

Enriched Long-Read Sequencing of Co-circulating Viruses in Complex Samples

Mariana Meneguzzi ^{1,†}, Jonathan Bravo ^{2,†}, Tara N. Gaire ³, Peter M. Ferm ¹,
Montserrat Torremorell ¹, Christina Boucher ^{2,†}, Noelle R. Noyes ^{1,†*}

¹Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, 1333 Gortner Avenue, St. Paul, MN 55108, USA

²Department of Computer & Information Science & Engineering, University of Florida, 432 Newell Drive, Gainesville, FL 32611, USA

³School of Veterinary Medicine, Texas Tech University, 7671 Evans Drive, Amarillo, TX 79119, USA

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: nnoyes@umn.edu.

Associate editor: Ana Carolina Junqueira

Abstract

At present, no single workflow is available for quick and accurate identification and analysis of genomes of various viruses present together in a field or clinical sample, particularly when followed by long-read sequencing. Our work addressed this limitation by combining targeted enrichment with long-read, real-time sequencing. Using a panel of probes targeting 16,069 complete viral genomes, we validated this workflow (termed TELSVirus) on complex sample matrices collected from pigs and compared its performance to traditional methods including real-time reverse transcription polymerase chain reaction and shotgun metagenomics. Using serial dilutions of samples with known viral status, we observed that TELSVirus generated viral reads for dilutions up to 10^{-9} . TELSVirus was able to detect viral targets when shotgun metagenomic long- and short-read datasets did not and when rRT-PCR results were undetermined. Finally, we performed TELSVirus on 144 oral fluid samples collected in the field, which are highly complex and diverse samples used for viral surveillance in swine. We identified a high prevalence of relatively understudied viruses, often found co-circulating with better-characterized viruses. In many cases, TELSVirus generated ultra-deep genome coverage, allowing for further genomic epidemiological investigations, although bioinformatic methods need further development to work robustly with TELSVirus data. Our results support using TELSVirus for rapid detection and genomic characterization of multiple low-abundance viruses from single samples using long-read sequencing.

Keywords: long-read, viruses, probe-capture, target enrichment, minion

Introduction

Accurate, high-resolution co-sequencing of multiple viral genomes from complex samples has long been considered the holy grail of viral surveillance (Lipkin and Firth 2013; McGinnis et al. 2016; Lalonde et al. 2020; Rockett et al. 2022). To date, available methods all have one (or more) shortcomings (Houldcroft et al. 2017). Traditionally, polymerase chain reaction (PCR)-based assays have been considered the “gold standard” for surveillance applications; however, PCR requires prior knowledge of the viruses to be targeted, limiting the detection of novel and fast-evolving viral variants (Edwards and Gibbs 1994; Gaudin and Desnues 2018). Metagenomic sequencing is capable of comprehensively sequencing almost every genome within a sample but faces certain limitations, particularly in its ability to detect low-abundance targets (Frey and Bishop-Lilly 2015). This issue is most noticeable when dealing with samples from host organisms or environmental sources where the desired microbial or viral ribonucleic or deoxyribonucleic acid (DNA/RNA) is only a small fraction of the total genetic content. In these cases, the abundance of host or non-target genetic material can significantly overwhelm the genomes of interest, making it challenging to detect them, especially when they are present at very

low levels (Steward et al. 2013; Dávila-Ramos et al. 2019). This inability to detect rare or low-abundance organisms within metagenomic data severely limits our understanding of microbial ecosystems and constrains our ability to monitor pathogens and identify emerging infectious disease threats.

Efforts to enhance the detection of rare genomic sequences amidst a backdrop of predominantly host or environmental DNA can involve virus isolation and/or PCR amplification of specific viral targets before conducting whole-genome sequencing (WGS) of the virus (Taylor et al. 2020; Tulloch et al. 2021). Although useful, these techniques can alter virus population structures and may not be well-suited for the discovery and surveillance of emerging and re-emerging viruses (Lee et al. 2013; Sari et al. 2019). Additionally, culture- and PCR-based assays are tailored for specific viruses, restricting each workflow to the detection of one or maybe several viruses. This limits the investigation of co-infections and restricts viral genomic surveillance to a handful of targets.

Probe-based assays can overcome these limitations by enabling the capture of diverse viral genomes within total extracted nucleic acids, followed by sequencing and genomic reconstruction. This approach has been successfully used for virome analysis using short-read sequencing platforms from human

Received: February 11, 2025. Revised: June 30, 2025. Accepted: September 5, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

clinical samples targeting several human viruses (Briese et al. 2015; Metsky et al. 2019; Paskey et al. 2019; Rehn et al. 2021; Wylezich et al. 2021; Kapel et al. 2023; Ceballos-Garzon et al. 2024). These previous studies investigated only a small number of samples with known viral status and used probe sets that targeted only a few human viruses. Very few studies directly compared probe-based assays to more common surveillance methods such as PCR-based methods. Additionally, these previous datasets were generated from paired-end short-read sequencing platforms, which make it difficult to reconstruct full-length genomes with phased variant analysis.

Long-read sequencing can overcome these inherent limitations of short-read data, but the use of long-read sequencing with probe-capture workflows is not well described. Very recently, two studies reported the successful use of SARS-CoV-2 probe sets to detect and reconstruct coronavirus genomes from human clinical samples (Nieuwenhuijse et al. 2022; Pogka et al. 2022). Aside from these studies, no studies were identified that combined probe-based capture with long-read sequencing for viral detection. Moreover, to our knowledge, co-sequencing of multiple probe-captured viruses from clinical or environmental samples using long-read sequencing has not yet been reported. Achieving this type of sequencing could significantly advance viral surveillance, because long sequences support more robust genomic characterization, particularly regarding variant phasing and genome reconstruction (Chaisson et al. 2019; Zaragoza-Solas et al. 2022). Therefore, a critical need remains to more extensively investigate the utility and efficacy of probe-based capture techniques in combination with long-read sequencing, particularly for more complex clinical samples with unknown status, and using probe sets that target a greater number of viruses.

To address this need, we designed a multi-virus probe set and combined its use with long-read sequencing on the Oxford Nanopore Technologies (ONT) platform, in an integrated workflow called “TELSVirus” (Target-Enriched Long-read Sequencing of Virus) (see Fig. 1). We validated TELSvirus using several benchmarking methods, including serial dilutions of known virus from different field-relevant sample matrices, followed by direct comparison to standard methods such as real-time reverse transcription PCR (rRT-PCR). To further assess performance in a more challenging scenario, we then used TELSvirus on complex samples likely to harbor multiple co-circulating viruses. Through this set of experiments, we demonstrate TELSvirus’ capacity to simultaneously sequence multiple co-circulating viruses from complex population-level samples used for routine surveillance.

Material and Methods

Probe Design

The probe set used in this work was designed to encompass a list of 44 swine viruses identified by experts in swine epidemiology and virology (see Table S1). The criteria for virus selection were based on their economic importance to the swine industry (e.g. porcine reproductive and respiratory syndrome virus [PRRSV], rotavirus, circovirus, porcine epidemic diarrhea virus [PEDV], and senecavirus), as well as their zoonotic and emergent potential (e.g. influenza viruses, rabies, astrovirus). Additionally, we included in the probe set genomes from viruses that are not currently circulating in commercial US swine herds and thus were highly unlikely to be in any of the samples analyzed in this study. These viruses were included to aid in the assessment of TELSvirus’ potential for false-

positive detections, and encompassed swine herpesvirus, nipah virus, Ebola Reston virus, and chikungunya virus. None of these viruses were detected in any of the samples analyzed using TELSvirus. All full-length genomes for each virus available from NCBI (www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239&sort=taxonomy) were downloaded in June 2021. Briefly, the respective family of each virus was selected, and the “Taxonomy” option was chosen. The virus “Family” name was then selected, and the “View and Analyze sequences in NCBI Virus” option was chosen, in which virus sequences were filtered for Nucleotide Completeness (RefSeq Genome Completeness == “Complete”). Subsequently, all complete genomes were selected and downloaded. For influenza A, B, and C, nucleotide sequences were downloaded from the NCBI influenza virus resource (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=genomeset>), selecting only genomes that identified the host as swine. These sequences were used as input for probe design via the Syotti algorithm (Alanko et al. 2022) with the following parameters: length of probe = 120 bp; number of potential mismatches between probe and target = 30 bp; and coverage of target genomes = 100%. The mismatch parameter value was selected based on manufacturer recommendations. The final probe set comprised 16,069 unique probes which together covered 100% of each of the input viral genomes. Probes were manufactured by Agilent Technologies Inc.

Sample Selection

Five nasal swabs collected from weaned pigs in a commercial swine farm as part of a previous study (Lopez-Moreno et al. 2022) were used to compare enriched vs unenriched libraries for influenza A virus (IAV). Three of these samples were considered positive for IAV and two were considered negative, based on results from previous testing of the samples using an rRT-PCR protocol targeting the conserved matrix gene (Slomka et al. 2010). RNA was extracted from each of the five samples, followed by complementary DNA (cDNA) synthesis and library preparation (for details, see RNA Extraction and cDNA Synthesis). cDNA from these samples was subjected to probe capture and long-read sequencing (see TELSvirus Workflow), and the subsequent libraries here on were termed “TELSvirus libraries.” A separate aliquot of cDNA from the same samples was also directly long-read sequenced and termed “untargeted libraries.” The only difference between the TELSvirus libraries and untargeted libraries was the use of the probe capture and enrichment process for the TELSvirus libraries.

