BMC Genomics

RESEARCH

Open Access

Sensitive, flexible, and affordable serum RNA sequencing for pathogen detection on the Oxford Nanopore platform

Philippe Selhorst[1,2*], Emilie Van Vyve[1], Francesca Falconi-Agapito[1,3], Joachim Mariën[4,5] and Kevin K. Ariën[1,6]

Abstract

Metagenomic sequencing for pathogen detection has traditionally suffered from low sensitivity due to the overwhelming presence of host nucleic acids. Commercial host-depletion kits are often prohibitively expensive and limited to specific species, hindering adoption in resource-limited settings, where the burden of zoonotic diseases is highest. To address this, we optimized and combined Sequence-Independent Single Primer Amplification (SISPA) with Depletion of Abundant Sequences by Hybridization (DASH), establishing a low-cost metagenomic protocol on the Oxford Nanopore sequencing platform. Our approach can be adapted to any species to detect microbial RNAs in serum samples at PCR-range sensitivity, outperforming existing methods in the field.

Keywords SISPA, CRISPR/Cas-9, Metagenomics, DASH, Sequencing, mNGS, Virus, RNA, Host-depletion, Oxford Nanopore, Pathogen

Background

Pathogen sequencing has become a major cornerstone of public health surveillance and disease control by increasing our understanding of pathogen biology, epidemiology, diversity and evolution [1]. Whole-genome sequences are generally obtained through targeted methods such as PCR amplicon sequencing. Although these methods are sensitive and inexpensive, they require prior knowledge of the pathogen, perform poorly on highly variable genomes, and are unable to detect co-infections or novel pathogens. Metagenomic sequencing on the other hand, does not suffer the same drawbacks as it comprises the untargeted sequencing of all genomic material present in a sample. Although these untargeted methods have revolutionized microbiome research, clinical diagnostics, and virus discovery [2], one major challenge remains the overwhelming presence of host nucleic acids in the sample which severely reduces their sensitivity for pathogen detection. This necessitates the introduction of additional host-depletion or target enrichment steps resulting in more complex and expensive protocols in comparison to the traditional, targeted methods.

One of the more frequently used metagenomic sequencing methods is based on reverse transcription (RT) and second strand synthesis of all RNA using a

*Correspondence:
Philippe Selhorst
pselhorst@itg.be
[1]Department of Biomedical Sciences, Virology Unit, Institute of Tropical Medicine, Antwerp, Belgium
[2]The Outbreak Research Team, Institute of Tropical Medicine, Antwerp, Belgium
[3]Universidad Peruana Cayetano Heredia, Instituto de Medicina Tropical Alexander Von Humboldt, Grupo de Virología, Unidad de Epidemiología Molecular, Lima, Peru
[4]Department of Biomedical Sciences, Virus Ecology Unit, Institute of Tropical Medicine, Antwerp, Belgium
[5]Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium
[6]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

random nonamer with a known overhang and the subsequent amplification of the double-stranded (ds) complementary DNA (cDNA) by Sequence-Independent Single Primer Amplification (SISPA) [3]. This method was successfully used to retrieve full genomes of dengue virus (DENV), and chikungunya virus (CHIKV) [4] and its application on the Oxford Nanopore (ONT) minION platform allowed for a rapid, in-field outbreak investigation of Lassa virus in Nigeria [5]. Although this protocol showed potential for use in resource-limited settings, it allows for several improvements. Most importantly, the lack of a host-depletion step in this protocol severely limits its ability to detect and sequence low-abundance pathogens in blood such as Zika virus (ZIKV) and West Nile virus or to catch infections during convalescence.

A commonly used method for host-depletion in viral metagenomics is the use of nucleases (Benzonase + micrococcal nuclease) [6] prior to extraction to digest unprotected, free nucleic acids in the sample that are not packaged in virions. Although cheap and easy to perform, to what extent different viruses are protected and how it compares to other methods, remains unclear. In contrast to nucleases, multiple commercial kits exist that can selectively deplete ribosomal RNAs (rRNAs). One example is Qiagen's FastSelect reagent, a pool of locked nucleic acid-containing DNA oligos with extremely high melting temperatures that can be hybridized to the unwanted RNA to block subsequent RT. However, commercial kits are often cost-prohibitive as they can easily double the cost per sample sequenced. Moreover they are only available for certain key species such as human, mouse, and rat, making them inflexible for research on other organisms.

A more promising and affordable host depletion method is DASH (Depletion of Abundant Sequences by Hybridization) which was pioneered by Gu et al. [7] to deplete human mitochondrial rRNAs (12S and 16S) in cell extracts. Single-stranded guide RNAs (sgRNA), each designed with complementarity to a 20-nt long genomic target, are complexed to Streptococcus pyogenes CRISPR-associated nuclease 9 (SpCas9) which will induce a double strand break upon recognition of the target sequence by the sgRNA. Loi et al. [8] then expanded this method to recognize human cytoplasmic rRNAs (18S, 5.8S, 28S) for single cell sequencing.

In this paper we present an improved SISPA-based metagenomics protocol (i.e., SISPA 2.0) for the Oxford Nanopore platform, with increased pathogen sensitivity and diagnostic reliability at a substantially lower cost. We further offer valuable insights into the intricacies, best practices, and common pitfalls of the SISPA and DASH methods.

