

Sequence analysis

ViMOP: a user-friendly and field-applicable pipeline for untargeted viral genome nanopore sequencing

Nils Peter Petersen^{1,2,*,†}, Mia Le^{1,2,3,†}, Annick Renevey^{1,2}, Ehizojie Emua⁴, Sarah Ryter^{1,2}, Giuditta Annibaldis^{1,2}, Jacob Camara⁵, Sanaba Boumbaly⁵, Cyril Erameh⁴, Tanja Laske^{3,6}, Jan Baumbach^{3,7}, Philippe Lemey⁸, Stephan Günther^{1,2}, Sophie Durauffour^{1,2,†}, Liana Eleni Kafetzopoulou^{1,2,8,†}

¹Bernhard Nocht Institute for Tropical Medicine (BNITM), 20359 Hamburg, Germany

²German Center for Infection Research (DZIF), Partner Site Hamburg–Lübeck–Borstel–Riems, 20251 Hamburg, Germany

³Institute for Computational Systems Biomedicine, University of Hamburg, 22761 Hamburg, Germany

⁴Viral and Emergent Pathogens Control and Research, Irrua Specialist Teaching Hospital, 310115 Irrua, Edo State, Nigeria

⁵Virology Research Center/Laboratory of Viral Hemorrhagic Fevers, 001 Conakry, Guinea

⁶Leibniz Institute of Virology, Viral Systems Modeling, 20251 Hamburg, Germany

⁷Department of Mathematics and Computer Science, Computational Biomedicine Lab, University of Southern Denmark, 5230 Odense, Denmark

⁸Department of Microbiology, Immunology and Transplantation, KU Leuven, Rega Institute, 3000 Leuven, Belgium

*Corresponding author. Bernhard Nocht Institute for Tropical Medicine (BNITM), Bernhard-Nocht-Straße 74, 20359 Hamburg, Germany

E-mail: nils.petersen@bnitm.de

† = equal contribution.

Associate Editor: Can Alkan

Abstract

Motivation: Untargeted, also known as metagenomic, nanopore sequencing is a powerful tool for virus genomic surveillance, particularly in resource-limited settings and when paired with the portability of the MinION device (Oxford Nanopore Technologies, ONT). However, a major bottleneck for global access is the absence of a user-friendly software capable of efficiently analyzing untargeted nanopore sequencing data to generate high-quality consensus genomes.

Results: We share ViMOP, a pipeline built on our long-term experience in nanopore field sequencing. The pipeline emphasizes field user-friendliness, flexibility and versatility to analyze reads generated directly from human clinical samples. The software assembles *de novo* contigs, matches contigs to known viral references and uses them to assemble consensus genomes. Executed with a single Nextflow command or via the EPI2ME Desktop interface (ONT), results are summarized in an HTML report. ViMOP, through its user-centered design, lowers the barrier to high-quality virus genome reconstruction and advances capacity for genomic surveillance.

Availability and implementation: ViMOP is freely available for non-commercial use (<https://github.com/opr-group-bnitm/vimop> and <https://zenodo.org/records/17913089>), along with the associated database (<https://zenodo.org/records/17652512>), the scripts used to generate it (<https://zenodo.org/records/17632662>) and benchmarking code (<https://zenodo.org/records/17633185>).

1 Introduction

Genomic surveillance has proven essential for identifying and responding to emerging viral disease outbreaks such as with mpox and COVID-19 (Attwood *et al.* 2022, Kinganda-Lusamaki *et al.* 2025). Rapid characterization of viral agents, alongside dense coverage of consensus genomes to monitor virus evolution, has substantially informed outbreak control efforts. This approach has facilitated the reconstruction of chains of transmissions, provided insights into the origin(s) of virus (re-)emergence, and enabled the monitoring of potential vaccine escape variants (Quick *et al.* 2016, Sissoko *et al.* 2017, Harvey *et al.* 2021, Keita *et al.* 2021, Kinganda-Lusamaki *et al.* 2021). Today, real-time genomic data sharing serves as a cornerstone for public health surveillance and evidence-based decision making.