Three nasal swabs and three serum samples obtained from two different biobanks were also used to investigate TELSvirus’ limit of detection by comparing the results generated by TELSvirus to traditional rRT-PCR Ct values and to non-enriched long-read sequencing. The nasal swabs were collected from weaned pigs on commercial swine farms as part of previous work (Lopez-Moreno et al. 2022) and had previously tested positive for IAV via rRT-PCR targeting the conserved IAV matrix gene (Slomka et al. 2010). The serum samples were obtained from pigs experimentally infected with PRRSV L1C 1-4-4 as part of a previous study and had been confirmed to be positive for PRRSV via rRT-PCR using previously published primers (Nirmala et al. 2021). RNA from the raw samples was extracted (see RNA Extraction and cDNA Synthesis) and subjected to 10-fold serial dilutions by serially adding 100 µL of the original sample suspension to 900 µL of

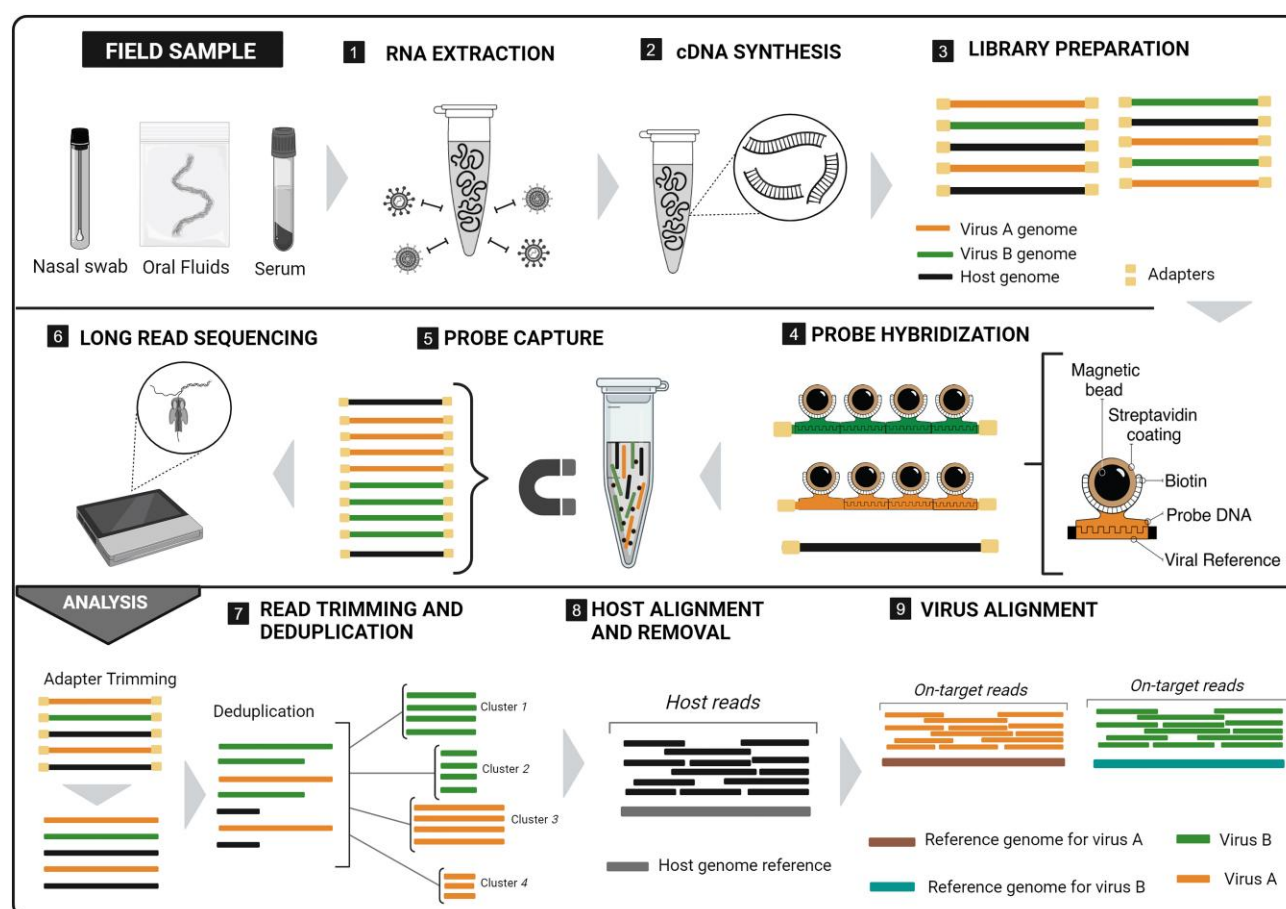


Fig. 1. TELSVirus overview. Utilizing field samples, RNA was extracted (1), and cDNA synthesis was performed (2). Then, library preparation was executed (3) followed by the hybridization process in which custom-designed biotinylated 120-mer probes (4) were used to capture the targeted viruses with streptavidin-coated magnetic beads (5). Captured fragments were amplified and submitted for sequencing using the minION platform from Oxford Nanopore Technologies (6). Resulting TELSVirus reads were trimmed for adapters and deduplicated (7). Subsequently, reads that aligned to the host genome (i.e. *S. scrofa*) were removed (8). The remaining reads were then aligned to the viral reference genomes, termed on-target reads (9). Reads that did not align to either the host (*S. scrofa*) genome or to any genomes in the viral reference database, were referred to as unmapped reads. Figure was adapted from [Slizovskiy et al. \(2022\)](#), with permission from the copyright holder.

Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific). Virus-specific rRT-PCR was performed four times for each dilution (technical replicates), targeting IAV in the nasal swabs using the previously described primers and rRT-PCR conditions ([Slomka et al. 2010](#)) and targeting PRRSV in the serum samples using the previously referenced primers and rRT-PCR conditions ([Nirmala et al. 2021](#)). Then, cDNA was synthesized from each dilution (see [RNA Extraction and cDNA Synthesis](#)), and each dilution was submitted to TELSVirus (see TELSVirus Workflow) and to non-enriched long-read sequencing (i.e. standard shotgun metagenomic sequencing). An average of the TELSVirus and non-enriched metagenomic read counts and rRT-PCR Ct values for each replicate was obtained and compared among each dilution for nasal swabs and serum samples. Additionally, a standard curve was generated for both IAV and PRRSV using serial dilutions of RNA transcripts with known viral copy numbers. This curve was used to convert Ct values into estimated viral copy numbers per dilution. The efficiency and R^2 value of each standard curve were assessed to ensure accurate quantification.

To assess the performance of TELSVirus on complex samples with unknown viral status, we used oral fluid (OF) samples collected from 63 wean-to-finish (WTF) herds located in

the Midwestern United States between 2017 and 2018, which were recently described ([Angulo et al. 2023](#)). OF samples are collected by hanging ropes within pens of pigs; the pigs chew on the ropes over a period of time, and then the ropes are collected and the fluid is manually extracted ([Prickett, et al. 2008](#)). In this way, OF samples represent a composite sample from a population of individual hosts, and they are used by swine veterinarians for routine viral surveillance due to the convenient collection ([Henao-Diaz et al. 2020](#)). For validating TELSVirus, we targeted OF samples that had been collected when PRRSV infection was at its highest incidence. Within this set of samples, we randomly selected 144 samples for analysis. The samples in this set had been previously tested for PRRSV by rRT-PCR ([Nirmala et al. 2021](#)).

RNA Extraction and cDNA Synthesis

RNA extraction for all the selected samples (serum, nasal swabs, and OF samples) was conducted using the MagMAX™ Total RNA Isolation Kit (Thermo Fisher Scientific AM1836), following the manufacturer's protocol (MagMAX™-96 Viral RNA Isolation Kit, n.d.). Quantitative measurements of total extracted RNA were

performed using Qubit 4 Fluorometric Quantification (Invitrogen, Thermo Fisher Scientific, Massachusetts, United States). After RNA quantification, cDNA synthesis was performed using the template-switching reverse transcription (RT) enzyme mix (NEB M0466) and the primeSTAR GLX DNA polymerase (Takara R050A), previously described (Schroeder et al. 2022). Briefly, 4 μ L of RNA, random primer, and dNTP were incubated for 5 min at 70 °C in a thermocycler, and then combined with the template-switching RT buffer, enzyme mix, and oligo, and then again incubated in a thermocycler for 90 min at 42 °C and 5 min at 85 °C. Lastly, cDNA amplification occurred using the primeSTAR GLX DNA polymerase, 5X primeSTAR GLX buffer, and dNTP mixture following thermocycler conditions of 98 °C for 1 min, 98 °C for 10 s, and 68 °C for 2.5 min for 30 cycles. All extraction and cDNA steps were performed in a separate location with a dedicated clean hood with unidirectional sample flow. Separate sets of pipettes, gowns, and filter tips were used for each step.

Negative Controls

During extraction of the nasal swab, serum, and OF samples, five extraction blanks consisting of elution buffer only were placed in different positions of each 96-well plate. For the cDNA synthesis step, an RNA extraction blank (consisting of reagents only) and a no-template control (NTC, consisting of water only) were included in each plate. All extraction blanks and NTCs were measured using the Qubit 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Massachusetts, United States), and all were measured as undetectable.