Materials and methods

Spiking of serum samples

Blood collected in BD Vacutainer serum tubes, was centrifuged for 15 min at 1000 g, after which the serum was pooled, and centrifuged a second time for 5 min at 2000 g. To evaluate changes to the sequencing protocol, serum pools from healthy individuals were frozen at −80 °C and spiked when needed with small amounts of virus culture supernatants or with virus from a high copy number clinical sample.

DASH design

Following the method described by Loi et al. [8], 20 bp sgRNA recognition sequences (Supp. Table 1, bold type) were designed by directing the software CHOPCHOPv2 [9] to a Kraken 2 build of RefSeq's viral database (https://benlangmead.github.io/aws-indexes/k2, accessed 2023) instead of the human genome. CHOPCHOP then identifies and scores potential sgRNA recognition sites in the provided targets (18S, 5.8S, 28S, 12S, 16S, ATP6, ND4) while also providing a list of possible off-targets. Only sgRNAs with the highest efficiency score for which zero off-targets with up to three mismatches exist in the viral database (excluding phages), were considered for selection every 100 to 150 bp on both sense and antisense strands. To further evaluate off-target cleavage in cellular pathogens, we used Cas-OFFinder 2 to identify sites in the non-redundant 16S/18S SILVA database (SILVA_138.2_SSURef_NR99_tax_silva.fasta (https://www.arb-silva.de/) for a selection of pathogens commonly detected in human serum (Supp.Table 2). Sites with more than 3 mismatches or a mismatch in the seed region (i.e. the 8 bases closest to the PAM region) were excluded and only one hit per species and sgRNA was retained.

sgRNA preparation

The DNA template of each sgRNA was generated from two single-stranded (ss) DNA oligos ordered from Integrated DNA Technologies (IDT) and listed in Supp. Table 1. The 54–55 bp sgRNA-specific forward primer (5´-TTC TAA TAC GAC TAC ATA TA(G)(sgRNA-20)GTT TTA GAG CTA GA-3´) contains a T7 RNA polymerase-specific promotor, followed by the 20 bp sgRNA sequence and a 14 bp overlapping sequence that is complementary to the universal reverse primer. A G nucleotide is added when the sgRNA sequence does not start with a G, to satisfy the sequence requirements of the T7 promoter. The 80 bp universal reverse primer (5´-AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT TAA CTT GCT ATT TCT AGC TCT AAA AC-3´) contains the sgRNA scaffold that associates with the Cas9 protein and has a 14 bp overlap with the forward primer. Both oligos were assembled and amplified into dsDNA using the KAPA HiFi HotStart

kit (07958927001, Roche). The amplicon concentrations were measured using the Qubit 4 High Sensitivity dsDNA kit (Q32851, Thermo Fisher Scientific), subsequently pooled at equimolar concentrations, and purified using the DNA Clean & ConcentratorTM-25 µg kit (Zymo, Cat# D4033). In vitro transcription was performed on 75 ng of DNA using the HiScribeTM T7 Quick High Yield RNA Synthesis Kit according to the provided protocol, including DNase treatment, addition of DTT, and overnight incubation. The produced sgRNAs were purified using the Monarch RNA cleanup kit (T2040, NEB), eluted in water or IDTE pH 7.5 buffer (11–01–02–02, IDT), and measured using the Qubit 4 broad range RNA kit (Q10210, Thermo Fisher Scientific) after which they were aliquoted and stored at −80 °C for further use.

