Influenza via Genomic Machine Learning (by Liam Brierley)

Avian influenza is currently a high-risk threat in Europe. The 2021/22 outbreak was the largest yet observed and several countries stringently controlled domestic birds in response. Zoonotic capability has been observed for 14 subtypes, most recently H3N8 in 2021 [109]. Infections have so far been minimally transmissible between humans though continued zoonotic transmission events (e.g., in the UK, Spain, and Russia [110]) have increased concerns about the potential for new lineages to emerge that may spread more widely.

Several seminal modelling studies have demonstrated that machine learning algorithms can be trained directly on genome sequence data to make adequate predictions about which virus species may represent future zoonoses [101,111]. However, few models of zoonotic potential have addressed the wide variation between influenza A virus subtypes.

Brierley et al. used NCBI GenBank and GISAID to source over 18,000 whole genome sequences of avian influenza from 122 subtypes. To prevent over-fitting models to well-sampled lineages, they used MMSeqs2's Linclust algorithm [112] to collapse these to 3958 non-zoonotic clusters and 88 zoonotic clusters sharing $\geq 90\%$ sequence identity across $\geq 80\%$ mutual genome coverage. They selected random forest methods based on performance in previous efforts to classify influenza sequences [113].

Taking 4046 cluster-representative sequences as the training set, Brierley et al. trained random forests to distinguish zoonotic from non-zoonotic clusters based on genome composition of nucleotides, dinucleotides, and codons. Training procedures featured inner loops of 10-fold cross-validation nested within an outer loop of hold-one-out cross-validation applied to cluster representatives.

Random forest models distinguished zoonotic status with good separability (AUC = 0.95, F1 = 0.99), albeit with low absolute probabilities for some zoonotic sequences. Performance was sufficiently generalisable across subtypes, i.e., zoonotic cluster representatives predicted with the strongest confidence covered not only well-represented subtypes (e.g., H5N1, H7N9) but also poorly-represented subtypes (e.g., H10N8). The most informa-

tive model features were primarily biases in dinucleotide usage, particularly GC, AC, and CG dinucleotides.

Brierley et al. demonstrated genomic machine learning models can be tailored to identify which lineages of circulating avian influenza have potential to become zoonotic in future. These methods also have potential to highlighted weaker, unrecognised signals of viral adaptation associated with human infectivity across whole genomes. If trained appropriately, computational learning frameworks can inform more reactive strategies to prevent zoonotic infection by suggesting key genomic sites to monitor viral evolution in wildlife and generating risk estimates for newly identified viruses as soon as sequences are available [114].

2.10. Viral Pathogenesis

Jenna Kelly (Bern, Switzerland) hosted the session about viral pathogenesis, which refers to the mechanisms by which viruses cause disease in their hosts. Computational tools play a significant role in understanding viral pathogenesis by enabling the analysis of viral genomic data, protein structures, host-virus interactions, and immune responses. These tools help identify viral virulence factors, study the molecular basis of pathogenesis, and predict the impact of viral mutations on disease outcomes. Computational modeling and simulation assist in unraveling the complex dynamics of viral infections and aid in the development of targeted therapies and vaccines. The integration of computational tools with experimental approaches accelerates our understanding of viral pathogenesis and aids in effective disease management strategies.

Sex Differences in Respiratory Virus Infections (by Sebastian Beck)

Respiratory viruses, such as influenza A viruses or coronaviruses, remain the major causative agents of acute respiratory distress syndrome (ARDS) that is associated with high morbidity and mortality. Retrospective analyses of the SARS-CoV-2 pandemic revealed old age, underlying comorbidities (e.g., obesity or diabetes) and in particular male sex as high-risk factors for severe or even fatal COVID-19 [115]. Similarly, a male bias towards severe infections was also observed for avian H7N9 influenza that emerged in China in 2013 [116]. Factors that mediate sex disparity upon respiratory virus infection may include gender aspects (e.g., social behaviour), fixed genetic predispositions, or dynamic changes in sex hormones, such as testosterone and estradiol. Beck et al. have recently shown that H7N9 avian influenza hits the metabolic HPG (hypothalamic–pituitary–gonadal) axis in men but not in women, leading to a significant reduction in circulating testosterone levels, which is associated with fatal outcome, see Figure 10 [117]. Using a mouse model for influenza infection, the authors further demonstrated a causal link between viral infection and testosterone depletion. Interestingly, men suffering from severe COVID-19 disease also presented reduced testosterone levels, as reported by us and others [118]. Strikingly, long-term monitoring of COVID-19 patients who recovered from acute infection, revealed that up to 30% still show low testosterone levels one year after recovery [119]. Collectively, these data highlight the need to further study the effect of respiratory virus infections on the endocrine system in more detail. Herein, monitoring sex hormone levels as novel biomarkers for disease severity may be crucial for individualized patient therapy in the future.

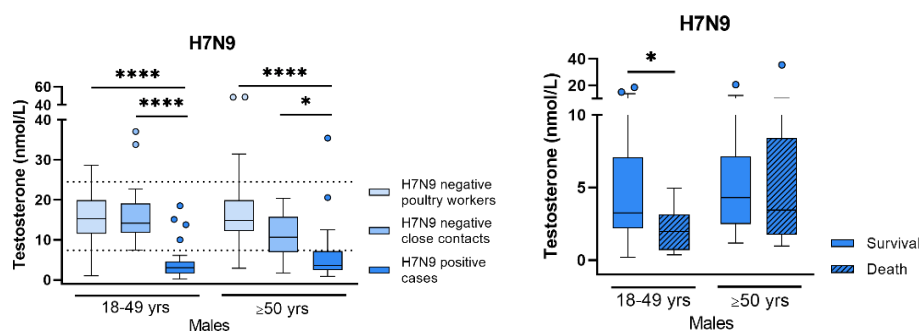


Figure 10. Low testosterone levels are a hallmark of severe and even fatal avian H7N9 influenza infection in men. *— $p < 0.05$; ****— $p < 0.0001$. Modified from [117].