Short-Read Non-enriched Metagenomic Sequencing

One OF sample was sequenced on the AVITI™ System (Element Biosciences) using one lane of a 2 × 150 CB Freestyle Med kit (2 × 150 bp, dual 8 bp index) and sequenced for 32 h. Basecalling was performed with AVITI OS v3.1.0 and demultiplexing with Element Bioscience's bases2fastq software version 2.0.0. Raw FASTQ files were quality-assessed with FastQC (Brown et al. 2017) and trimmed with Cutadapt (Martin 2011). Reads aligning to the *Sus scrofa* reference genome were removed and the remaining non-host reads were aligned to the viral database using Burrows–Wheeler Alignment tool (BWA) (Li and Durbin 2009).

TELSVirus Workflow (Target-Enriched Long-Read Sequencing)

Target Hybridization and Enrichment

Following cDNA synthesis, the SureSelect XT HS2 DNA System (Agilent Technologies, Santa Clara, California, United States) protocol was carried out per manufacturer's protocol, except for the fragmentation step which was not performed. A total of 200 ng of DNA in 50 μ L was used to initiate the repair and dA-tailing of the cDNA ends, followed by ligation of the SureSelect XT HS2 Adaptor Oligo. Then, samples were purified using AMPure XP beads (Agencourt Biosciences Corp., Beverly, Massachusetts, United States), amplified by PCR, and purified again. The quality of pre-capture libraries was verified using an Agilent TapeStation 4200 (Agilent, Santa Clara, California, United States).

Probe hybridization was performed on amplified cDNA libraries with the addition of custom-designed biotinylated

probes to each library, along with 25% RNase block solution and hybridization buffer. The custom-designed viral probes were added at the concentration recommended by Agilent Technologies (normalized during manufacturing) and used without further dilution, as per manufacturer instructions. Incubation proceeded at 95 °C for 5 min, 65 °C for 10 min, 65 °C for 1 min, and 37 °C for 3 s for 60 cycles. Subsequent capture was performed by placing the samples on a plate mixer (1200 rpm) for 30 min at room temperature with MyONE streptavidin T1 beads (Invitrogen Co, Waltham, Massachusetts, United States). Once the capture incubation period concluded, beads were collected and washed using reagents and established temperatures and durations, per manufacturer's protocol.

For post-capture processing, the entire final wash elution volume (25 μ L) was used for amplification, followed by AMPure XP bead (Agencourt Biosciences Corp., Beverly, Massachusetts, United States) purification. All libraries were verified using Agilent TapeStation 4200 (Agilent, Santa Clara, California, United States), and a final volume of 23 μ L from each sample was used for subsequent nanopore library preparation.

Nanopore Library Preparation

After enrichment, library preparation was performed in batches of 12 samples, using the PCR barcoding kit SQK-PB004 for all samples except the non-enriched metagenomic dilution series, which were processed using the SQK-LSK114 kit. Library preparation proceeded as per manufacturer's instructions except for the fragmentation step, which was not performed. End-prep of DNA molecules was carried out using the NEBNext Ultra II End repair/dA-tailing Module (New England Biolabs Inc., Ipswich, United States), with incubation at 20 °C for 5 min followed by 65 °C for 5 min. Samples were purified using AMPure XP beads (Agencourt Biosciences Corp., Beverly, Massachusetts, United States) on a magnetic separation rack. Adapters were ligated to each sample using Blunt/TA Ligase Master Mix (New England Biolabs Inc., Ipswich, United States) following another purification using AMPure XP beads (Agencourt Biosciences Corp., Beverly, Massachusetts, United States). The concentration of each sample was measured by Qubit 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Massachusetts, United States), and a final concentration of 0.2 ng/ μ L in a 50 μ L volume was taken forward to amplification and a third purification using AMPure XP beads. Each sample was eluted in a final volume of 10 μ L and quantified using the Qubit 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Massachusetts, United States). For each sample, 5 μ L of the final library was combined into pools of 12 samples. Pools were subjected to concentration using Vacufuge Plus concentrator (Eppendorf 620-187566), to reach a final pooled volume of 10 μ L for subsequent sequencing.

MinION Sequencing and Basecalling

Pooled, barcoded libraries were sequenced on a new FLO-MIN106D flow cell with R9 sequencing chemistry, except for the non-enriched metagenomic sequencing of the dilution series libraries, which were sequenced using R10 chemistry due to discontinuation of R9 chemistry. Prior to sequencing, each flow cell was checked to ensure that >1,000 active pores were available for sequencing. Flow cells were primed and loaded into an ONT MinION Mk1C or GridION device and each run was allowed to proceed for

22 h. Sequencing runs were initiated and monitored using MinKNOW software, v20.10.3. Reads were basecalled in real time using the fast basecalling model in Guppy, with the exception of the non-enriched metagenomic dilution series libraries, which were basecalled using the high accuracy mode in Dorado.

Bioinformatic Analysis

Quality Filtering and Deduplication

The demultiplexed FASTQ files were merged by sample and then the barcode sequences were removed using a custom script that takes as input FASTA file that listed all barcode sequences, and a specified crop length which indicates the minimum number of bases to be trimmed from both ends of each read. First, the reads were scanned for exact barcode matches, and then the position of the barcode was identified as being either at the beginning, end, or middle of the read. Reads with barcodes in the middle were deemed chimeric and were split into two segments (read_A and read_B) at the barcode site. For barcodes located at the start or end, the trim length was adjusted to account for the entire region. Subsequently, any reads shorter than 50 base pairs post-barcode removal were discarded.

Trimmed and filtered reads were then subjected to deduplication, which is necessary for target-enriched data because the enrichment process involves multiple rounds of PCR amplification, which can create technical duplicates that appear as multiple reads in the downstream sequence data (Noyes et al. 2017). Briefly, reads with the same length were first sorted into separated individual bins. Bins were then merged to create clusters of reads with <10% difference in length. Then, all reads from one cluster were pairwise aligned with the Blast-Like Alignment Tool (BLAT) (Kent 2002). Reads were considered duplicated if the span of all the hit/query High-scoring Segment Pairs (HSPs) were greater than or equal to 90% of the total hit/query length. Sets of duplicate reads were accumulated, and deduplicated FASTQ files were generated by randomly retaining a single read for each duplicated set from the original library FASTQ. For these analyses, the pysam Python module was used to parse Sequence Alignment Map (SAM) files, and the SeqIO and SearchIO modules from Biopython were used to parse FASTA/FASTQ files and PSL files, respectively.

Identification of On-Target, Host, and Unmapped Reads

After trimming and deduplication, the remaining reads were aligned to the reference genome of *S. scrofa*, the domestic pig, using the Minimap2 aligner (Kent 2002). The alignment process was configured to exclude secondary alignments from the SAM file and to preserve the CIGAR string within the SAM file for subsequent analysis, using parameters `-secondary=no -a`. All other Minimap2 parameters were set to their default. Reads that aligned to the *S. scrofa* genome were considered and reported as “host reads.” Next, the remaining (non-host) reads were aligned to the viral reference database, i.e. the database of 58,221 accessions that were sourced from NCBI for probe design. To reduce spurious alignments, we removed polyA tails from reference genomes by removing from the end of genomes all contiguous strings of “A’s” longer than 10. In addition, two genomes contained a polyA tail with a single non-A base within the tail; these regions were removed manually from the fasta file. Finally, eight genomes contained a long stretch of C’s in the head of

the genome; these regions were also removed manually. Alignment to the modified reference fasta file was done using Minimap2 with parameter settings to enforce that no secondary alignments were included in the SAM file, and that the query minimizer discarding was disabled (`-secondary=no -q-occ-frac 0 -f 0 -a`). Lastly, any supplementary alignments were removed to avoid spurious duplicate counts. After this filtering, any reads that aligned to the viral database were considered to be “on-target.” Reads that did not align to either the *S. scrofa* reference genome or the viral database were considered “unmapped.” The status of each targeted virus in each sample was defined as “detected” (positive) if at least one read aligned to any accession in the viral reference database for the given virus; if there were no alignments, the status was defined as “not detected” (negative).

Calculation of Genome Fraction and Mean Coverage Depth

Non-host reads were aligned to the viral reference database for a second time. This alignment used Minimap2 with parameters that enforced secondary alignment inclusion in the output SAM file and query minimizer discarding was disabled (i.e. `-q-occ-frac 0 -f 0 -a`) to facilitate selection of viral variants. The genome fraction for each accession was calculated by dividing the total number of aligned bases by the total bases in the accession (i.e. the size of the genome). The mean coverage depth for each accession was calculated by summing the number of reads aligned at each position in the accession and dividing by the total number of positions in the accession (i.e. the size of the genome). Both genome fraction and mean coverage depth were computed across all viral accessions and all samples.

Taxonomic Assignment of Viruses

Alignment and coverage statistics were performed at the accession level. However, viral richness and diversity results were not reported at the accession level due to extensive sequence homology across accessions, particularly within virus types, which can lead to extensive inflation of diversity counts. To circumvent this problem, we instead reported richness and diversity values at the virus level. To accomplish this, we grouped accessions according to virus name and selected the accession with the highest genome fraction to represent the virus. In this way, we were able to describe viral richness and diversity at the level of virus type, thus circumventing inter-genome homology within virus type.