Sample preparation

Nucleic acids were extracted from 200 µL spiked serum using Macherey Nagel's NucleoSpin Virus Mini kit (740,983.50) following the manufacturer's protocol but replacing carrier RNA with 8 µl of linear polyacrylamide (LPA, AM9520, Thermo Fisher Scientific), using 400 µL instead of 200 µL lysis buffer VL, and eluting with 50 µL water in Lo-Bind Eppendorf tubes. 44 µL of extract was transferred to tube strips (Thermo Fisher Scientific) to which a mix of 5 µL TURBO buffer and 1 µL TURBO DNase (AM2238, Thermo Fisher Scientific) was added and subsequently incubated for 30 min at 37 °C. Reactions were cleaned and concentrated to 15 µL using 1.5 × RNACleanXP beads (A63987, Beckman Coulter) and two 500 µL ethanol washes. To allow for detection of dsRNA viruses, an RNA denaturation step at 95 °C for 1 min, followed by rapid cooling to 4 °C, is performed after adding 1 µL of 100 µM solA primer (5′-GTT TCC CAC TGG AGG ATA NNN NNN NNN-3′) and 1 µL of 25 mM dNTPs (R1121, Thermo Fisher Scientific) to the 15 µL of cleaned RNA. A mix of 5 µL of Maxima H- RT buffer (EP0753, Thermo Fisher Scientific), 0.5 µL of RiboLock RNase Inhibitor (EO0382, Thermo Fisher Scientific), and 1.25 µL of Maxima H- enzyme is subsequently added to the denatured RNA up to a total volume of 25 µL. RT was performed as follows: 2 min at 10 °C, 2 min at 20 °C, 2 min at 30 °C, 2 min at 40 °C, 30 min at 50 °C, 5 min at 95 °C, 4 °C. After RT, 0.5µL of thermostable RNaseH (H39500, Lucigen) was added to the reaction, mixed, and incubated for 20 min at 50 °C. Second strand synthesis was performed by adding a mix of 2.4 µL water, 0.4 µL buffer, and 1.2 µL Klenow Fragment (3′–5′ exo-) (M0212L, NEB) to the 25 µL RT reaction and incubating at 37 °C for 1 h, followed by 15 min heat inactivation at 75 °C and cooling to 4 °C. The resulting 29 µL reaction was subjected to DASH depletion by first adding 1.5 µl of Albumin buffer consisting of water, 0.15 µL of 20 mg/ml Albumin (B9200S, NEB), 0.15 µL of 1 M MgCl2, and 0.3 µL of 5 M NaCl. Then, 10 µL of ribonucleoprotein complex is added and incubated for 2.5 h or overnight at 37 °C. This ribonucleoprotein complex was formed by a 10 min preincubation at 37 °C of 1 µL Cas9 buffer, 2300 ng sgRNAs, and 3.4 µL Cas9 enzyme (M0386S, NEB) added up to 10 µL with water. Following DASH, Cas9 is inactivated by adding 2 µL of a 1 in 5 dilution of proteinase K (19,134, Qiagen) and incubating at 37 °C for 10 min followed by 95 °C for 10 min. The 42 µL reactions were cleaned using 1 × AmpureXP beads (A63881, Beckman Coulter) and eluted with 13 µL water. 12.5 µL of PCR mix is then added containing 2.5 µL LA PCR buffer II, 2.5 µL of 25 mM MgCl2, 2.5 µL of 20 µM custom barcode primer (Supp. Table 3), 1.25 µL of 10 mM dNTPs (N0447L, NEB), and 0.25 µL (1.25U) of TaKaRa LA polymerase (RR002A, TaKaRa Bio) and 3.5 µL water. Cycling conditions consisted of an initial denaturation at 98 °C for 30 s followed by 26 cycles of 94 °C for 15 s, 52 °C for 20 s, 72 °C for 2 min, and a final elongation at 72 °C for 10 min. Amplified cDNA was purified using 1 × AmpureXP beads, eluted in 25 µL EB buffer (19,086, Qiagen), and quantified using the Qubit High Sensitivity dsDNA kit.

ONT library preparation and sequencing

The amplified DNA was end-prepped using the Ultra II End Repair/dA-Tailing Module (E7546L, NEB) by pooling 6.4 ng of each of four samples added up to a maximum of 40 µL with water. 10 µL end-prep mix was then added consisting of 1.7 µL water, 5.84 µL Ultra II buffer and 2.5 µL Ultra II enzyme. Incubation was as follows: 10 min at 22 °C, 5 min at 65 °C, 4 °C. All end-prep reactions were pooled and cleaned using 1 × AmpureXP beads in a single Lo-Bind tube with two washes of 500 µL ethanol and eluted in 25 µL of water. Adapter ligation was performed by adding 25 µL blunt/TA ligase master mix (M0367L, NEB) and 5 µL of ligation adapter (SQK-LSK114, ONT) to 62 ng of eluate followed by 15 min incubation at 22 °C. Reactions were cleaned using 0.6 × AmpureXP beads and two washes of 125 µL SFB (SQK-LSK114, ONT) and eluted in 15 µL EB (SQK-LSK114, ONT). The latter was then loaded onto an R10.4.1 flow cell (ONT) after mixing with 37.5 µL SB and 25.5 µL LIB (SQK-LSK114, ONT). Sequencing data were collected over 24–48 h.

Data processing and analysis

Raw sequence data (pod5 files) were basecalled in high accuracy mode with a minimum quality of Q9, and demultiplexed using ONT's Guppy v6.5.7 or Dorado v0.91 algorithm requiring barcodes on both ends. After demultiplexing, fastq files were QC-ed using NanoPlot v1.42 and filtered to remove short (< 200 bp) and low complexity reads using PRINSEQ++ while the flanking solB sequence was trimmed using a custom python v3.6

script. Filtered reads were then aligned to a fasta file containing multiple target reference sequences (host or viral) using minimap2 v2.17 with the "-x map-ont" and "–secondary = no" setting. Samtools v1.18 view, sort, and index were used to convert the resulting bam file into a sorted, indexed sam file for each target, filtering out supplementary alignments. Sequencing depth for each target was extracted using Samtools depth including zero coverage positions and deletions while the average read length and total bases were obtained using Samtools stats. The percentage virus in the library was calculated per read count and per base count. Depth plots were generated using matplotlib. Basecalled fastq files were also uploaded to the online metagenomic sequence analysis platform Chan Zuckerberg ID (https://czid.org/) to exclude external contamination.