2.11. Metagenomics for Identifying and Tracking Potential Zoonotic Viruses

Metagenomics has emerged as a powerful tool for identifying and tracking potential zoonotic viruses, which can jump from animals to humans. By sequencing genetic material from diverse environmental samples, such as animal reservoirs or their habitats, metagenomics allows for the detection of viral sequences that may be novel or closely related to known zoonotic viruses. This approach aids in surveillance efforts, enabling the early detection of emerging viruses and facilitating proactive public health interventions to mitigate potential zoonotic disease outbreaks. Metagenomics provides a comprehensive and unbiased approach to monitor viral diversity and identify potential sources of zoonotic transmission, ultimately contributing to global health security. Jelle Matthijssens (Leuven, Belgium) hosted this session.

Discovering and Tracking Potential Zoonotic Species from Metagenomic Samples with a Capture-Based Oriented Pipeline (by Maria Tarradas-Aleman)

From the dawn of Next Generation Sequencing (NGS) technologies, those strategies have become crucial in the study of microbial communities from environmental samples. However, there are still some challenges to overcome, either from biological and computational perspectives, to characterize their virome composition. Viral metagenomics has to deal with low quality sequences, possible sample biases (due to chemical inhibitors, degradation, etc.), challenging data analysis, and more specifically the lack of standardized regions for classification, the arduous purification of enough biomass for sequencing, and the limited completeness of the available virus databases. In addition, most of the viral particles found in environmental samples correspond to bacteriophages, which further complicates the detection of specific viral families and species [120].

The proposed approach to overcome some of those issues focuses on the use of capture probes specifically designed to hybridate a set of species of interest, with the aim to enrich the sample with their genomic sequences and similar ones [104]. For a specialized bioinformatic analysis of these datasets, Tarradas-Aleman et al. introduced CAPTVRED <https://github.com/MarTarAl> (Capture-based metagenomics Analysis Pipeline for tracking ViRal species from Environmental Datasets), a NextFlow [121] automated pipeline purposely designed to provide comprehensive results of capture-based metagenomics datasets. The pipeline includes a pre-filtering stage to discard non-viral sequences, taking advantage of a curated viral database, which also excludes phage viral sequences, as reference. Unlike other available protocols, CAPTVRED offers the flexibility to adjust almost any parameter at each step, making it adaptable to the unique characteristics of viral metagenomic datasets.

The virome present in a set of samples retrieved from sewage and bat guano have been already analyzed with this pipeline; moreover, sequences obtained by whole-genome shotgun and probe-based viral capture approaches have been also considered, in order to assess the performance of the capture kit, as well as for the pipeline. The results show an increased number of assigned viral contigs in the capture approach (using RVDB database),

which also recalls higher coverage and similarity with respect to reference sequences of potentially zoonotic viruses.

The paper's highlights include a strong emphasis on bioinformatics methods specifically designed for viruses. Most contributions of the conference address challenges in ultra-massive data analysis and an increasingly prominent role of artificial intelligence (AI) in the field. Nevertheless, the tools introduced in this publication underscore the pressing demand for further development of virus-specific bioinformatics tools, a need that has not been prominently addressed at other bioinformatics conferences. Conversely, the viruses featured in this conference have been thoroughly analyzed using specially tailored bioinformatic tools, reflecting a unique and vital synergy that does not typically find expression in virus-focused conferences.

Author Contributions: Conceptualization, F.H.; writing—original draft preparation, all authors; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the German Federal Ministry of Health (BMG): ZMI5-2521DAT400 (MarMac). We obtained funding by the EU: Marie Skłodowska-Curie ITN 955974 (VIROINF, YW, GLL, LS); NextGenerationEU/PRTR (SFE); HIVERA—249697 (ABA); by FCT through funds to GHTM-UID/Multi/04413/2013 and GHTM-UID/04413/2020 (ABA), MigrantHIV (PTDC/ DTP-EPI/7066/2014, ABA) and Integriv (PTDC/SAU- INF/31990/2017, ABA). We thank MICIN/AEI/10.13039/501100011033: SGL2021-03-009, SGL2021-03-052 (SFE). Funding was also received by the DFG: Microverse (EXC 2051—390713860, ManMar, CE), iDIV (FZT 118, KL), 448974291 (MN), CRC 1278 (CE), NFDI4microbiota (NFDI 28/1—460129525, ST, KL), as well as by BMBF: VirBio 01KI2106 (MN). The work was also supported by an award from The Pandemic Institute & CSL Seqirus Influenza Research Partnership (LB), the TAB (SARSrapid 2020-FGR-0052, CE), the LPI (CE), and the National Science Foundation (CAREER-2141862, AH). Further support was gained by JSPS KAKENHI (JP18H05506, JP21J22509, LN) and Russian Science Foundation (22-15-00230, YV).