Genome Assembly and Haplotyping

De novo assembly of TELSVirus reads was attempted with the long-read assembler metaFlye (Kolmogorov et al. 2020), but assemblies were not successful due to the tool’s input read length requirement of 1,000 bp. Haplotyping was attempted using RVHaplo and Strainline (Cai and Sun 2022; Luo et al. 2022), but the former tool did not scale with the on-target read depth generated by TELSVirus, and the latter tool had a longer input read requirement than the read lengths generated by TELSVirus.

Results

Probe Set Description

The probe panel used in this study covered 44 viruses with importance for the swine industry. The final probe set consisted of 19,136 unique probes, which together covered 100% of

each of the 16,069 input genomes. After applying the manufacturer-recommended boosting process for GC content, the final probe set contained 59,952 probes. The number of viruses with complete genomes varied widely across each virus, with a low of just one for each of porcine sapovirus, sendai virus, and porcine adenovirus B to a high of 5,995 for influenza virus (Table S1).

TELSVirus Successfully Enriched Target Viruses

To assess the efficacy of TELSvirus compared to non-enriched metagenomic sequencing, we used nasal swabs with known IAV status based on rRT-PCR results. Non-enriched metagenomic sequencing of IAV-positive samples generated only 0.2% to 0.7% of total reads that aligned to the IAV genome (i.e. “on-target reads”). By contrast, TELSvirus libraries generated from IAV-positive samples contained a relatively high proportion of IAV-aligned reads, ranging from 59% to 88%, demonstrating enrichment efficacy (Table 1; Table S2). In addition, TELSvirus libraries demonstrated higher genome fraction and coverage depth than untargeted libraries (Fig. 2; Table S3).

For the samples that were IAV-negative based on rRT-PCR, both TELSvirus and non-enriched metagenomic libraries contained a small number of reads that mapped to the influenza genome (Table 1). If the rRT-PCR results were accurate and the samples were truly devoid of IAV nucleic acids, then these reads could represent false positives due to cross-contamination of the positive and negative samples during sample preparation. However, extraction and NTC blanks processed alongside the samples all yielded undetectable nucleic acids, suggesting minimal cross-contamination. Alternatively, the rRT-PCR results may have been false negatives, for example, if the IAV nucleic acid abundance was below the limitation of detection of the rRT-PCR assay, or if the rRT-PCR target was not present but other segments of the genome were. Sequence homology between IAV and other viruses co-present in the samples could have also led to false-positive results from untargeted and TELSvirus libraries; however, in such a scenario, one would expect high coverage depth in only the very small homologous section of the IAV genome. The coverage profile of the IAV genomes instead showed multiple regions of very low coverage across multiple IAV segments (Fig. 2), indicating that inter-virus homology was likely not the cause of the alignments.

Performance of TELSvirus Across Dilution Series Highlights Its Capacity to Detect Low-Abundance Viruses

To estimate the limit of detection of TELSvirus compared to rRT-PCR, we performed serial dilutions of IAV- and PRRSV-positive samples together with a standard curve for each virus. Based on this analysis, TELSvirus was able to generate IAV and PRRSV reads for all dilutions when run in triplicate, whereas the rRT-PCR results were undetermined for dilutions with mean Ct values greater than 40 and estimated viral load less than $3.81\text{E}+03$ (Fig. 3). Additionally, TELSvirus produced robust on-target rates (i.e. >60%) for dilutions up to 4 for IAV and 6 for PRRSV. For both experiments, these dilutions corresponded to a Ct value of approximately 30 and estimated viral load of $3.34\text{E}+08$ to $1.06\text{E}+03$. The TELSvirus on-target values for higher dilutions were more variable and typically less than 50%. Despite this lower on-target proportion, the sequencing depth afforded by TELSvirus generated robust genome coverage, even with a high percentage of host and

unmapped reads. Even for dilutions 7-9, coverage depth of PRRSV often exceeded 100 reads, while coverage depth for dilutions 1-3 was consistently higher than 50,000 reads (see Fig. 4 and Table 2).

When the same dilution series samples were subjected to long-read non-enriched metagenomic sequencing, we observed on-target rates ranging from 0% to 86% for IAV and 0% to 70% for PRRSV (Fig. 3). Across the 26 non-enriched IAV dilution libraries, 7 contained 0 alignments to any targeted viruses (i.e. 0% on-target), and 9 contained <1% on-target reads. Across the 26 non-enriched PRRSV dilution libraries, 3 yielded a 0% on-target rate and 10 yielded <1% on-target reads. For both IAV and PRRSV samples, a significant reduction in on-target reads occurred after dilution 3, corresponding to a Ct between 22 and 29 and estimated viral load of $1.59\text{E}+05$ to $1.43\text{E}+08$ (Tables S5 and S6). Additionally, the ability to recover IAV and PRRSV reads from non-enriched metagenomic libraries was highly variable even across replicates within each dilution.

In addition to the nasal swabs and serum samples, one of the OF samples was subjected to short-read non-enriched metagenomic sequencing on the AVITI™ System (Element Biosciences). A total of 16,054,942 reads were generated, 8,269,358 of which were retained after host removal (51.5%). From the non-host reads, 1388 reads (0.0086%) aligned to the viral database reference. In comparison, the same sample when submitted to TELSvirus generated a total of 385,090 on-target reads out of 429,883 total reads (89.58%). From the short-read, non-enriched sample, we identified 10 unique viruses. In the same sample subjected to TELSvirus, we identified 23 unique viruses.

Additionally, the OF samples were tested by rRT-PCR for PRRSV. For this comparison, rRT-PCR results were considered positive for PRRSV when the Ct value was below or equal to 42. TELSvirus results were considered positive for PRRSV if at least one read aligned to any of the PRRSV genomes within the viral reference database. Based on this analysis, 35 of the 144 OF samples were considered positive by both TELSvirus and rRT-PCR testing, and 32 samples were considered negative by both methods. Conversely, 77 of the 144 samples had discordant results when comparing TELSvirus and rRT-PCR. Of these, 6 were considered positive by TELSvirus due to the generation of reads that aligned to at least one PRRSV reference genome, but were considered negative by rRT-PCR. Finally, 71 samples were considered negative by TELSvirus because none of the generated reads aligned to any of the reference PRRSV genomes, whereas the rRT-PCR Ct values indicated the presence of the virus.

TELSvirus Supports Detection and Characterization of Multiple Viruses From Complex Field Samples

To extensively validate the TELSvirus workflow and to investigate its ability to detect and characterize multiple viruses from complex field samples, 144 OF samples were subjected to the TELSvirus workflow and sequenced across 13 sequence runs. The percentage of on-target reads per sample ranged from 0.44% to 95.01%, with a median of 60.71%. Additionally, we observed very low levels of host-aligned reads (range 0.005% to 6.17%) and highly variable percentages of unmapped reads (range 3.95% to 99.55%) across the samples. These results highlighted the success of the assay in enriching targeted viruses while depleting host nucleic acids within a complex sample matrix.

Table 1 IAV mapping statistics, comparing TELSVirus and untargeted libraries, for nasal swabs with known IAV status (based on rRT-PCR testing) collected from weaned pigs at commercial swine farms

Sample ID	rRT-PCR status	TELSVirus libraries				Untargeted libraries			
		Input reads (N)	On-target reads (%)	Host reads (%)	Unmapped reads (%)	Input reads (N)	On-target reads (%)	Host reads (%)	Unmapped reads (%)
1	Positive (Ct 18.63)	629347	85.12	11.30	3.58	58799	0.24	74.95	24.82
2	Positive (Ct 17.97)	527171	87.77	8.46	3.76	79757	0.70	72.88	26.43
3	Positive (Ct 18.99)	38826	58.75	21.34	19.91	100205	0.22	79.32	20.46
4	Negative (Ct undetermined)	11146	0.29	50.02	49.69	127200	0.11	79.61	20.28
5	Negative (Ct undetermined)	10243	0.15	70.99	28.87	60339	0.22	68.74	31.03

Mapping statistics encompass alignments to all eight IAV segments.

Using the on-target viral reads, we were able to detect multiple viruses in OF samples (Table 3). Across 144 samples, a total of 65 distinct viruses and viral segments were identified via alignment to the viral reference database. The mean richness per sample was 21.3 (range 4 to 44). Porcine astrovirus 2 and 4, porcine bocavirus, and porcine sapelovirus 1 were the most common viruses, identified in >95% of OF samples. Typically, detection of these viruses was accompanied by a high genome fraction and high coverage depth, although there was a wide range of values with some outlier samples (Table 3). These results demonstrate the robust nature of the genomic data generated by TELSVirus, particularly for highly prevalent and abundant viruses. A number of viruses achieved moderately high genome fraction (>60% to 100%) with sample prevalence ranging between 45% and 95%, including segments of influenza A, atypical porcine pestivirus, porcine astrovirus 3, PRRSV, rotavirus, and porcine epidemic diarrhoea virus. The remaining detected viruses were identified in a smaller number of samples and with less coverage, including porcine delta coronavirus, bovine circovirus, and porcine adenovirus 1 (Table S4).

Discussion

In this work, we demonstrated that the TELSVirus workflow enriched for targeted viruses within metagenomic nucleic acids extracted from diverse sample types. Our findings reinforce previous studies that demonstrate the efficacy of biotinylated probes for viral target enrichment (Briese et al. 2015; Lee et al. 2017; Paskey et al. 2019; Wylezich et al. 2021; Schuele et al. 2022). However, prior research was primarily limited to short-read sequencing and focused on a small number of viruses. Our results expand the use of enrichment by demonstrating its continued efficacy when used with a panel of 44 distinct viruses and when combined with long-read sequencing.