Outbreaks often occur in endemic regions with limited resources (Kinganda-Lusamaki *et al.* 2021, Koundouno *et al.* 2022, Kinganda-Lusamaki *et al.* 2025). Global disparities persist in access to genomic surveillance infrastructure, and solutions to facilitate broad access and field applicability must be prioritized (Brito *et al.* 2022, Carter *et al.* 2022). Nanopore-based sequencing using the MinION device (Oxford nanopore technologies, ONT) offers a portable approach suited for such settings (Quick *et al.* 2016, Quick *et al.* 2017, Kafetzopoulou *et al.* 2019, Keita *et al.* 2021, de Vries *et al.* 2022). We previously demonstrated the feasibility of untargeted (metagenomic) nanopore sequencing and on-site analysis during outbreak emergencies in low- and middle-income countries (LMICs) (Kafetzopoulou *et al.* 2019, Keita *et al.* 2021, Koundouno *et al.* 2022).

Received: 17 June 2025; Revised: 12 December 2025; Accepted: 17 December 2025

© The Author(s) 2025. Published by Oxford University Press.

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

Recurring challenges in on-site field data analysis include the highly demanding technical setup in environments with unreliable power supply and internet access, as well as the need of advanced bioinformatics expertise. Operating bioinformatics tools from the command line requires expertise that may not be available when trained personnel are scarce. Also, factors such as field-sample quality (e.g. lack of cold chain) and sample type (e.g. blood, swab or seminal fluid) may lead to large fractions of host-reads, short reads or uneven read coverage, challenges that analysis pipelines should be equipped to handle. Ultimately, generating high-quality consensus genomes is a requirement for further downstream molecular analysis, and sharing with public health authorities (Carter et al. 2022).

Few tools support automated analysis of untargeted viral nanopore data (Table 1, available as [supplementary data](#) at *Bioinformatics* online). None fully satisfied our requirements for offline use, simple setup and operation by non-experts. Vir-MinION and VirPipe support viral detection but do not automate the transition to reference-based assembly, instead relying on *de novo* reconstruction that is often fragmented for untargeted sequencing data (Mastriani et al. 2022, Kim et al. 2023). VirDetector is limited when host reads, segmented viral genomes, or co-infections are present (Kaiser et al. 2025). INSaFLU-TELEVIR is the only option with a graphical interface, yet still requires command-line installation, separates detection and assembly, and depends on a large local database (Santos et al. 2024).

We thus developed ViMOP (Virus Metagenomics for Outbreaks Pipeline), a user-friendly, automated pipeline for virus detection and reference-guided genome assembly of untargeted sequencing reads from human clinical or animal samples. Integration into ONT's EPI2ME framework enables command-line-free operation, setup and updates. Alternatively, it runs with a single Nextflow command. Docker containers grant independence from the host operating system. ViMOP runs offline on a laptop (≥ 16 CPU cores, 30 GB RAM recommended) enabling flexible deployment.

2 Materials, methods, and results

2.1 Broad overview of the nanopore virus surveillance pipeline

ViMOP detects known viral species and generates consensus genomes from basecalled nanopore reads (FASTQ) (Fig. 1; see

Table 2, available as [supplementary data](#) at *Bioinformatics* online for tool choice rationales). Following separation of viral from non-viral reads, viral reads are assembled *de novo* into contigs. These are used to identify closely related reference genomes from a local database. Viral consensus sequences are subsequently generated via reference-guided assembly. All results, including consensus sequences and alignments, are output along with an HTML report.

2.2 Installation

ViMOP can be installed via EPI2ME Desktop by entering the GitHub repository URL. After installation, the backend database can be downloaded using the setup functionality. Alternatively, these steps can be done in a single Nextflow command.