Quantitative PCRs

For Phocine Distemper Virus (PDV), CHIKV, ZIKV, and DENV quantitation, RT-qPCR was performed with primers and probes as previously described [10–12]. Briefly, a 83 bp H fragment of PDV, or a 77 bp nsP1 fragment of CHIKV, or a 112 bp NS1 fragment of DENV-1, or a 77 bp E fragment DENV2, or a 107 bp NS1 fragment of ZIKV were amplified in 45 cycles by adding 2.5 µL RNA in a 12.5 µL reaction using the one-step SensiFAST™ Probe No-ROX Kit (BIO-76001, Meridian), which includes reverse transcriptase, and RiboSafe RNase inhibitor. Cycling conditions were 10 min at 45 °C, 2 min at 95 °C, followed by 45 cycles of 5 s at 95 °C and 20 s at 60 °C (or 56 °C for ZIKV specifically). For quantitation of human 18S and 16S a SYBR green RT-qPCR was performed using the one-step SensiFAST™ SYBR No-ROX Kit (BIO-72001, Meridian) with the same cycling and reaction conditions as above and using the following primers: 16S_qPCR_hs_Fwd (5´-TGC TTA GCC CTA AAC CTC AA CA-3´), 16S_qPCR_hs_Rev (5´-GGG TTT GCT GAA GAT GGC G-3´), 18S_v2_Fwd (5´-CGG CGA CGA CCC ATT CGA AC-3´), and 18S_v2_Rev (5´-GAA TCG AAC CCT GAT TCC CCG TC-3´).

Results

DASH design for viral and broader pathogen diagnostics

The sgRNAs from Gu et al. and Loi et al. were designed for human genomics and hence to only cleave rRNAs while leaving the human transcriptome intact. To apply DASH to virus detection, we redesigned the sgRNAs avoiding off-target effects in any of the known viruses (excluding bacteriophages). In addition to targeting rRNAs (12S, 16S, 18S, 5.8S, 28S), the sgRNA panel was expanded to include the mitochondrial genes ATP6 (ATP synthase membrane subunit 6) and ND4 (NADH dehydrogenase subunit 4) as they were observed at high levels in non-depleted libraries. Although cellular pathogen diagnostics (i.e. bacteria, fungi, protozoa, helminths etc.), are not the primary use case of our protocol, the current DASH panel was also evaluated for off-target cleavage within 16S and 18S rRNA sequences of respectively bacterial and eukaryotic species that can be detected in human serum (Supp. Table 2). This is important because 16S/18S rRNAs are the most abundant transcripts in cellular pathogens and often the only detectable remnants of their transcriptome in serum samples with low pathogen load, while also serving as widely used markers for species identification. Due to significant sequence similarities with human 18S, only for eukaryotic serum pathogens off-targets were found, which we listed in Supp. Table 4. Of note, most of those off-targets hits were caused by the same five sgRNAs indicating that sufficient sequence variation exists to find sgRNAs unique to humans and thus that a revised panel can be designed with limited reduction in panel redundancy.

DASH depletion can outcompete Qiagen's FastSelect -rRNA reagent

We then introduced DASH into the SISPA protocol between the TURBO DNase and PCR amplification step (Fig. 1) since Cas9, which is a single-turnover enzyme, will be more efficient when DNA quantities are low. To evaluate the efficacy of host depletion, we first compared DASH to the commercially available FastSelect -rRNA Reagent (Qiagen) which was hybridized to the RNA after TURBO DNase but prior to RT.

Figure 2A shows that DASH outcompetes Qiagen's FastSelect reagent with rRNA levels being depleted to 1% or below resulting in higher levels of the ssRNA virus PDV, and of the non-depleted mitochondrial transcripts ND1/2 (NADH dehydrogenase subunit 1/2), COI (Cytochrome c oxidase subunit I), and CYB (cytochrome B). A 2.5 h DASH further appeared equivalent if not better than an overnight incubation. Since DASH works on the DNA level, this could potentially make the TURBO DNase step redundant, especially when used in combination with nuclease treatment prior to extraction. We therefore compared different combinations of DASH, TURBO DNase, and nuclease treatments.

Figure 2B shows that for PDV, adding DASH resulted in a 10 × higher sensitivity compared to the original SISPA protocol (which only uses TURBO DNase post-extraction) or 6.9 × higher compared to the combination of pre-extraction nuclease treatment followed by TURBO DNase treatment post-extraction. Similar gains in detection were observed for e.g. the mitochondrial transcripts COI and CYB as well as the cytoplasmic transcripts B2M (β2 microglobulin) and ACTB (Actin beta). Surprisingly, pre-extraction nuclease treatment was only 1.14 × better at enriching than TURBO DNase treatment. The latter only digests free DNA, whereas nuclease treatment will