The TELSVirus enrichment allowed for deep sequence coverage of targeted genomes, which in many cases exceeded 1,000× across most of the genome (Figs. 2 and 4). This type of data is especially useful for reconstructing haplotypes, detecting recombination events, and resolving complex genome architecture. However, current bioinformatic assembly and haplotyping tools are not suited to the unique characteristics of the TELSVirus data, namely, the ultra-deep coverage and medium-length reads. Specifically, the N50 produced by

TELSVirus was approximately 500 bp, which fell below the minimum requirement of 1,000 bp for tools such as *MetaFlye* (Kolmogorov et al. 2020). Short-read assemblers and single-genome long-read assemblers were also unsuitable. For haplotyping, we tested RVHaplo and Strainline (Cai and Sun 2022; Luo et al. 2022), but both tools were unable to handle the large volume of on-target reads combined with the medium-sized read lengths. Thus, the development of tailored bioinformatic tools will be crucial for fully utilizing TELSVirus data for whole genome analysis and haplotyping, which is especially important for viruses that are difficult to culture, including PRRSV (Lalonde et al. 2020). Additionally, TELSVirus read lengths could potentially be increased through optimized molecular methods; however, in many cases, the sample quality is the limiting factor in obtaining high molecular weight extractions, especially for field samples.

Standard rRT-PCR methods are important benchmarks for viral detection since these methods are often robustly validated and highly sensitive. The TELSVirus workflow demonstrated good performance across serial dilutions where rRT-PCR Ct values were high and viral copies low. This was observed most readily in lower-complexity samples such as serum (Fig. 3). Even at higher Ct values (i.e. >30 and “undetermined”), we observed consistently high TELSVirus on-target rates with high horizontal and vertical read coverage (Fig. 4 and Table 2). In contrast, long-read shotgun metagenomic data from the same dilution series yielded highly variable results, with some libraries failing to detect the targeted virus at loads > 1.00E+03 copies per µL. Our findings contrast with those of a recent study that used shotgun metagenomic short-read sequencing on PRRSV-positive samples positive and failed to recover complete PRRSV genomes when corresponding rRT-PCR values exceeded 30 (Vandenbussche et al. 2021). Moreover, our study used a relatively shallow sequencing depth by multiplexing 8 to 12 samples on a single ONT R9 flow cell. By reducing the number of samples per flow cell, TELSVirus would very likely achieve even higher coverage statistics, including for samples with very low viral copy number.

An essential aspect of the TELSVirus workflow is the capability to pause or extend the sequencing run of the ONT device “on-the-fly.” This feature allows users to adjust the sequencing duration based on the desired level of genome coverage (Lin et al. 2022). This flexibility is a significant advantage of the TELSVirus assay. Additionally, TELSVirus’

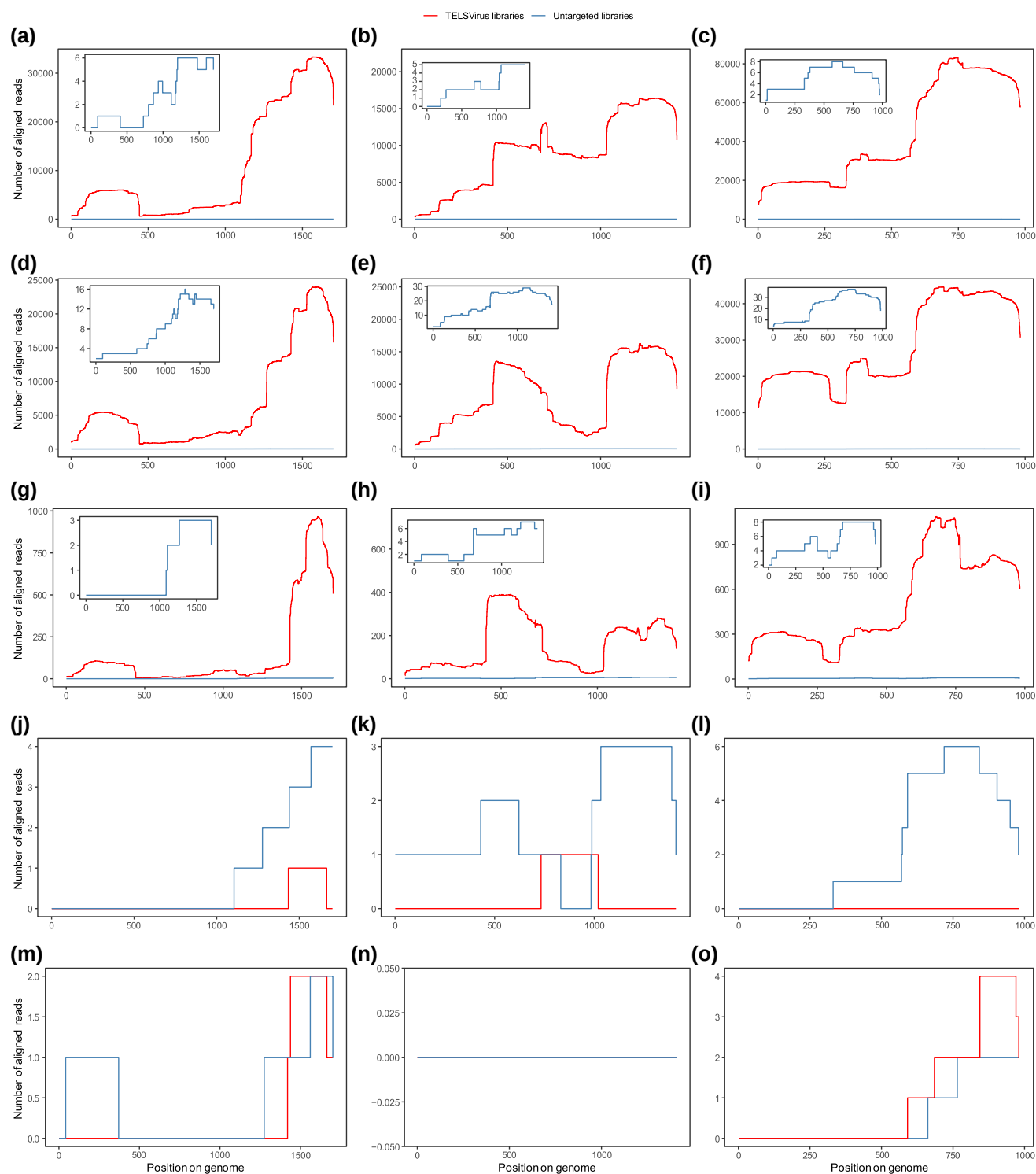


Fig. 2. Coverage profiles for IAV, comparing coverage generated by TELSVirus libraries (red lines) vs untargeted libraries (blue lines). The x axis of each panel represents genome position, and the y axis depicts the number of aligned reads. Coverage profiles for untargeted libraries are shown as a subplot within each plot so that the low coverage can be visualized on the subplot y axes. Each column represents a segment, with segment 4 (encoding for the HA protein) in the left-hand column (a, d, g, j, m); segment 6 (encoding for the NA protein) in the middle column (b, e, h, k, n), and segment 7 (encoding for the M1/M2 proteins) in the right-hand column (c, f, i, l, o). Each row represents one of five samples (a to c = sample 1 IAV-positive; d to f = sample 2 IAV-positive; g to i = sample 3 IAV-positive; j to l = sample 4 IAV-negative; m to o = sample 5 IAV-negative; see Table 1). NCBI accession numbers MT233941, MT269490, and MT269491 were used as references IAV segments 4, 6, and 7, respectively.

estimated cost and workflow bench and bioinformatic time is not notably different compared to shotgun metagenomics. TELSVirus can achieve a turnaround time of approximately 36 h from sample receipt to finished analysis, 22 h of which is the sequencing run-time and can be user-defined (Table S7). The estimated per-sample materials cost of TELSVirus is

\$300 to \$400 depending on the number of libraries multiplexed onto one flow cell (Table S8). A direct cost and runtime comparison of TELSVirus to shotgun metagenomics is difficult because such comparisons typically calculate the cost per sequenced base pair. However, the higher on-target rate of TELSVirus means that each base pair generated by