2.3 Database

ViMOP uses a structured backend database and provides functionalities for automated download and updates. Users can create their own database. Our database is tailored to hosts and viral species relevant to our research. It contains (i) sequence files for background read removal, (ii) a virus genome set, and (iii) a Centrifuge index for read and contig classification. For background read removal, our database comprises the human genome and transcriptome, two rodent genomes and one mosquito genome, as well as a set of reagent-associated sequences including common bacterial laboratory strains and expression vectors (Tables 3 and 4, available as [supplementary data](#) at *Bioinformatics* online). The virus genome set is built based on all GenBank virus genomes (Benson et al. 2013). A curated subset of species relevant to our projects is maintained to ensure quality and consistency (Table 5, available as [supplementary data](#) at *Bioinformatics* online). Genomes with (a) $>1\%$ of unknown bases (N), (b) $>98\%$ sequence identity, or (c) $<80\%$ of a reference genome size, or not aligning to any reference genome (considered misassigned) were excluded. For SARS-CoV-2, a pre-selection from RVDB (Goodacre et al. 2018) was used. All remaining GenBank viral genomes are included without curation, and may require additional review upon detection. Groups comprising all genomes in our database for selected species or families were defined to filter for virus reads (Tables 5 and 6, available as [supplementary data](#) at *Bioinformatics* online). Our Centrifuge index combines our virus genome set with sequences (RefSeq) of human, mouse,

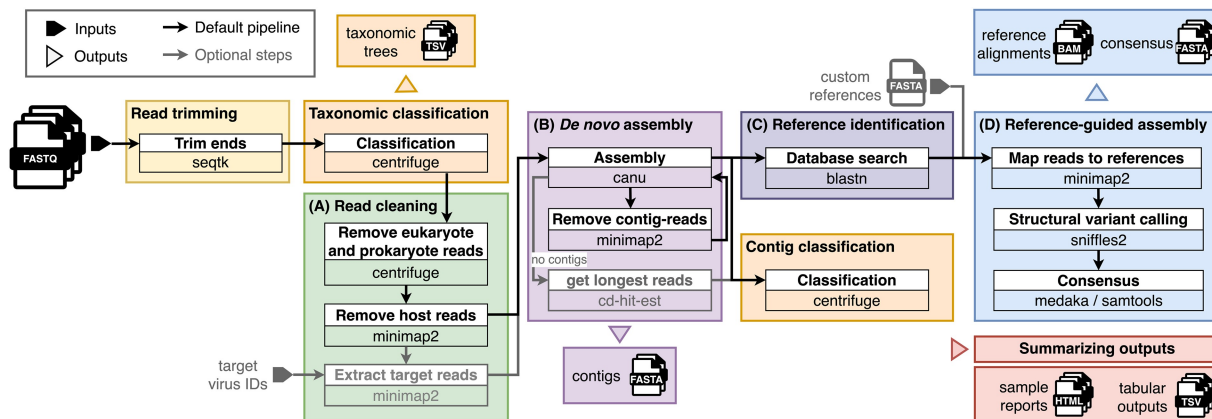


Figure 1. Schematic overview ViMOP. The workflow starts with cleaning and filtering the reads (A) which are then assembled into *de novo* contigs (B). Closely related virus genomes are identified (C) and used for a reference-guided assembly (D). Results are provided in an output directory with consensus sequences, alignments and TSV files and summarized in a comprehensive HTML report.

bacterial, and archaeal genomes (O’Leary *et al.* 2016). Database updates and extensions are implemented as needed or upon user request. In low-connectivity settings, users may restrict an update to individual components.

2.4 Read trimming

seqtk (Li, <https://github.com/lh3/seqtk>) trims 3’ and 5’ ends to remove primer sequences. The number of trimmed-bases is user-configurable.

2.5 Taxonomic classification and read cleaning

Non-viral reads are removed in three steps. First, the reads are taxonomically classified using Centrifuge (Kim *et al.* 2016). Reads classified as non-viral are removed, and an overview of the microbial content of the sample is displayed in the ViMOP report. Second, for thorough removal of non-viral reads, the remaining reads are mapped with minimap2 (Li 2018) against one or more sets of host genome sequences and reagent-associated sequences. The resulting output is called the cleaned read set. Third, ViMOP filters for viral reads. This is optional. The user can apply one or multiple filters in parallel to specifically map the reads against one or more virus species or family. All reads are mapped against target viral genomes using minimap2. The reads mapping to a viral reference are called target filtered reads and one read set is created for each filter. For both the set of cleaned reads, and each set of target-filtered reads, *de novo* assembly and reference genome searches are performed.