Acknowledgments: The conference organizers thank all speakers, poster presenters, chairs and participants for creating a highly interactive scientific environment and making the meeting a success. The conference organisers acknowledge sponsorship from Promega Biotech Ibérica SL. We thank Milena Žarković, Ruben Rose, Christina Martínez Christophersen, Lea Paltzow, Mohsen M. Hegab, Anna Wansorra, Akash Srivastava, Andi Krumbholz, Pablo A. Gutiérrez, João M.F. Silva, Juan C. Muñoz-Sánchez, Maria J. Olmo-Uceda, José A. Oteo and Carlos P. Garay, Yanhua Xu, Dominik Wojtczak, Matthew Baylis, Joaquin Mould-Quevedo, Markus Braun, Gibran Rubio Quintanares, Pauline Santos, Erica Margiotta, Janice Brückmann, Csaba Miskey, Johannes Blümel, Renate König, Liam Childs, Frank O. Aylward, Alexander Lukashev, Artem Orlov, Celia Jakob, Hardin Bolte, Christian Höner zu Siederdisen, Martin Schwemmler.

Conflicts of Interest: The authors declare no conflict of interest. The sponsors had no role in the decision to publish this report nor in the selection process for oral presentations.

References

1. Ibrahim, B.; McMahon, D.P.; Hufsky, F.; Beer, M.; Deng, L.; Mercier, P.L.; Palmarini, M.; Thiel, V.; Marz, M. A new era of virus bioinformatics. *Virus Res.* **2018**, *251*, 86–90. [[CrossRef](#)] [[PubMed](#)]
2. Hufsky, F.; Ibrahim, B.; Beer, M.; Deng, L.; Mercier, P.L.; McMahon, D.P.; Palmarini, M.; Thiel, V.; Marz, M. Virologists-Heroes need weapons. *PLoS Pathog.* **2018**, *14*, e1006771. [[CrossRef](#)] [[PubMed](#)]
3. Hufsky, F.; Beslic, D.; Boeckaerts, D.; Duchene, S.; González-Tortuero, E.; Gruber, A.J.; Guo, J.; Jansen, D.; Juma, J.; Kongkitimanon, K.; et al. The International Virus Bioinformatics Meeting 2022. *Viruses* **2022**, *14*, 973. [[CrossRef](#)]
4. Hendrix, R.W.; Smith, M.C.; Burns, R.N.; Ford, M.E.; Hatfull, G.F. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2192–2197. [[CrossRef](#)] [[PubMed](#)]
5. Koskella, B.; Brockhurst, M.A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **2014**, *38*, 916–931. [[CrossRef](#)] [[PubMed](#)]
6. Brüssow, H.; Canchaya, C.; Hardt, W.D. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 560–602. [[CrossRef](#)] [[PubMed](#)]
7. Cobián Güemes, A.G.; Youle, M.; Cantú, V.A.; Felts, B.; Nulton, J.; Rohwer, F. Viruses as winners in the game of life. *Annu. Rev. Virol.* **2016**, *3*, 197–214. [[CrossRef](#)]