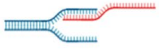
	SISPA 1.0	SISPA 2.0
Nuclease ✂ RNA/DNA	• None	• Not recommended
RNA extraction 	• Qiagen viral RNA mini	• MN Nucleospin virus
Turbo DNase ✂ DNA	• 30 min 37°C in 50 uL	• 30 min 37°C in 50 uL
Beads cleanup 	• 1.8x RNAClean XP beads	• 1.5x RNAClean XP beads • <1.5x depending on virus
Reverse Transcription 	• Denaturation 5 min 65°C • Superscript IV w/ RNaseOUT	• Denaturation 1 min 95°C • Maxima H- w/ Ribolock • Denaturation 5 min 95°C • RNaseH 50°C 20 min
Second strand synthesis 	• 2x 8 min Sequenase 2.0 • By adding 14 uL to RT reaction	• 45-60 min Klenow 3-5 exo- • By adding 4 uL to RT reaction • 15 min 75°C Klenow inactivation
Beads cleanup 	• RNAClean XP beads 1.0x • On 39 uL reaction	• None
DASH 	• None	• 2.5 hours or overnight • Inactivation recommended
Beads cleanup 	• None	• 1.0x RNAClean XP beads • On 40ul reaction
SISPA PCR 	• 100 uL reaction • SolB primer • AccuTaq LA	• 25 uL reaction • PCR barcodes • Takara Taq LA
Beads cleanup 	• 1.0x Ampure XP • On individual samples	• 1.0x Ampure XP • On individual samples
End prep & dA tailing 	• On individual samples	• On pools of 4 samples
Beads cleanup 	• 1.0x Ampure XP in 0.2ml tubes • On individual samples	• 1.0x Ampure XP in one 1.5ml tube • On pool of pools
Barcode ligation 	• On individual samples • Post-PCR contamination possible up to this point	• None, barcoded during PCR • No post-PCR contamination risk
Beads cleanup 	• 1.0x Ampure XP in one 1.5ml tube • On pool of samples	• None
Adapter ligation 	• Native adapter (NA)	• Ligation adapter (LA)

Fig. 1 Overview of the different steps in the original SISPA 1.0 and the improved SISPA 2.0 protocols

degrade both free RNA and DNA so we expected superior depletion. Although cytoplasmic rRNA (18S, 28S), and cytoplasmic transcripts B2M and ACTB were better depleted when using pre-extraction nuclease treatment, the detection of mitochondrial rRNA (16S) and transcripts (COI, CYB, ATP6, ND4) actually increased, indicating that their RNA was protected prior to extraction. Even when combining pre-extraction nuclease with post-extraction TURBO DNase, significant amounts of rRNA reads remain present in the sequencing library, suggesting they cannot effectively be removed by nuclease digestion.

In Fig. 2C, we tested whether we could drop the TURBO DNase step by combining nuclease treatment with DASH and whether a DNase or nuclease treatment is needed at all by only performing DASH. Interestingly, DASH by itself performed the worst with an 86 × drop in virus detection compared to DASH + TURBO which showed 2.9 × higher sensitivity than DASH + nucleases. However, DASH by itself did deplete its targets (18S, 28S, 12S, 16S, ATP6, ND4) efficiently. This suggests that the