2.6 De novo assembly

Reads are assembled into contigs using Canu (Koren *et al.* 2017). An iterative approach was implemented to account for uneven genome abundance (e.g. in a co-infection, one viral species might be less abundant than the other), increasing the likelihood for broad viral species detection. At each iteration, reads mapping against any generated contigs are removed. The remaining reads are further used for another Canu run. This is done until a pre-defined maximum number of iterations is reached, or until no contig is created or no reads are left. Assembled contigs are then used for reference identification. A special routine is integrated in ViMOP for Canu runs of target filtered read sets that did not yield any contigs. In this case, reads are clustered using CD-HIT (Fu *et al.* 2012) and then used as queries, analogous to contigs.

2.7 Reference identification

Contigs are matched to reference genomes from the backend database using BLAST (Altschul *et al.* 1990). For each contig, the top hit is selected using the default BLAST ranking with ties being resolved by the database’s internal ordering. Each unique hit is then used for reference-guided assembly. Users can manually add additional reference sequences in a FASTA file.

2.8 Reference-guided assembly

ViMOP assembles a consensus sequence for each reference identified. All reads are mapped to the reference sequence and large insertions and deletions are detected with Sniffles2 (Smolka *et al.* 2024). These are integrated into a first draft genome using BCFTools (Danecek *et al.* 2021). All reads are mapped to this draft, and the final consensus is built using either Medaka (ONT, <https://github.com/nanoporetech/medaka>) or Samtools (Danecek *et al.* 2021) to call single-

nucleotide variants and small insertions and deletions. Per default, ViMOP uses Medaka if a model that fits the respective basecaller is available, and Samtools if not. In both cases, a minimum depth (default 20) is required to call bases. Samtools additionally requires a minimum proportion (70%) for a nucleotide to be called. Positions below the thresholds are masked with Ns. Samtools may alter the length of N-masked areas due to insertion or deletion artifacts from sparse reads. Therefore, a custom script replaces deletions that are not covered by the minimum depth of reads with Ns and removes insertions of Ns.

2.9 Contig classification

Contigs are classified using Centrifuge. This can help to identify contigs that were not assigned to a reference or are only partially aligned to the reference hit, e.g. if these are from bacterial origin. The Centrifuge classification is only used for reporting and not for the assembly of the reference target selection.

2.10 Output

ViMOP produces consensus sequences (FASTA), contigs (FASTA), alignment files (BAM), tables as tab-delimited files, and a separate HTML report for each sample. The user-friendly report, organized in three sections, provides information on the outcome of all the pipeline steps. The first section includes read distributions, read filtering statistics, and an overview of the Centrifuge read classification. The second section lists the viruses identified and statistics on genome recovery. The third section lists all *de novo* assembled contigs with the corresponding BLAST hits, and Centrifuge classification.

2.11 Benchmark

We compared ViMOP against INSaFLU-TELEVIR and VirDetector using simulated viral reads, mixed host–virus datasets, and real samples from the Sequence Read Archive (Supplementary Methods, Results, and Tables 7–12, available as supplementary data at Bioinformatics online). While VirDetector failed to automatically assemble many of the genomes, ViMOP and INSaFLU-TELEVIR show comparable results in most cases. Comparison on simulated data shows, that ViMOP is more robust when it comes to high mutation levels. Different to INSaFLU, ViMOP also recovers partial genomes below 50% completeness. On the downside, ViMOP has longer runtimes.

3 Conclusions and outlook

ViMOP simplifies untargeted nanopore sequencing data analysis. By integrating field-applicability, accessibility, maintainability and versatility, ViMOP enables broad usage for virus detection and genome assembly. The graphical interface, offline capability, and HTML reports make it particularly attractive for small teams focused on laboratory implementation. ViMOP integrates 8-years of experience of nanopore sequencing development in laboratories across Sub-Saharan Africa, and within the European Mobile Laboratory (<https://www.emlab.eu>). Integrated into our laboratory network, ViMOP will be continuously improved. ViMOP aims to strengthen global outbreak preparedness by supporting rapid, decentralized virus genomic surveillance where it is most urgently needed.