8. Howard-Varona, C.; Lindback, M.M.; Bastien, G.E.; Solonenko, N.; Zayed, A.A.; Jang, H.; Andreopoulos, B.; Brewer, H.M.; Glavina Del Rio, T.; Adkins, J.N.; et al. Phage-specific metabolic reprogramming of virocells. *ISME J.* **2020**, *14*, 881–895. [[CrossRef](#)]
9. Suttle, C.A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **2007**, *5*, 801–812. [[CrossRef](#)]
10. Li, M.; Wang, C.; Guo, Q.; Xu, C.; Xie, Z.; Tan, J.; Wu, S.; Wang, P.; Guo, J.; Fang, Z.; et al. More Positive or More Negative? Metagenomic Analysis Reveals Roles of Virome in Human Disease-Related Gut Microbiome. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 846063. [[CrossRef](#)]
11. Ren, J.; Song, K.; Deng, C.; Ahlgren, N.A.; Fuhrman, J.A.; Li, Y.; Xie, X.; Poplin, R.; Sun, F. Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **2020**, *8*, 64–77. [[CrossRef](#)] [[PubMed](#)]
12. Auslander, N.; Gussow, A.B.; Benler, S.; Wolf, Y.I.; Koonin, E.V. Seeker: Alignment-free identification of bacteriophage genomes by deep learning. *BioRxiv* **2020**. [[CrossRef](#)]
13. Guo, J.; Bolduc, B.; Zayed, A.A.; Varsani, A.; Dominguez-Huerta, G.; Delmont, T.O.; Pratama, Akbar Adje +and Gazitúa, M.C.; Vik, D.; Sullivan, M.B.; et al. VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **2021**, *9*, 37. [[CrossRef](#)] [[PubMed](#)]
14. Camargo, A.P.; Nayfach, S.; Chen, I.M.A.; Palaniappan, K.; Ratner, A.; Chu, K.; Ritter, S.J.; Reddy, T.B.K.; Mukherjee, S.; Schulz, F.; et al. IMG/VR v4: An expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **2023**, *51*, D733–D743. [[CrossRef](#)]
15. Wu, L.Y.; Pappas, N.; Wijesekara, Y.; Piedade, G.J.; Brussaard, C.P.; Dutilh, B.E. Benchmarking Bioinformatic Virus Identification Tools Using Real-World Metagenomic Data across Biomes. *BioRxiv* **2023**. [[CrossRef](#)]
16. Katz, K.; Shutov, O.; Lapoint, R.; Kimelman, M.; Brister, J.; O’Sullivan, C. The Sequence Read Archive: A decade more of explosive growth. *Nucleic Acids Res.* **2022**, *50*, D387–D390. [[CrossRef](#)]
17. Courtot, M.; Cherubin, L.; Faulconbridge, A.; Vaughan, D.; Green, M.; Richardson, D.; Harrison, P.; Whetzel, P.L.; Parkinson, H.; Burdett, T. BioSamples database: An updated sample metadata hub. *Nucleic Acids Res.* **2019**, *47*, D1172–D1178. [[CrossRef](#)]
18. The Cost of Sequencing a Human Genome. 2022. Available online: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> (accessed on 9 July 2023).
19. Edgar, R.C.; Taylor, J.; Lin, V.; Altman, T.; Barbera, P.; Meleshko, D.; Lohr, D.; Novakovsky, G.; Buchfink, B.; Al-Shayeb, B.; et al. Petabase-scale sequence alignment catalyses viral discovery. *Nature* **2022**, *602*, 142–147.
20. Wolf, Y.I.; Kazlauskas, D.; Iranzo, J.; Lucía-Sanz, A.; Kuhn, J.H.; Krupovic, M.; Dolja, V.V.; Koonin, E.V. Origins and evolution of the global RNA virome. *MBio* **2018**, *9*, e02329-18. [[CrossRef](#)]
21. Zayed, A.A.; Wainaina, J.M.; Dominguez-Huerta, G.; Pelletier, E.; Guo, J.; Mohssen, M.; Tian, F.; Pratama, A.A.; Bolduc, B.; Zablocki, O.; et al. Cryptic and abundant marine viruses at the evolutionary origins of Earth’s RNA virome. *Science* **2022**, *376*, 156–162. [[CrossRef](#)]
22. Sakaguchi, S.; Urayama, S.I.; Takaki, Y.; Hirotsuna, K.; Wu, H.; Suzuki, Y.; Nunoura, T.; Nakano, T.; Nakagawa, S. NeoRdRp: A Comprehensive Dataset for Identifying RNA-dependent RNA Polymerases of Various RNA Viruses from Metatranscriptomic Data. *Microbes Environ.* **2022**, *37*, ME22001. [[CrossRef](#)] [[PubMed](#)]
23. Aylward, F.O.; Moniruzzaman, M.; Ha, A.D.; Koonin, E.V. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol* **2021**, *19*, e3001430. [[CrossRef](#)] [[PubMed](#)]
24. Moniruzzaman, M.; Weinheimer, A.R.; Martinez-Gutierrez, C.A.; Aylward, F.O. Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* **2020**, *588*, 141–145. [[CrossRef](#)] [[PubMed](#)]
25. Ha, A.D.; Moniruzzaman, M.; Aylward, F.O. High Transcriptional Activity and Diverse Functional Repertoires of Hundreds of Giant Viruses in a Coastal Marine System. *mSystems* **2021**, *6*, e0029321. [[CrossRef](#)]
26. Moniruzzaman, M.; Martinez-Gutierrez, C.A.; Weinheimer, A.R.; Aylward, F.O. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* **2020**, *11*, 1710. [[CrossRef](#)]
27. Ha, A.D.; Moniruzzaman, M.; Aylward, F.O. Assessing the biogeography of marine giant viruses in four oceanic transects. *ISME Commun.* **2023**, *3*, 43. [[CrossRef](#)]
28. Chojnacki, J.; Staudt, T.; Glass, B.; Bingen, P.; Engelhardt, J.; Anders, M.; Schneider, J.; Iler, B.; Hell, S.W.; Usslich, H.G. Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* **2012**, *338*, 524–528. [[CrossRef](#)]
29. Chojnacki, J.; Waithe, D.; Carravilla, P.; Huarte, N.; Galiani, S.; Enderlein, J.; Eggeling, C. Envelope glycoprotein mobility on HIV-1 particles depends on the virus maturation state. *Nat. Commun.* **2017**, *8*, 545. [[CrossRef](#)]
30. Chojnacki, J.; Eggeling, C. Super-resolution fluorescence microscopy studies of human immunodeficiency virus. *Retrovirology* **2018**, *15*, 41. [[CrossRef](#)]
31. Chojnacki, J.; Eggeling, C. Zooming in on virus surface protein mobility. *Future Virol.* **2018**, *13*, 225–227. [[CrossRef](#)]
32. Gutierrez, P.A.; Elena, S.F. Single-cell RNA-sequencing data analysis reveals a highly correlated triphasic transcriptional response to SARS-CoV-2 infection. *Commun. Biol.* **2022**, *5*, 1302. [[CrossRef](#)] [[PubMed](#)]
33. Ravindra, N.G.; Alfajaro, M.M.; Gasque, V.; Huston, N.C.; Wan, H.; Szigeti-Buck, K.; Yasumoto, Y.; Greaney, A.M.; Habet, V.; Chow, R.D.; et al. Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. *PLoS Biol.* **2021**, *19*, e3001143. [[CrossRef](#)] [[PubMed](#)]
34. Triana, S.; Metz-Zumaran, C.; Ramirez, C.; Kee, C.; Doldan, P.; Shahraz, M.; Schraivogel, D.; Gschwind, A.R.; Sharma, A.K.; Steinmetz, L.M.; et al. Single-cell analyses reveal SARS-CoV-2 interference with intrinsic immune response in the human gut. *Mol. Syst. Biol.* **2021**, *17*, e10232. [[CrossRef](#)] [[PubMed](#)]