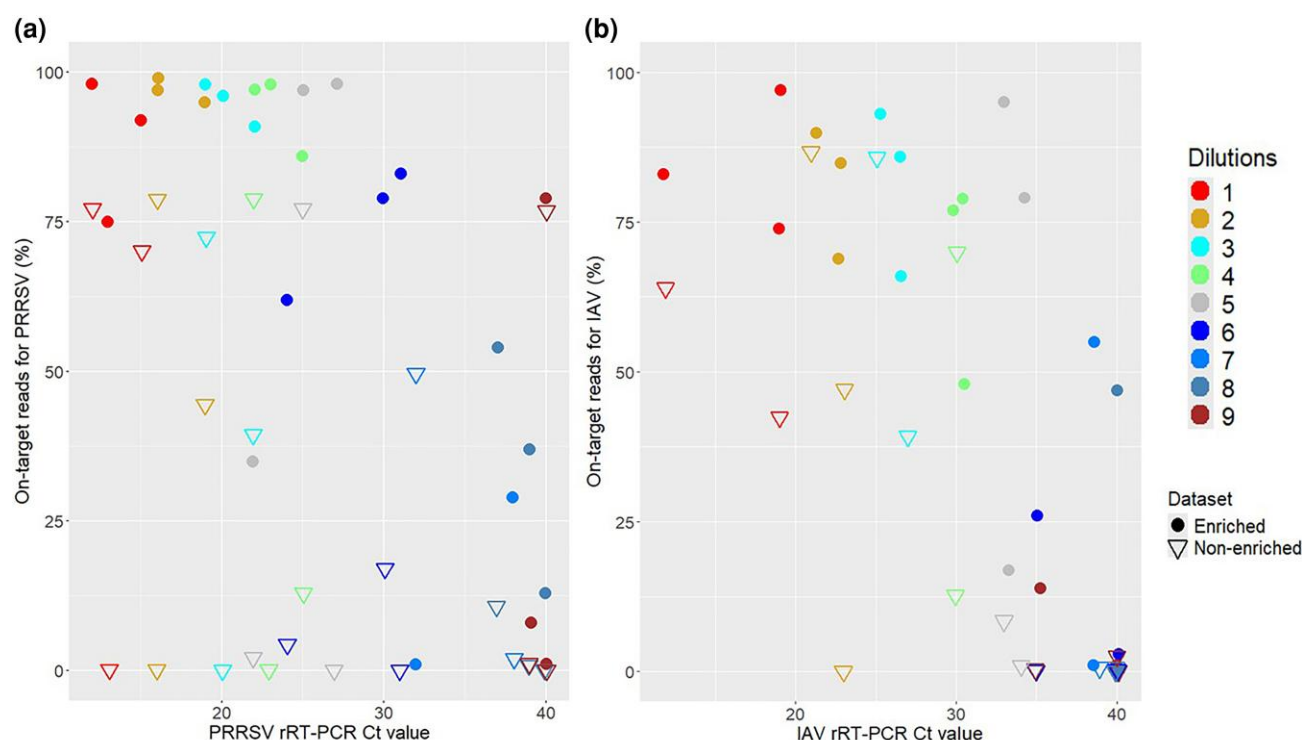


Fig. 3. Comparison of rRT-PCR (x axis) and TELSVirus (y axis) performance on dilution series of samples positive for a) PRRSV and b) IAV. Each dot and triangle represents a single biological replicate, with the rRT-PCR value representing the mean of four technical rRT-PCR replicates, and the TELSVirus results representing one technical TELSVirus replicate. Dots and triangles are colored by dilution number (1 to 9). One PRRSV TELSVirus library (dilution 7) and one IAV TELSVirus library (dilution 7) were removed due to failed basecalling.

TELSVirus has more information than each base pair generated in a shotgun metagenomic dataset. Thus, an appropriate cost comparison is highly dependent on the on-target rate achieved by both the shotgun metagenomic and the TELSVirus workflows, which is difficult to predict for any given target and sample matrix.

It is crucial to acknowledge that TELSVirus did not achieve the same level of performance in OF samples as compared to rRT-PCR, as indicated by a relatively high rate of false-negative detections. However, the false-negative rate is highly dependent on the Ct value used as the threshold for determining a positive sample via rRT-PCR. If the Ct threshold is lowered, TELSVirus' false-negative rate decreases, with only slight increases in the false-positive rate (Table S9). The Ct threshold for PRRSV rRT-PCR results is not well established within the scientific literature, and therefore, we provided a sensitivity analysis for this threshold at differing values (Table S9).

Even at low Ct thresholds (i.e. 35), TELSVirus continued to generate false-negative results for PRRSV in the OF samples, which could be problematic for surveillance purposes. We hypothesized that the occurrence of false negatives can be attributed to two main factors: firstly, the relatively low sequencing depth of 12 samples per flow cell; and secondly, the extensive viral diversity present in the OF samples. Notably, the majority of the OF samples were found to contain more than ten targeted viruses, several of which achieved very high coverage depth (Table 3). In such cases, these viruses would dominate the sequencing pores for most of the run. Consequently, any PRRSV nucleic acids present in the library might not be sequenced adequately, leading to false-negative results. The robust PRRSV coverage obtained from the serum samples in the dilution series experiment supports this hypothesis, as it demonstrates that we were able to detect PRRS from enriched

samples that lacked other targeted viruses and which would have otherwise outcompeted PRRSV for sequencing resources. The use of DMEM as the background diluent for the serial dilutions allowed us to be confident in the pan-viral negative status of the dilutions for other ubiquitous targeted viruses. However, the use of DMEM also artificially reduced the complexity of the sample matrix, making the dilutions less representative of actual field samples with very low viral burden. Balancing such trade-offs is a common challenge for metagenomic investigations, and future work will need to perform precise validation assays using multiple diverse sample matrices.

For samples with high complexity such as the OF samples, there are two potential approaches to improve the detection ability of TELSVirus: firstly, increasing the depth of sequencing can help capture a more comprehensive snapshot of the viral content, including lower abundance viruses like PRRSV. Secondly, adjusting the probe design to exclude viruses that typically occur in high copy numbers, such as astrovirus and bocavirus, can reduce competition for sequencing resources. Future research should focus on optimizing both the bait design panel and sequencing strategy aiming for a more targeted and efficient detection of low-abundance viruses in complex sample matrices.

In addition to false-negative results, we found that both TELSVirus and untargeted metagenomic sequencing generated a small number of alignments to IAV and PRRSV genomes that were considered absent from a given sample based on rRT-PCR results (Table 1 and Fig. 2). These findings are not unexpected for metagenomic datasets, as they can arise from a variety of technical and biological factors. Technical factors include well-documented phenomena such as index hopping and cross-sample contamination (Ezpeleta et al. 2022; Rollin et al. 2023). However, index hopping is not a documented

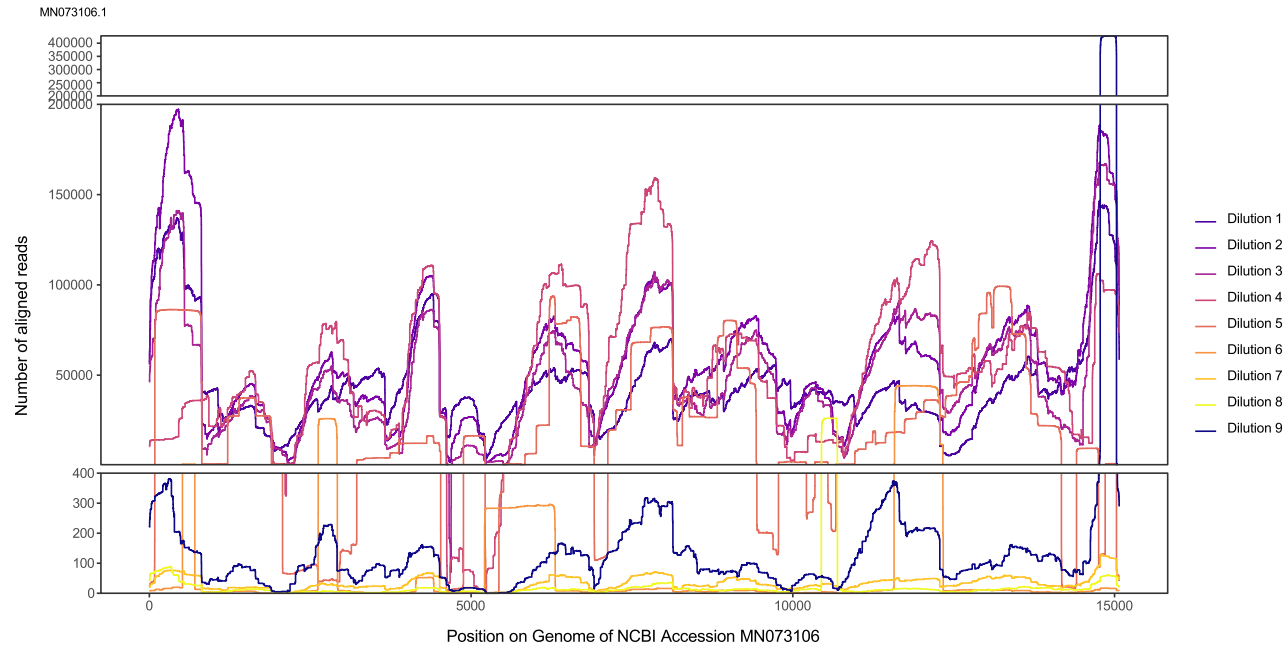


Fig. 4. Coverage profiles for PRRSV NCBI accession MN073106 for one of the three PRRSV-positive serum samples subjected to serial dilution. Each coverage profile line represents one of the nine dilutions. The x axis represents the genome position, and the y axis represents the number of aligned reads (coverage depth). Note that the y axis increments have been split into thirds to allow visualization of each dilution series’ coverage profile (i.e. 0–400 reads, 400–200,000 reads, and 200,000–400,000 reads).