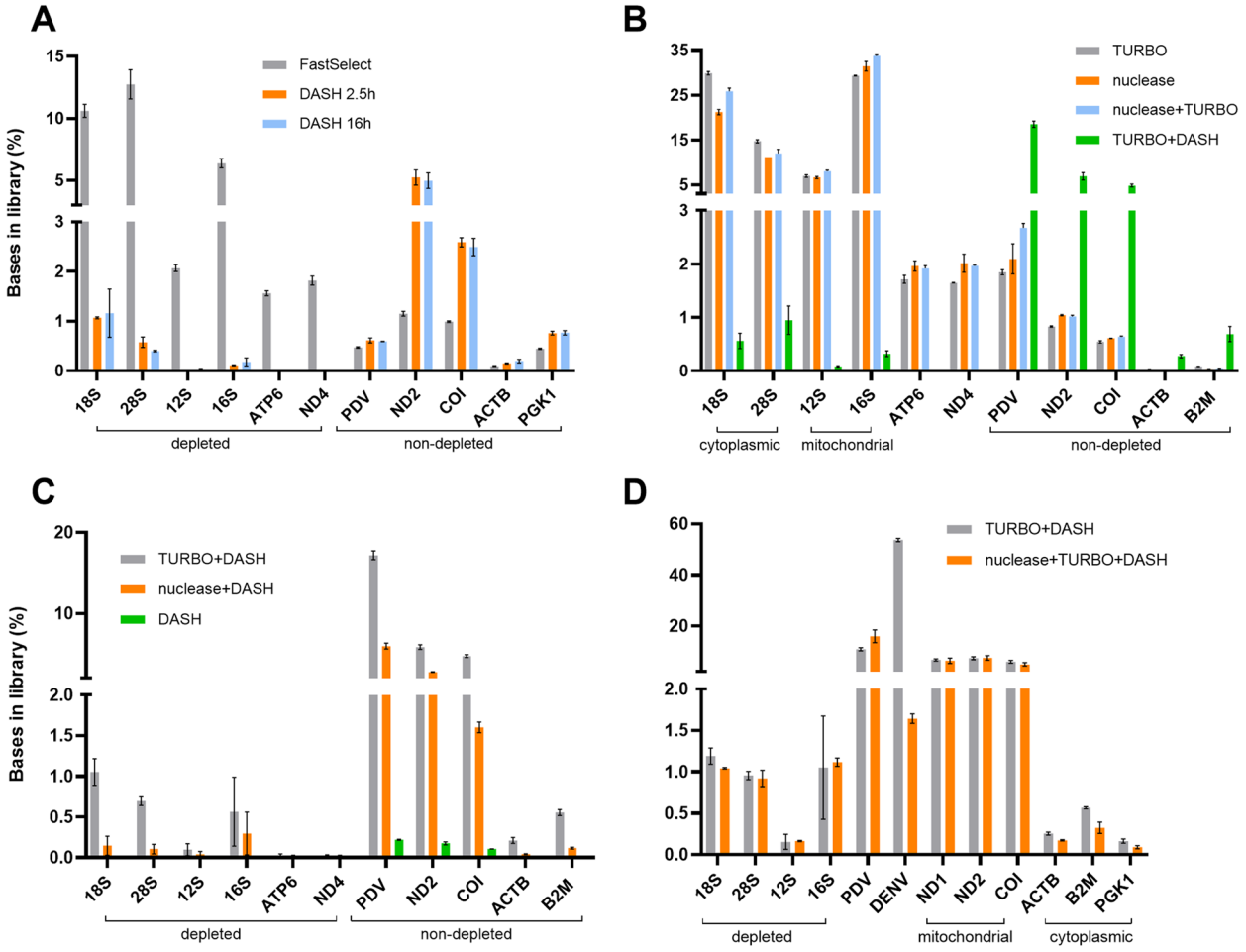


Fig. 2 Effect of different host depletion methods and their combinations on the detection of depleted and non-depleted, representative RNAs in metagenomic sequencing libraries of human serum samples spiked with the RNA viruses PDV (Phocine distemper virus) and/or DENV (Dengue virus). **A)** Qiagen's commercial FastSelect reagent was added prior to reverse transcription to remove DNA derived from human cytoplasmic (18S, 28S) and mitochondrial (12S, 16S) ribosomal RNAs. DASH was performed after second strand synthesis for 2.5h or 16h to remove DNA derived from 18S, 28S, 12S, 16S, and from mitochondrial transcripts ATP6 (ATP synthase membrane subunit 6) and ND4 (NADH dehydrogenase subunit 4). ND1/2 and COI are mitochondrial transcripts NADH dehydrogenase subunit 1/2 and cytochrome c oxidase subunit I. ACTB and PGK1 are cytoplasmic transcripts actin beta and phosphoglycerate kinase 1. **B)** TURBO DNase treatment was performed prior to reverse transcription to digest all co-extracted DNA molecules. Nuclease treatment, consisting of a cocktail of Benzonase and micrococcal nuclease, was performed prior to nucleic acid extraction to remove all unprotected DNA and RNA molecules in the serum sample. DASH was performed as above to deplete 18S, 28S, 12S, 16S, ATP6, and ND4. B2M is cytoplasmic transcript β2 microglobulin. **C+D)** Continuation of panel B but performed in different sequencing experiments

presence of host DNA in serum samples has the biggest impact on virus detection and that a significant amount of that DNA is unaffected by pre-extraction nuclease treatment. Finally, we investigated whether the combination of nuclease, TURBO, and DASH adds any value. Figure 2D shows that pre-extraction nuclease digestion resulted in a dramatic, 34 × drop in virus detection for a clinical DENV sample. For lab grown PDV virus however, detection sensitivity increased by 1.45 ×. Notably, pre-extraction nucleases again did not seem to reduce mitochondrial RNAs although they did cause a 1.7 × drop in cytoplasmic RNAs. We therefore hypothesized that during serum preparation, cell-free mitochondria [13] were not effectively removed as suggested by literature where mitochondrial isolation procedures recommend 15,000 g for 5 to 20 min [14].

Pre-extraction nuclease treatment is inefficient and can degrade viral nucleic acids

We confirmed the above observations through qPCR (Fig. 3) showing an 11 Ct increase when the same clinical sample of DENV was exposed to nucleases whereas lab grown PDV only lost 0.3 Ct, suggesting most of the DENV RNA was unprotected. In addition, nuclease digestion barely affected mitochondrial 16S rRNA concentration, whereas it decreased the concentration of cytoplasmic 18S rRNA by 1 Ct. To evaluate the removal of mitochondria, an additional serum centrifugation for

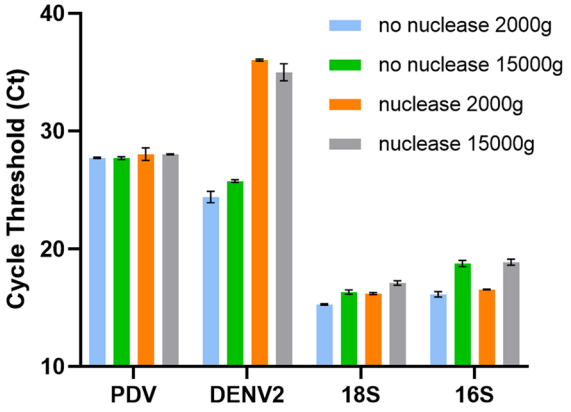


Fig. 3 The effect of pre-extraction nuclease treatment on qPCR detection of the single-stranded RNA viruses PDV (Phocine distemper virus), DENV2 (Dengue virus serotype 2), cytoplasmic ribosomal RNA 18S and mitochondrial ribosomal RNA 16S after sample centrifugation at 2,000g versus 15,000g