35. Burke, J.M.; St Clair, L.A.; Perera, R.; Parker, R. SARS-CoV-2 infection triggers widespread host mRNA decay leading to an mRNA export block. *RNA* **2021**, *27*, 1318–1329. [[CrossRef](#)] [[PubMed](#)]
36. Finkel, Y.; Gluck, A.; Nachshon, A.; Winkler, R.; Fisher, T.; Rozman, B.; Mizrahi, O.; Lubelsky, Y.; Zuckerman, B.; Slobodin, B.; et al. SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis. *Nature* **2021**, *594*, 240–245. [[CrossRef](#)]
37. Herzog, V.A.; Reichholf, B.; Neumann, T.; Rescheneder, P.; Bhat, P.; Burkard, T.R.; Wlotzka, W.; von Haeseler, A.; Zuber, J.; Ameres, S.L. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **2017**, *14*, 1198–1204. [[CrossRef](#)]
38. Schofield, J.A.; Duffy, E.E.; Kiefer, L.; Sullivan, M.C.; Simon, M.D. TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* **2018**, *15*, 221–225. [[CrossRef](#)]
39. Riml, C.; Amort, T.; Rieder, D.; Gasser, C.; Lusser, A.; Micura, R. Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 13479–13483. [[CrossRef](#)]
40. Jürges, C.; Dölken, L.; Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **2018**, *34*, i218–i226. [[CrossRef](#)]
41. Rummel, T.; Sakellaridi, L.; Erhard, F. grandR: A comprehensive package for nucleotide conversion RNA-seq data analysis. *Nat. Commun.* **2023**, *14*, 3559. [[CrossRef](#)]
42. Goldberg, A.V. An Efficient Implementation of a Scaling Minimum-Cost Flow Algorithm. *J. Algorithms* **1997**, *2*, 1–29. [[CrossRef](#)]
43. Nishimura, L.; Fujito, N.; Sugimoto, R.; Inoue, I. Detection of Ancient Viruses and Long-Term Viral Evolution. *Viruses* **2022**, *14*, 1336. [[CrossRef](#)] [[PubMed](#)]
44. Nishimura, L.; Sugimoto, R.; Inoue, J.; Nakaoka, H.; Kanzawa-Kiriyama, H.; Shinoda, K.I.; Inoue, I. Identification of ancient viruses from metagenomic data of the Jomon people. *J. Hum. Genet.* **2021**, *66*, 287–296. [[CrossRef](#)] [[PubMed](#)]
45. Biller, S.J.; Schubotz, F.; Roggensack, S.E.; Thompson, A.W.; Summons, R.E.; Chisholm, S.W. Bacterial vesicles in marine ecosystems. *Science* **2014**, *343*, 183–186. [[CrossRef](#)] [[PubMed](#)]
46. Linney, M.D.; Schvarcz, C.R.; Steward, G.F.; DeLong, E.F.; Karl, D.M. A method for characterizing dissolved DNA and its application to the North Pacific Subtropical Gyre. *Limnol. Oceanogr. Methods* **2021**, *19*, 210–221. [[CrossRef](#)]
47. Abecasis, A.B.; Geretti, A.M.; Albert, J.; Power, L.; Wait, M.; Vandamme, A.M. Science in court: The myth of HIV fingerprinting. *Lancet Infect. Dis.* **2011**, *11*, 78–79. [[CrossRef](#)]
48. Pineda-Peña, A.C.; Theys, K.; Stylianou, D.C.; Demetriades, I.; SPREAD/ESAR Program; Abecasis, A.B.; Kostrikis, L.G. HIV-1 infection in Cyprus, the Eastern Mediterranean European frontier: A densely sampled transmission dynamics analysis from 1986 to 2012. *Sci. Rep.* **2018**, *8*, 1702. [[CrossRef](#)]
49. Pimentel, V.F.; Pingarilho, M.; Sole, G.; Alves, D.; Miranda, M.; Diogo, I.; Fernandes, S.; Pineda-Pena, A.; Martins, M.R.O.; Camacho, R.; et al. Differential patterns of postmigration HIV-1 infection acquisition among Portuguese immigrants of different geographical origins. *AIDS* **2022**, *36*, 997–1005. [[CrossRef](#)]
50. Pingarilho, M.; Pimentel, V.; Miranda, M.N.S.; Silva, A.R.; Diniz, A.; Ascensão, B.B.; Piñeiro, C.; Koch, C.; Rodrigues, C.; Caldas, C.; et al. HIV-1-transmitted drug resistance and transmission clusters in newly diagnosed patients in Portugal between 2014 and 2019. *Front. Microbiol.* **2022**, *13*, 823208. [[CrossRef](#)]
51. Klitting, R.; Kafetzopoulou, L.E.; Thiery, W.; Dudas, G.; Gryseels, S.; Kotamarthi, A.; Vrancken, B.; Gangavarapu, K.; Momoh, M.; Sandi, J.D.; et al. Predicting the evolution of the Lassa virus endemic area and population at risk over the next decades. *Nat. Commun.* **2022**, *13*. [[CrossRef](#)]
52. Dellicour, S.; Baele, G.; Dudas, G.; Faria, N.R.; Pybus, O.G.; Suchard, M.A.; Rambaut, A.; Lemey, P. Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. *Nat. Commun.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
53. Dellicour, S.; Lequime, S.; Vrancken, B.; Gill, M.S.; Bastide, P.; Gangavarapu, K.; Matteson, N.L.; Tan, Y.; du Plessis, L.; Fisher, A.A.; et al. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nat. Commun.* **2020**, *11*. [[CrossRef](#)] [[PubMed](#)]
54. Tsukiyama-Kohara, K.; Iizuka, N.; Kohara, M.; Nomoto, A. Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.* **1992**, *66*, 1476–1483. [[CrossRef](#)] [[PubMed](#)]
55. Ng, W.; Soto-Acosta, R.; Bradrick, S.; Garcia-Blanco, M.; Ooi, E. The 5' and 3' Untranslated Regions of the Flaviviral Genome. *Viruses* **2017**, *9*, 137. [[CrossRef](#)] [[PubMed](#)]
56. Ochsenreiter, R.; Hofacker, I.; Wolfinger, M. Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses* **2019**, *11*, 298. [[CrossRef](#)]
57. Sola, I.; Mateos-Gomez, P.A.; Almazan, F.; Zuñiga, S.; Enjuanes, L. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol.* **2011**, *8*, 237–248. [[CrossRef](#)]
58. Yang, D.; Leibowitz, J.L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **2015**, *206*, 120–133. [[CrossRef](#)]
59. Madhugiri, R.; Karl, N.; Petersen, D.; Lamkiewicz, K.; Fricke, M.; Wend, U.; Scheuer, R.; Marz, M.; Ziebuhr, J. Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology* **2018**, *517*, 44–55. [[CrossRef](#)]
60. Dadonaite, B.; Gilbertson, B.; Knight, M.L.; Trifkovic, S.; Rockman, S.; Laederach, A.; Brown, L.E.; Fodor, E.; Bauer, D.L.V. The structure of the influenza A virus genome. *Nat. Microbiol.* **2019**, *4*, 1781–1789. [[CrossRef](#)]
61. Jones, J.E.; Le Sage, V.; Padovani, G.H.; Calderon, M.; Wright, E.S.; Lakdawala, S.S. Parallel evolution between genomic segments of seasonal human influenza viruses reveals RNA-RNA relationships. *eLife* **2021**, *10*, e66525. [[CrossRef](#)]