Acknowledgements

The authors thank the Ministry of Health of the Republic of Guinea and Nigeria, the healthcare workers and all those who contributed to diagnostics activities and in-country surveillance efforts. We thank the National Center for Biotechnology Information (NCBI) for providing GenBank and the Sequence Read Archive resources and all contributors for sharing their data. The BNITM acknowledges punctual technical assistance from Adam Koletis. We thank David O'Connor for sharing his reagent sequence data set.

Author contributions

Nils Peter Petersen (Conceptualization [equal], Data curation [equal], Formal analysis [lead], Investigation [equal], Methodology [lead], Project administration [equal], Software [lead], Supervision [equal], Validation [equal], Visualization [equal], Writing—original draft [lead], Writing—review & editing [equal]), Mia Le (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [lead], Validation [equal], Visualization [lead], Writing—original draft [equal]), Annick Renevey (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [equal], Software [lead], Validation [equal], Visualization [equal], Writing—review & editing [equal]), Ehizojie Emua (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [supporting], Software [supporting], Validation [equal]), Sarah Ryter (Formal analysis [equal], Investigation [equal], Validation [equal]), Giuditta Annibaldis (Investigation [equal], Project administration [equal], Supervision [equal]), Jacob Camara (Formal analysis [equal], Investigation [equal], Validation [equal]), Sanaba Boumbaly (Investigation [equal], Resources [equal]), Cyril Erameh (Investigation [equal], Resources [equal]), Tanja Laske (Resources [supporting], Supervision [equal], Writing—review & editing [supporting]), Jan Baumbach (Resources [supporting], Supervision [equal]), Philippe Lemey (Resources [equal], Supervision [equal]), Stephan Günther (Conceptualization [equal], Funding acquisition [equal], Resources [equal]), Sophie Duraffour (Conceptualization [equal], Funding acquisition [lead], Investigation [equal], Project administration [lead], Resources [equal], Supervision [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [lead]), and Liana Eleni Kafetzopoulou (Conceptualization [lead], Data curation [equal], Formal analysis [equal], Investigation [lead], Methodology [lead], Project administration [equal], Software [equal], Supervision [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [lead])

Supplementary data

[Supplementary data](#) is available at *Bioinformatics* online.

Conflict of interest: None declared.

Funding

The work was supported by the German Federal Ministry of Health through support of the WHO Collaborating Centre for Arboviruses and Hemorrhagic Fever Viruses at the Bernhard-Nocht-Institute for Tropical Medicine [agreement ZMV I1-2517WHO005], the Global Health Protection

Program [GHPP, agreements ZMV I1-2517GHP-704, ZMVI1-2519GHP704, and ZMI1-2521GHP921 until end of 2022, and from 2023 agreements ZMI5-2523GHP006 and ZMI5-2523GHP008] and the COVID-19 surge fund [BMG ZMVI1-2520COR001]. The work was supported by the European Union through the Civil Protection Pool Upgrade or Repair of Response Capacities [UCPM-2022-ECPP-URC, agreement ECHO/REP/INT/SUB/2022/883987/Adaptation/DE/EmLab] and the Research and Innovation Program of the European Union under H2020 grant agreement n°871029-EVA-GLOBAL. The BNITM is a member of the German Center for Infection Research (DZIF, partner site Hamburg–Lübeck–Borstel–Riems, Hamburg, Germany) and all works performed in this study have been supported by DZIF. P.L. acknowledges support from the Research Foundation—Flanders (“Fonds voor Wetenschappelijk Onderzoek—Vlaanderen”, G051322N and G005323N). L.E.K. acknowledges support from the Research Foundation—Flanders (“Fonds voor Wetenschappelijk Onderzoek—Vlaanderen”, 12X9222N and G005323N). This work was supported by the Federal Ministry of Education and Research (BMBF) and the Universität Hamburg with funds of the Excellence Strategy of the Federal Government and the Länder [grants to M.L. and T.L.]. This work was supported by the Leibniz ScienceCampus InterACT, funded by the BWFGB Hamburg and the Leibniz Association (W75/2022 InterACT and “Hamburg- X Infektionsforschung”) [to J.B.].