5 min at 15,000 g was performed prior to nuclease digestion. This decreased 18S by 1 Ct whereas 16S decreased by 2.5 Ct aligning with our hypothesis of 16S being associated with mitochondria that pellet at higher speeds. The decrease of 18S at higher centrifugation speeds while only showing a 1 Ct loss after nuclease treatment, suggests that a significant part of serum 18S is protected which is in line with studies reporting it to occur in extracellular vesicles [15]. To further evaluate the risk of viral RNA degradation when using nuclease treatment, we compared Ct levels before and after treatment for several clinical samples and lab grown viruses (Supp. Table 5). In line with lab grown PDV, a freshly frozen DENV2 lab virus was resistant to nuclease treatment. However, for 5 clinical samples of DENV2, which were collected in two different field studies in the Democratic Republic of the Congo and Peru, Ct values for viral RNAs increased up to Δ14.5 indicating complete degradation. Interestingly, clinical samples of CHIKV, DENV1/2, and ZIKV, which were collected from returning travelers in Belgium under ideal conditions, were less impacted showing a maximum Ct increase of 2.4. This suggests that samples, especially when stored under field conditions, can contain substantial amounts of likely disintegrated virus and hence unprotected viral RNA sensitive to nucleases.

Denaturation prior to RT is essential for dsRNA virus detection

RT usually starts with an RNA denaturation step (5 min, 65 °C) in presence of primer followed by rapid cooling to 4 °C to reduce RNA secondary structure. However, to enable the detection of dsRNA viruses this temperature is too low. Our experiments show a dramatic 281 × and 220 × increase in detection of respectively Rotavirus (spiked from a clinical sample) and Orthoreovirus 3 (spiked as part of the ATCC-MSA-2008 Virome Virus Mix) when a 1 min incubation at 95 °C is introduced. (Fig. 4A). Of note, although this higher denaturation temperature should help with clearing secondary structure, we observed a 1.41×, 1.31×, and 1.26 × lower detection of the ssRNA viruses DENV2, PDV, and TBEV (Tick Borne Encephalitis Virus) (Fig. 4B). However, this drop in detection was no longer observed when the RT enzyme was inactivated at 95 °C instead of the recommended 85 °C (for Maxima H Minus), with TBEV being even better detected. This suggested that in addition to pre-RT denaturation, the denaturation of the RNA–DNA hybrids post-RT plays an important role in virus detection.

Denaturation temperature post-RT influences virus and host detection

The original SISPA protocol cycled to 4 °C after RT and hence did not include an enzyme inactivation step nor an RNaseH or a 95 °C denaturation step to remove the

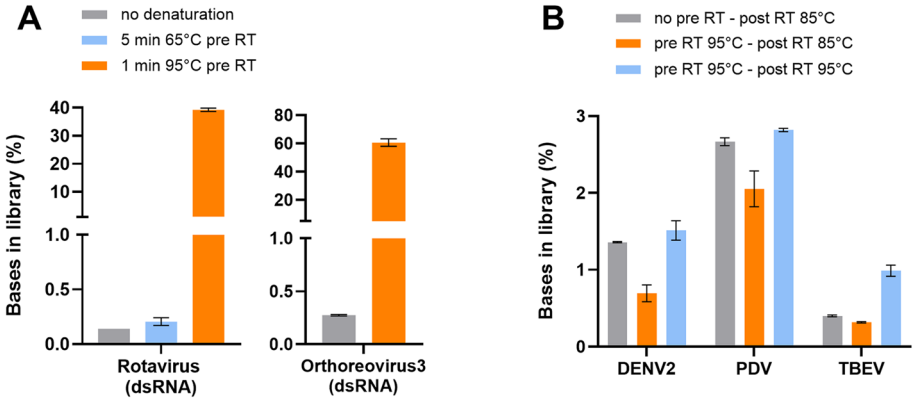


Fig. 4 **A)** Effect of a heat denaturation step prior to reverse transcription (RT) on the detection of double-stranded (ds)RNA viruses Orthoreovirus and Rotavirus in individually spiked metagenomic sequencing libraries of TURBO+DASH-depleted human serum samples. **B)** The effect of pre-RT heat denaturation on single-stranded RNA viruses PDV (Phocine Distemper Virus), TBEV (Tick Borne Encephalitis virus), and DENV2 (Dengue virus 2) at different denaturation/heat inactivation temperatures following RT