62. Jakob, C.; Lovate, G.L.; Desirò, D.; Gießler, L.; Smyth, R.; Marquet, R.; Lamkiewicz, K.; Marz, M.; Schwemmle, M.; Bolte, H. Sequential disruption of SPLASH-identified vRNA–vRNA interactions challenges their role in influenza A virus genome packaging. *Nucleic Acids Res.* **2023**, *51*, 6479–6494. [[CrossRef](#)] [[PubMed](#)]
63. Will, S.; Reiche, K.; Hofacker, I.L.; Stadler, P.F.; Backofen, R. Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering. *PLoS Comput. Biol.* **2007**, *3*, e65. [[CrossRef](#)] [[PubMed](#)]
64. Bach, S.; Demper, J.C.; Biedenkopf, N.; Becker, S.; Hartmann, R.K. RNA secondary structure at the transcription start site influences EBOV transcription initiation and replication in a length- and stability-dependent manner. *RNA Biol.* **2021**, *18*, 523–536. [[CrossRef](#)] [[PubMed](#)]
65. Mühlberger, E.; Trommer, S.; Funke, C.; Volchkov, V.; Klenk, H.D.; Becker, S. Termini of All mRNA Species of Marburg Virus: Sequence and Secondary Structure. *Virology* **1996**, *223*, 376–380. [[CrossRef](#)] [[PubMed](#)]
66. Worobey, M.; Holmes, E.C. Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **1999**, *80 Pt 10*, 2535–2543. [[CrossRef](#)] [[PubMed](#)]
67. Simmonds, P. Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. *J. Virol.* **2006**, *80*, 11124–11140. [[CrossRef](#)]
68. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **2015**, *1*, vev003. [[CrossRef](#)]
69. Vakulenko, Y.; Deviatkin, A.; Drexler, J.F.; Lukashev, A. Modular Evolution of Coronavirus Genomes. *Viruses* **2021**, *13*, 1270. [[CrossRef](#)]
70. Lukashev, A.N. Recombination among picornaviruses. *Rev. Med. Virol.* **2010**, *20*, 327–337. [[CrossRef](#)]
71. Desselberger, U. Other Than Noroviruses. *Viruses* **2019**, *11*, 286. [[CrossRef](#)]
72. Bosch, A.; Pintó, R.M.; Guix, S. Human astroviruses. *Clin. Microbiol. Rev.* **2014**, *27*, 1048–1074. [[CrossRef](#)] [[PubMed](#)]
73. Vakulenko, Y.A.; Orlov, A.V.; Lukashev, A.N. Patterns and Temporal Dynamics of Natural Recombination in Noroviruses. *Viruses* **2023**, *15*, 372. [[CrossRef](#)] [[PubMed](#)]
74. Lu, Z.; Zhang, Q.C.; Lee, B.; Flynn, R.A.; Smith, M.A.; Robinson, J.T.; Davidovich, C.; Gooding, A.R.; Goodrich, K.J.; Mattick, J.S.; et al. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **2016**, *165*, 1267–1279.
75. Aw, J.; Shen, Y.; Wilm, A.; Sun, M.; Lim, X.; Boon, K.L.; Tapsin, S.; Chan, Y.S.; Tan, C.P.; Sim, A.; et al. In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol. Cell* **2016**, *62*, 603–617. [[CrossRef](#)] [[PubMed](#)]
76. Sage, V.L.; Kanarek, J.P.; Snyder, D.J.; Cooper, V.S.; Lakdawala, S.S.; Lee, N. Mapping of Influenza Virus RNA–RNA Interactions Reveals a Flexible Network. *Cell Rep.* **2020**, *31*.
77. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
78. Aggarwal, R. Hepatitis E: The endemic perspective. *Clin. Liver Dis.* **2013**, *2*, 240–244. [[CrossRef](#)]
79. Shukla, P.; Nguyen, H.T.; Torian, U.; Engle, R.E.; Faulk, K.; Dalton, H.R.; Bendall, R.P.; Keane, F.E.; Purcell, R.H.; Emerson, S.U. Cross-species infections of cultured cells by hepatitis E virus and discovery of an infectious virus–host recombinant. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2438–2443. [[CrossRef](#)]
80. Nguyen, H.T.; Torian, U.; Faulk, K.; Mather, K.; Engle, R.E.; Thompson, E.; Bonkovsky, H.L.; Emerson, S.U. A naturally occurring human/hepatitis E recombinant virus predominates in serum but not in faeces of a chronic hepatitis E patient and has a growth advantage in cell culture. *J. Gen. Virol.* **2012**, *93*, 526–530. [[CrossRef](#)]
81. Lhomme, S.; Abravanel, F.; Dubois, M.; Sandres-Saune, K.; Mansuy, J.M.; Rostaing, L.; Kamar, N.; Izopet, J. Characterization of the polyproline region of the hepatitis E virus in immunocompromised patients. *J. Virol.* **2014**, *88*, 12017–12025. [[CrossRef](#)]
82. Biedermann, P.; Klink, P.; Nocke, M.K.; Papp, C.P.; Harms, D.; Kebelmann, M.; Thürmer, A.; Choi, M.; Altmann, B.; Todt, D.; et al. Insertions and deletions in the hypervariable region of the hepatitis E virus genome in individuals with acute and chronic infection. *Liver Int. Off. J. Int. Assoc. Study Liver* **2023**, *43*, 794–804. [[CrossRef](#)] [[PubMed](#)]
83. Gorris, M.; van der Lecq, B.M.; van Erpecum, K.J.; de Bruijne, J. Treatment for chronic hepatitis E virus infection: A systematic review and meta-analysis. *J. Viral Hepat.* **2021**, *28*, 454–463. [[CrossRef](#)] [[PubMed](#)]
84. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **2011**, *17*, 10. [[CrossRef](#)]
85. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
86. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)] [[PubMed](#)]
87. 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. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [[CrossRef](#)] [[PubMed](#)]
88. Steinegger, M.; Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [[CrossRef](#)]
89. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)]
90. Ontiveros, R.J.; Stoute, J.; Liu, K.F. The chemical diversity of RNA modifications. *Biochem. J.* **2019**, *476*, 1227–1245. [[CrossRef](#)]