Table 2 Genome fraction (%) and coverage depth (reads) for PRRSV NCBI accession MN073106 across nine serial dilutions, for one of the three PRRSV-positive serum samples subjected to serial dilution

Dilution	PRRSV genome fraction (%)	PRRSV coverage depth (reads)
1	100	5420
2	100	5248
3	100	4941
4	100	4669
5	99.8	3277
6	88.2	415
7	98.8	24
8	98.3	109
9	98.5	180

problem for ONT sequencing, and we took measures to successfully minimize cross-contamination, as evidenced by the lack of detectable nucleic acid in our extraction blanks and NTCs. Additionally, cross-contamination occurring prior to hybridization would result in a high on-target rate because the contaminating nucleic acids would be captured and amplified during the TELSVirus procedure. However, in all cases, the on-target rates in the false-positive samples were extremely low (<1%). Therefore, cross-contamination is considered an unlikely source of the target reads. In addition to these technical artifacts, biological factors can also lead to false-positive results, and these factors can be divided into three categories. Firstly, the TELSVirus workflow may possess a lower limit of detection compared to rRT-PCR. This implies that results typically regarded as the “gold standard” negative findings from rRT-PCR might actually be false negatives, suggesting the TELSVirus results could be the more accurate reflection of viral presence. Determining the true accuracy of each test result without extensive, multi-modal testing of each sample is challenging. However, it is important to note that comprehensive evaluations

of diagnostic tests have sometimes found rRT-PCR to be less sensitive than anticipated, particularly for PRRSV detection (Wagstrom et al. 2000). Secondly, it is possible that both test results are accurate, even if they seem to conflict. This can occur when the primer target of the rRT-PCR assay is absent in the sample, but other segments of the genome are present. For instance, in this study, the matrix gene was the target of the rRT-PCR assay for IAV. If the matrix gene was not present in the sample, but other IAV segments were, then rRT-PCR results would be negative, while TELSVirus results would be positive. However, based on our analysis, this scenario was unlikely to be causing the discrepant results in samples that tested negative for IAV based on rRT-PCR, because we found alignments to the matrix gene in our TELSVirus results, suggesting that the matrix gene was indeed present (Fig. 3; Table S3). Thirdly, sequence homology between the virus-of-interest and other organisms present in the sample can cause false-positive metagenomic results. In this scenario, nucleic acids from non-IAV organisms align to the IAV genomes because of a homologous locus. This well-documented issue is difficult to eliminate because of incomplete reference databases combined with extensive sequence homology across phylogenetically distant organisms (Doster et al. 2019; Ye et al. 2019; Sun et al. 2023). The sequence homology issue would typically present within the genome coverage profile as a locus of relatively high coverage depth across a very small segment of the genome. In analyzing our data, we did not find the aforementioned coverage profile for either IAV or PRRSV (as depicted in Fig. 3 and Fig. S1), indicating that sequence homology did not significantly contribute to false-positive results in our study. Consequently, we infer that the alignments to IAV and PRRSV in samples deemed negative by rRT-PCR were probably due to technical factors or lower limit of detection afforded by the TELSVirus method. Future studies could augment the utility of TELSVirus by extensively validating its sensitivity, specificity, positive and negative predictive values, and limit of detection. Such studies will need to estimate these values for

Table 3 TELSVirus results for virus prevalence, maximum and minimum genome fraction, and coverage depth, obtained from OF samples collected between 2017 and 2018 from Midwestern US commercial swine farms

Virus	Number of positive samples (N = 144)	Genome fraction (%)		Coverage depth (reads)	
		Maximum	Minimum	Maximum	Minimum
Porcine astrovirus 4	144	99.58	10.33	1848.51	0.11
Porcine bocavirus	144	100	35.67	3527.71	0.71
Porcine sapelovirus 1	144	89.47	2.48	125.43	0.02
Porcine astrovirus 2	143	99.08	7.86	552.87	0.08
Influenza A virus (H1N1)	141	100	4.48	2685.37	0.04
Porcine torovirus	141	98.02	0.54	1840.37	0.01
Respirovirus suis	139	100	0.87	3891.60	0.01
Teschovirus A	137	69.14	1.07	22.50	0.01
Porcine kobuvirus	132	40.48	1.55	6.20	0.02
Atypical porcine pestivirus	130	100	0.66	1366.41	0.01
Porcine hemagglutinating encephalomyelitis virus	130	99.89	0.53	1997.80	0.01
Porcine astrovirus 3	121	96.76	1.36	262.73	0.01
Suid betaherpesvirus 2	115	45.87	0.08	9.33	0.00
Influenza A virus	107	100	3.94	5633.48	0.04
Porcine adenovirus 5	106	44.17	0.25	76.49	0.03
Rotavirus A	105	100	2.91	2995.51	0.03
Human rotavirus A	90	98.42	2.15	21.43	0.02
Transmissible gastroenteritis virus	90	91.05	0.36	473.46	0.004
Porcine circovirus 2	81	89.69	5.21	202.23	0.05
Porcine respiratory coronavirus	78	94.39	0.55	530.60	0.01

Only viruses with a prevalence of at least 50% are shown. For full results, see [Table S4](#).

each targeted virus individually, and across different sample matrices obtained from populations with known prevalence.

Lastly, robust coverage statistics were observed for understudied viruses such as porcine bocavirus, porcine sapelovirus, porcine astrovirus 2 and 4, porcine respirovirus, and porcine torovirus ([Table 3](#)). Given that these viruses are commonly found in both healthy and clinically affected pigs ([Prpić et al. 2024](#)), their high prevalence in the collected OF samples was anticipated. However, the pathogenic nature of these viruses and their potential involvement in complex disease etiologies remain uncertain. This underscores the need for further research to understand the implications of these viruses for swine health and disease management, as well as to increase our understanding of the potential health impacts of co-circulating viruses. Hence, the outcome from novel enrichment workflows, such as TELSVirus, could contribute to such future studies by enhancing the comprehension of the genomic diversity of under-characterized and potentially pathogenic viruses.

Supplementary material

Supplementary material is available at [Molecular Biology and Evolution](#) online.

Acknowledgments

We thank Dr. Albert Rovira, Dr. Christopher Faulk, and Dr. Declan Schroeder for their constructive feedback during the development of this project. We thank Evan Kipp for the technical assistance in Nanopore sequencing. We thank Dr. Jose Angulo, Dr. Gustavo Lopez-Moreno, Dr. Claudio Marcello Melini, Dr. Joaquin Alvarez Norambuena, and My Yang for providing the samples used in this project.

Author Contributions

N.R.N., C.B., and M.T. conceived and supervised the project. M.M. performed probe capture and Nanopore sequencing

and conducted descriptive and statistical analysis. J.B. developed the bioinformatic pipeline. P.M.F. assisted with data analysis and data management. T.N.G. generated data visualizations and assisted with descriptive and statistical analysis. All authors contributed to writing the manuscript.

Funding

This work was supported by the Swine Health Information Center (SHIC) (grant number 22-071 to M.T.); a 2022 Resident & Graduate Student Research Grant from the College of Veterinary Medicine at the University of Minnesota (internal grant, no grant number assigned to M.M.); the National Institutes of Health (NIH) National Institute of Allergy and Infectious Disease (NIAID) (grant number 1R01AI141810-01 to C.B. and grant number 1R01AI173928-01 to N.R.N.); and by the USDA National Institute of Food and Agriculture (grant number 024555 to N.R.N.).

Conflict of Interest

We declare no competing interests that might be perceived as contributing to potential bias in this manuscript.

Data Availability

The bioinformatic methods and scripts used in this study are freely available at <https://github.com/jonathan-bravo/TELSVirus>. All raw data from ONT MinION sequencing experiments described here are deposited in the NCBI Sequence Read Archive under BioProject ID PRJNA1087754.

References

- Alanko JN *et al.* Syotti: scalable bait design for DNA enrichment. *Bioinformatics*. 2022;38:i177–i184. <https://doi.org/10.1093/bioinformatics/btac226>.
- Angulo J, Yang M, Rovira A, Davies PR, Torremorell M. Infection dynamics and incidence of wild-type porcine reproductive and