Data availability

All software and resources associated with ViMOP are available via GitHub and Zenodo as cited in the Availability and Implementation section of the abstract. Benchmarking data consisted of publicly available SRA datasets and simulated reads. Simulated reads can be reproduced and SRA datasets downloaded using the benchmarking scripts provided with the benchmarking package.

References

- Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- Attwood SW, Hill SC, Aanensen DM *et al.* Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nat Rev Genet* 2022;23:547–62.
- Benson DA, Cavanaugh M, Clark K *et al.* GenBank. *Nucleic Acids Res* 2013;41:D36–42.
- Brito AF, Semenova E, Dudas G *et al.*; Swiss SARS-CoV-2 Sequencing Consortium. Global disparities in SARS-CoV-2 genomic surveillance. *Nat Commun* 2022;13:7003.
- Carter LL, Yu MA, Sacks JA *et al.* Global genomic surveillance strategy for pathogens with pandemic and epidemic potential 2022–2032. *Bull World Health Organ* 2022;100:239–239A.
- Danecek P, Bonfield JK, Liddle J, *et al.* Twelve years of samtools and bcftools. *Gigascience* 2021;10:giab008
- de Vries EM, Cogan NOI, Gubala AJ *et al.* Rapid, in-field deployable, avian influenza virus haemagglutinin characterisation tool using MinION technology. *Sci Rep* 2022;12:11886.
- Fu L, Niu B, Zhu Z *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150–2.
- Goodacre N, Aljanahi A, Nandakumar S, *et al.* A reference viral database (RVDB) TO enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *mSphere* 2018; 3:00069–18.

- Harvey WT, Carabelli AM, Jackson B *et al.*; COVID-19 Genomics UK (COG-UK) Consortium. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;**19**:409–24.
- Kafetzopoulou LE, Pullan ST, Lemey P *et al.* Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* 2019;**363**:74–7.
- Kaiser NL, Groschup MH, Sadeghi B. VirDetector: A bioinformatic pipeline for virus surveillance using nanopore sequencing. *Bioinformatics* 2025;**41**:btaf029
- Keita AK, Koundouno FR, Faye M *et al.* Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks. *Nature* 2021;**597**:539–43.
- Kim D, Song L, Breitwieser FP *et al.* Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res* 2016;**26**:1721–9.
- Kim K, Park K, Lee S, *et al.* VirPipe: An easy-TO-use and customizable pipeline for detecting viral genomes from nanopore sequencing. *Bioinformatics* 2023;**39**:btad293
- Kinganda-Lusamaki E, Amuri-Aziza A, Fernandez-Nuñez N *et al.* Clade I mpox virus genomic diversity in the democratic republic of the Congo, 2018–2024: predominance of zoonotic transmission. *Cell* 2025;**188**:4–14 e16.
- Kinganda-Lusamaki E, Black A, Mukadi DB *et al.* Integration of genomic sequencing into the response to the Ebola virus outbreak in Nord Kivu, Democratic Republic of the Congo. *Nat Med* 2021;**27**:710–6.
- Koren S, Walenz BP, Berlin K *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017;**27**:722–36.
- Koundouno FR, Kafetzopoulou LE, Faye M *et al.* Detection of marburg virus disease in Guinea. *N Engl J Med* 2022;**386**:2528–30.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.
- Mastriani E, Bienes KM, Wong G *et al.* PIMGAVir and vir-minion: Two viral metagenomic pipelines for complete baseline analysis of 2nd and 3rd generation data. *Viruses* 2022;**14**:1260. <https://doi.org/10.3390/v14061260>.
- O’Leary NA, Wright MW, Brister JR *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;**44**:D733–45.
- Quick J, Grubaugh ND, Pullan ST *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;**12**:1261–76.
- Quick J, Loman NJ, Duraffour S *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;**530**:228–32.
- Santos JD, Sobral D, Pinheiro M *et al.*; TELEVIR Consortium. INSaFLU-TELEVIR: an open web-based bioinformatics suite for viral metagenomic detection and routine genomic surveillance. *Genome Med* 2024;**16**:61.
- Sissoko D, Duraffour S, Kerber R *et al.* Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study. *Lancet Glob Health* 2017;**5**:e80–e88.
- Smolka M, Paulin LF, Grochowski CM *et al.* Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* 2024;**42**:1571–80.