91. Lichinchi, G.; Zhao, B.S.; Wu, Y.; Lu, Z.; Qin, Y.; He, C.; Rana, T.M. Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. *Cell Host Microbe* **2016**, *20*, 666–673. [[CrossRef](#)]
92. Brocard, M.; Ruggieri, A.; Locker, N. m6A RNA methylation, a new hallmark in virus-host interactions. *J. Gen. Virol.* **2017**, *98*, 2207–2214. [[CrossRef](#)] [[PubMed](#)]
93. Hao, H.; Hao, S.; Chen, H.; Chen, Z.; Zhang, Y.; Wang, J.; Wang, H.; Zhang, B.; Qiu, J.; Deng, F.; et al. N6-methyladenosine modification and METTL3 modulate enterovirus 71 replication. *Nucleic Acids Res.* **2018**, *47*, 362–374. [[CrossRef](#)] [[PubMed](#)]
94. Williams, G.D.; Gokhale, N.S.; Horner, S.M. Regulation of Viral Infection by the RNA Modification N6-Methyladenosine. *Annu. Rev. Virol.* **2019**, *6*, 235–253. [[CrossRef](#)] [[PubMed](#)]
95. Burgess, H.M.; Depledge, D.P.; Thompson, L.; Srinivas, K.P.; Grande, R.C.; Vink, E.I.; Abebe, J.S.; Blackaby, W.P.; Hendrick, A.; Albertella, M.R.; et al. Targeting the m6A RNA modification pathway blocks SARS-CoV-2 and HCoV-OC43 replication. *Genes Dev.* **2021**, *35*, 1005–1019. [[CrossRef](#)]
96. Ulrich, J.U.; Lutfi, A.; Rutzen, K.; Renard, B.Y. ReadBouncer: Precise and scalable adaptive sampling for nanopore sequencing. *Bioinformatics* **2022**, *38*, i153–i160. [[CrossRef](#)]
97. Tausch, S.H.; Strauch, B.; Andrusch, A.; Loka, T.P.; Lindner, M.S.; Nitsche, A.; Renard, B.Y. LiveKraken—Real-time metagenomic classification of illumina data. *Bioinformatics* **2018**, *34*, 3750–3752. [[CrossRef](#)]
98. Näher, A.F.; Vorisek, C.N.; Klopfenstein, S.A.; Lehne, M.; Thun, S.; Alsalamah, S.; Pujari, S.; Heider, D.; Ahrens, W.; Pigeot, I.; et al. Secondary data for global health digitalisation. *Lancet Digit. Health* **2023**, *5*, e93–e101. [[CrossRef](#)]
99. Wittig, A.; Miranda, F.; Hölzer, M.; Altenburg, T.; Bartoszewicz, J.M.; Beyvers, S.; Dieckmann, M.A.; Genske, U.; Giese, S.H.; Nowicka, M.; et al. CovRadar: Continuously tracking and filtering SARS-CoV-2 mutations for genomic surveillance. *Bioinformatics* **2022**, *38*, 4223–4225. [[CrossRef](#)]
100. Nasri, F.; Kongkitimanon, K.; Wittig, A.; Cortés, J.; Brinkmann, A.; Nitsche, A.; Schmachtenberg, A.; Renard, B.; Fuchs, S. MpoxRadar: A worldwide MPXV genomic surveillance dashboard. *Nucleic Acids Res.* **2023**, *51*, W331–W337. [[CrossRef](#)]
101. Bartoszewicz, J.M.; Seidel, A.; Renard, B.Y. Interpretable detection of novel human viruses from genome sequencing data. *NAR Genom. Bioinform.* **2021**, *3*, lqab004. [[CrossRef](#)]
102. Pérot, P.; Fourgeaud, J.; Rouzaud, C.; Regnault, B.; Rocha, N.D.; Fontaine, H.; Pavec, J.L.; Dolidon, S.; Garzaro, M.; Chrétien, D.; et al. Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France. *Emerg. Infect. Dis.* **2023**, *29*, 286–293. [[CrossRef](#)] [[PubMed](#)]
103. Lewandowska, D.W.; Zagordi, O.; Geissberger, F.D.; Kufner, V.; Schmutz, S.; Böni, J.; Metzner, K.J.; Trkola, A.; Huber, M. Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. *Microbiome* **2017**, *5*, 94. [[CrossRef](#)]
104. Briese, T.; Kapoor, A.; Mishra, N.; Jain, K.; Kumar, A.; Jabado, O.J.; Lipkin, W.I. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *mBio* **2015**, *6*, e01491-15. [[CrossRef](#)] [[PubMed](#)]
105. LaPierre, N.; Mangul, S.; Alser, M.; Mandric, I.; Wu, N.C.; Koslicki, D.; Eskin, E. MiCoP: Microbial community profiling method for detecting viral and fungal organisms in metagenomic samples. *BMC Genom.* **2019**, *20*, 423. [[CrossRef](#)] [[PubMed](#)]