- respiratory syndrome virus in growing pig herds in the U.S. Midwest. *Prev Vet Med.* 2023;217:105976. <https://doi.org/10.1016/j.prevetmed.2023.105976>.
- Briese T *et al.* Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio.* 2015;6:e01491–e01415. <https://doi.org/10.1128/mBio.01491-15>.
- Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics.* 2017;33:3137–3139. <https://doi.org/10.1093/bioinformatics/btx373>.
- Cai D, Sun Y. Reconstructing viral haplotypes using long reads. *Bioinformatics.* 2022;38:2127–2134. <https://doi.org/10.1093/bioinformatics/btac089>.
- Ceballos-Garzon A, Comtet-Marre S, Peyret P. Applying targeted gene hybridization capture to viruses with a focus to SARS-CoV-2. *Virus Res.* 2024;340:199293. <https://doi.org/10.1016/j.virusres.2023.199293>.
- Chaisson MJP *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun.* 2019;10:1784. <https://doi.org/10.1038/s41467-018-08148-z>.
- Dávila-Ramos S *et al.* A review on viral metagenomics in extreme environments. *Front Microbiol.* 2019;10:2403. <https://doi.org/10.3389/fmicb.2019.02403>.
- Doster E *et al.* A cautionary report for pathogen identification using shotgun metagenomics; a comparison to aerobic culture and polymerase chain reaction for *Salmonella enterica* identification. *Front Microbiol.* 2019;10:2499. <https://doi.org/10.3389/fmicb.2019.02499>.
- Edwards MC, Gibbs RA. Multiplex PCR: advantages, development, and applications. *Genome Res.* 1994;3:S65–S75. <https://doi.org/10.1101/gr.3.4.S65>.
- Ezpeleta J *et al.* Robust and scalable barcoding for massively parallel long-read sequencing. *Sci Rep.* 2022;12:7619. <https://doi.org/10.1038/s41598-022-11656-0>.
- Frey KG, Bishop-Lilly KA. Chapter 15—next-generation sequencing for pathogen detection and identification. In: Sails A, Tang YW, editors. *Methods in microbiology, current and emerging technologies for the diagnosis of microbial infections*. Vol. 42: Academic Press; 2015. p. 525–554.
- Gaudin M, Desnues C. Hybrid capture-based next generation sequencing and its application to human infectious diseases. *Front Microbiol.* 2018;9:2924. <https://doi.org/10.3389/fmicb.2018.02924>.
- Henaio-Diaz A, Giménez-Lirola L, Baum DH, Zimmerman J. Guidelines for oral fluid-based surveillance of viral pathogens in swine. *Porcine Health Manag.* 2020;6:28. <https://doi.org/10.1186/s40813-020-00168-w>.
- Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol.* 2017;15:183–192. <https://doi.org/10.1038/nrmicro.2016.182>.
- Kapel N *et al.* Evaluation of sequence hybridization for respiratory viruses using the twist bioscience respiratory virus research panel and the OneCodex respiratory virus sequence analysis workflow. *Microb Genom.* 2023;9:001103. <https://doi.org/10.1099/mgen.0.001103>.
- Kent WJ. BLAT—The BLAST-like alignment tool. *Genome Res.* 2002;12:656–664. <https://doi.org/10.1101/gr.229202>.
- Kolmogorov M, Rayko M, Yuan J, Pevikov E, Pevzner P. metaFlye: scalable long-read metagenome assembly using repeat graphs. *bioRxiv.* 2020. <https://doi.org/10.1101/637637>.
- Lalonde C, Provost C, Gagnon CA. Whole-genome sequencing of porcine reproductive and respiratory syndrome virus from field clinical samples improves the genomic surveillance of the virus. *J Clin Microbiol.* 2020;58:e00097–e00020. <https://doi.org/10.1128/JCM.00097-20>.
- Lee HK *et al.* Comparison of mutation patterns in full-genome A/H3N2 influenza sequences obtained directly from clinical samples and the same samples after a single MDCK passage. *PLoS One.* 2013;8:e79252. <https://doi.org/10.1371/journal.pone.0079252>.
- Lee JS *et al.* Targeted enrichment for pathogen detection and characterization in three felid species. *J Clin Microbiol.* 2017;55:1658–1670. <https://doi.org/10.1128/JCM.01463-16>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Lin Y *et al.* Rapid PCR-based nanopore adaptive sequencing improves sensitivity and timeliness of viral clinical detection and genome surveillance. *Front Microbiol.* 2022;13:929241. <https://doi.org/10.3389/fmicb.2022.929241>.
- Lipkin WI, Firth C. Viral surveillance and discovery. *Curr Opin Virol.* 2013;3:199–204. <https://doi.org/10.1016/j.coviro.2013.03.010>.
- Lopez-Moreno G *et al.* Evaluation of dam parity and internal biosecurity practices in influenza infections in piglets prior to weaning. *Prev Vet Med.* 2022;208:105764. <https://doi.org/10.1016/j.prevetmed.2022.105764>.
- Luo X, Xiongbin K, Alexander S. Strainline: full-length de novo viral haplotype reconstruction from noisy long reads. *Genome Biol.* 2022;23:29. <https://doi.org/10.1186/s13059-021-02587-6>.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
- McGinnis J, Laplante J, Shudt M, George KS. Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. *J Clin Virol.* 2016;79:44–50. <https://doi.org/10.1016/j.jcv.2016.03.005>.
- Metsky HC *et al.* Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat Biotechnol.* 2019;37:160–168. <https://doi.org/10.1038/s41587-018-0006-x>.
- Nieuwenhuijsen DF *et al.* Towards reliable whole genome sequencing for outbreak preparedness and response. *BMC Genomics.* 2022;23:569. <https://doi.org/10.1186/s12864-022-08749-5>.
- Nirmala J *et al.* Evaluation of viral RNA extraction methods to detect porcine reproductive and respiratory syndrome and influenza A viruses from used commercial HVAC air filters from swine farms. *J Aerosol Sci.* 2021;151:105624. <https://doi.org/10.1016/j.jaerosci.2020.105624>.
- Noyes NR *et al.* Enrichment allows identification of diverse, rare elements in metagenomic resistome-virulome sequencing. *Microbiome.* 2017;5:142. <https://doi.org/10.1186/s40168-017-0361-8>.
- Paskey AC *et al.* Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. *BMC Genomics.* 2019;20:155. <https://doi.org/10.1186/s12864-019-5543-2>.
- Pogka V *et al.* Targeted virome sequencing enhances unbiased detection and genome assembly of known and emerging viruses—the example of SARS-CoV-2. *Viruses.* 2022;14:1272. <https://doi.org/10.3390/v14061272>.
- Prickett JR *et al.* Oral-fluid samples for surveillance of commercial growing pigs for porcine reproductive and respiratory syndrome virus and porcine circovirus type 2 infections. *J Swine Health Prod.* 2008;16:86–91. <https://doi.org/10.54846/jshap/565>.
- Prpić J, Keros T, Božiković M, Kamber M, Jemeršić L. Current insights into Porcine Bocavirus (PBoV) and its impact on the economy and public health. *Vet Sci.* 2024;11:677. <https://doi.org/10.3390/vetsci11120677>.
- Rehn A *et al.* Catching SARS-CoV-2 by sequence hybridization: a comparative analysis. *mSystems.* 2021;6:e00392–e00321. <https://doi.org/10.1128/msystems.00392-21>.
- Rockett RJ *et al.* Co-infection with SARS-CoV-2 Omicron and Delta variants revealed by genomic surveillance. *Nat Commun.* 2022;13:2745. <https://doi.org/10.1038/s41467-022-30518-x>.
- Rollin J, Rong W, Massart S. Cont-ID: detection of sample cross-contamination in viral metagenomic data. *BMC Biol.* 2023;21:217. <https://doi.org/10.1186/s12915-023-01708-w>.
- Sari G *et al.* Hepatitis E virus shows more genomic alterations in cell culture than in vivo. *Pathogens.* 2019;8:255. <https://doi.org/10.3390/pathogens8040255>.
- Schroeder DC *et al.* Two distinct genomic lineages of Sinaivirus detected in Guyanese Africanized honey bees. *Microbiol Resour Anounc.* 2022;11:e00512–e00522. <https://doi.org/10.1128/mra.00512-22>.
- Schuele L *et al.* Application of shotgun metagenomics sequencing and targeted sequence capture to detect circulating porcine viruses in

- the Dutch–German border region. *Transbound Emerg Dis*. 2022;69:2306–2319. <https://doi.org/10.1111/tbed.14249>.
- Slizovskiy IB *et al*. Target-enriched long-read sequencing (TELSeq) contextualizes antimicrobial resistance genes in metagenomes. *Microbiome*. 2022;10:185.
- Slomka MJ *et al*. Real-time reverse transcription (RRT)-polymerase chain reaction (PCR) methods for detection of pandemic (H1N1) 2009 influenza virus and European swine influenza A virus infections in pigs. *Influenza Other Respir Viruses*. 2010;4:277–293. <https://doi.org/10.1111/j.1750-2659.2010.00149.x>.
- Steward GF *et al*. Are we missing half of the viruses in the ocean? *ISME J*. 2013;7:672–679. <https://doi.org/10.1038/ismej.2012.121>.
- Sun Z *et al*. Removal of false positives in metagenomics-based taxonomy profiling via targeting Type IIB restriction sites. *Nat Commun*. 2023;14:5321. <https://doi.org/10.1038/s41467-023-41099-8>.
- Taylor MK *et al*. Amplicon-based, next-generation sequencing approaches to characterize single nucleotide polymorphisms of Orthohantavirus species. *Front Cell Infect Microbiol*. 2020;10:603817. <https://doi.org/10.3389/fcimb.2020.565591>.
- Tulloch RL, Kok J, Carter I, Dwyer DE, Eden JS. An amplicon-based approach for the whole-genome sequencing of human metapneumovirus. *Viruses*. 2021;13:499. <https://doi.org/10.3390/v13030499>.
- Vandenbussche F, Mathijs E, Tignon M, Vandersmissen T, Cay AB. WGS- versus ORF5-based typing of PRRSV: a Belgian case study. *Viruses*. 2021;13:2419. <https://doi.org/10.3390/v13122419>.
- Wagstrom EA, Yoon KJ, Cook C, Zimmerman JJ. Diagnostic performance of a reverse transcription-polymerase chain reaction test for porcine reproductive and respiratory syndrome virus. *J Vet Diagn Invest*. 2000;12:75–78. <https://doi.org/10.1177/104063870001200116>.
- Wylezich C *et al*. Next-generation diagnostics: virus capture facilitates a sensitive viral diagnosis for epizootic and zoonotic pathogens including SARS-CoV-2. *Microbiome*. 2021;9:51. <https://doi.org/10.1186/s40168-020-00973-z>.
- Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. *Cell*. 2019;178:779–794. <https://doi.org/10.1016/j.cell.2019.07.010>.
- Zaragoza-Solas A, Haro-Moreno JM, Rodriguez-Valera F, López-Pérez M. Long-read metagenomics improves the recovery of viral diversity from complex natural marine samples. *mSystems*. 2022;7:e00192–e00122. <https://doi.org/10.1128/msystems.00192-22>.