106. Goodacre, N.; Aljanahi, A.; Nandakumar, S.; Mikailov, M.; Khan, A.S. A Reference Viral Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection. *mSphere* **2018**, *3*, e00069-18. [[CrossRef](#)] [[PubMed](#)]
107. Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **2016**, *17*, 132. [[CrossRef](#)]
108. Liang, Q.; Bible, P.W.; Liu, Y.; Zou, B.; Wei, L. DeepMicrobes: Taxonomic classification for metagenomics with deep learning. *NAR Genom. Bioinform.* **2020**, *2*, lqaa009. [[CrossRef](#)]
109. Sit, T.H.C.; Sun, W.; Tse, A.C.N.; Brackman, C.J.; Cheng, S.M.S.; Tang, A.W.Y.; Cheung, J.T.L.; Peiris, M.; Poon, L.L.M. Novel Zoonotic Avian Influenza A(H3N8) Virus in Chicken, Hong Kong, China. *Emerg. Infect. Dis.* **2022**, *28*, 2009–2015. [[CrossRef](#)]
110. Adlhoch, C.; Fusaro, A.; Gonzales, J.L.; Kuiken, T.; Marangon, S.; Mirinaviciute, G.; Niqueux, E.; Stahl, K.; Staubach, C.; Terregino, C.; et al. Avian influenza overview December 2022–March 2023. *EFSA J.* **2023**, *21*, e07917.
111. Mollentze, N.; Babayan, S.A.; Streicker, D.G. Identifying and prioritizing potential human-infecting viruses from their genome sequences. *PLoS Biol.* **2021**, *19*, e3001390. [[CrossRef](#)]
112. Steinegger, M.; Soding, J. Clustering huge protein sequence sets in linear time. *Nat. Commun.* **2018**, *9*, 2542. [[CrossRef](#)] [[PubMed](#)]
113. Xu, Y.; Wojtczak, D. Dive into machine learning algorithms for influenza virus host prediction with hemagglutinin sequences. *Biosystems* **2022**, *220*, 104740. [[CrossRef](#)]
114. Carlson, C.J.; Farrell, M.J.; Grange, Z.; Han, B.A.; Mollentze, N.; Phelan, A.L.; Rasmussen, A.L.; Albery, G.F.; Bett, B.; Brett-Major, D.M.; et al. The future of zoonotic risk prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2021**, *376*, 20200358. [[CrossRef](#)] [[PubMed](#)]
115. Zhang, J.J.; Dong, X.; Liu, G.H.; Gao, Y.D. Risk and protective factors for COVID-19 morbidity, severity, and mortality. *Clin. Rev. Allergy Immunol.* **2023**, *64*, 90–107. [[CrossRef](#)] [[PubMed](#)]
116. Wang, X.; Jiang, H.; Wu, P.; Uyeki, T.M.; Feng, L.; Lai, S.; Wang, L.; Huo, X.; Xu, K.; Chen, E.; et al. Epidemiology of avian influenza A H7N9 virus in human beings across five epidemics in mainland China, 2013–17: An epidemiological study of laboratory-confirmed case series. *Lancet Infect. Dis.* **2017**, *17*, 822–832. [[CrossRef](#)]
117. Bai, T.; Chen, Y.; Beck, S.; Stanelle-Bertram, S.; Mounogou, N.K.; Chen, T.; Dong, J.; Schneider, B.; Jia, T.; Yang, J.; et al. H7N9 avian influenza virus infection in men is associated with testosterone depletion. *Nat. Commun.* **2022**, *13*, 6936. [[CrossRef](#)]

118. Schroeder, M.; Schaumburg, B.; Mueller, Z.; Parplys, A.; Jarczak, D.; Roedl, K.; Nierhaus, A.; de Heer, G.; Grensemann, J.; Schneider, B.; et al. High estradiol and low testosterone levels are associated with critical illness in male but not in female COVID-19 patients: A retrospective cohort study. *Emerg. Microbes Infect.* **2021**, *10*, 1807–1818. [[CrossRef](#)]
119. Salonia, A.; Pontillo, M.; Capogrosso, P.; Pozzi, E.; Ferrara, A.M.; Cotelessa, A.; Belladelli, F.; Corsini, C.; Gregori, S.; Rowe, I.; et al. Testosterone in males with COVID-19: A 12-month cohort study. *Andrology* **2023**, *11*, 17–23. [[CrossRef](#)]
120. Garner, E.; Davis, B.C.; Milligan, E.; Blair, M.F.; Keenum, I.; Maile-Moskowitz, A.; Pan, J.; Gnegy, M.; Liguori, K.; Gupta, S.; et al. Next generation sequencing approaches to evaluate water and wastewater quality. *Water Res.* **2021**, *194*, 116907. [[CrossRef](#)]
121. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. NextFlow enables reproducible computational workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.