RNA from the RNA–DNA hybrids generated during the RT reaction. We hypothesized the latter to be essential for the solA primer to optimally bind along the cDNA and initiate second strand synthesis. As expected, we observed that introducing a post-RT denaturation step, significantly impacts virus detection and host depletion, however, with different temperatures being optimal for different viruses or transcripts (Supp. Figure 1). Most likely, the length and GC content of the RNA–DNA hybrids determines the extent of denaturation at different temperatures as hybrids with a high GC*length score showed higher detection at 95 °C whereas targets with a low GC*length score were better detected at 85 °C (Spearman $r = -0.829$, $p < 0.0001$) (Supp. Table 6). We therefore evaluated the levels of three spiked viruses with increasing GC levels (PDV: 23%, DENV2: 36%, TBEV: 65%) at different post-RT temperatures (4 °C, 85 °C, 95 °C) with or without a pre-RT denaturation step. Figure 5 shows that virus detection and host depletion were highest when a pre-RT 95 °C denaturation step was combined with a 95 °C post-RT denaturation step. However, we also evaluated the use of RNaseH to degrade the RNA in RT hybrids as an alternative to enable efficient second strand synthesis but independent of the hybrid's GC content. Notably, combining pre-RT denaturation (95 °C for 1 min) with the use of a thermostable RNaseH post-RT (50 °C for 20 min), showed superior performance over the combination with post-RT denaturation (95 °C, 5 min). Moreover, combining all three steps, i.e. pre-RT, post-RT, and RNaseH yielded even better results with increased depletion and mostly higher levels of TBEV (Supp. Figure 2). To decrease protocol complexity, we finally evaluated the effect of adding RNaseH to the Klenow reaction at 37 °C. Although possible, this leads to a decrease in virus enrichment and host depletion (Supp. Figure 2).

Development of custom barcoding primers

To multiplex samples on the sequencer, Oxford Nanopore's native barcodes can be ligated to the SISPA PCR products after end-prep and dA-tailing of the dsDNA. However, this method is time consuming and expensive. More importantly, handling high quantities of amplified, unbarcoded DNA poses an intra-experimental contamination risk for clinical diagnostics. Therefore, we investigated whether the Nanopore barcode sequences could be incorporated into the dsDNA during PCR as an overhang attached to the solB primer. To this end, we had to adapt the barcodes_masked.fasta file and include a barcode_arrs_cust.toml file (Supp. data) for custom barcodes to be recognized by Guppy/Dorado. In our first

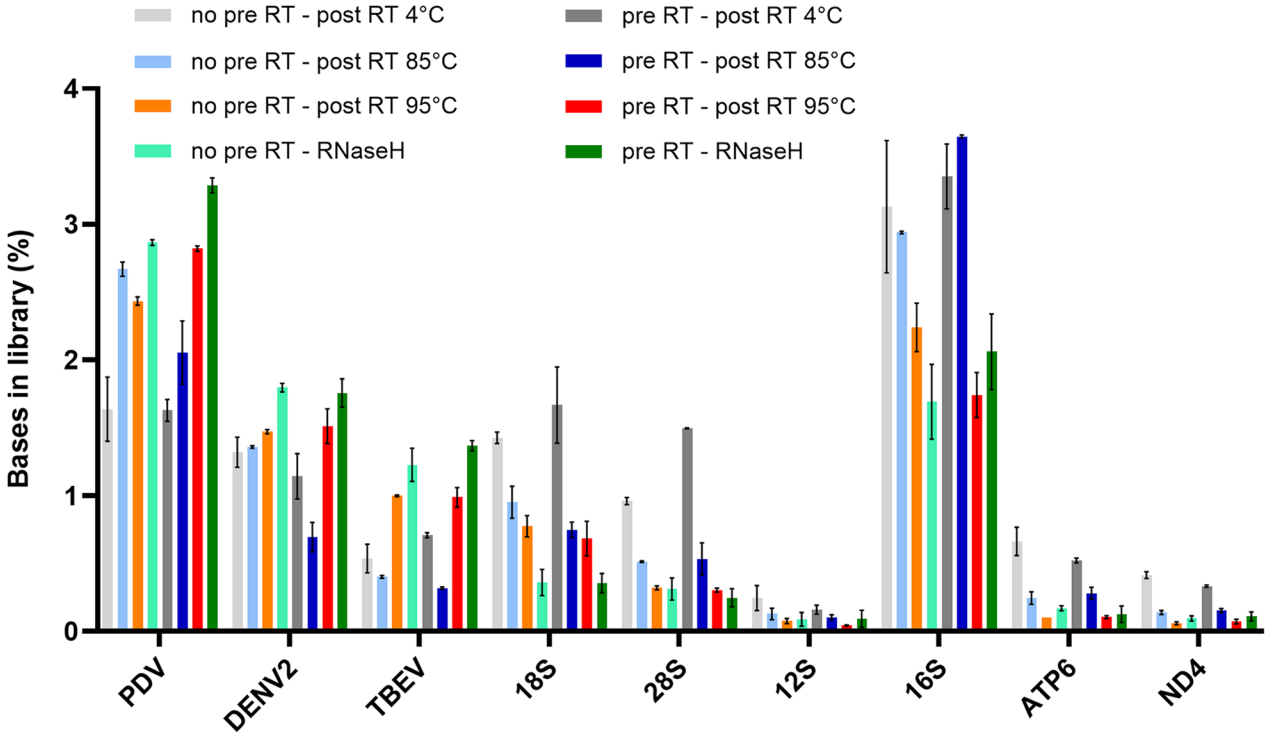


Fig. 5 Effect of pre-RT (reverse transcription) denaturation (1 min 95 °C) and post-RT denaturation (5 min at 85 °C or 95 °C) on the detection of the spiked single-stranded RNA viruses DENV2 (Dengue virus serotype 2), PDV (Phocine Distemper Virus), and TBEV (Tick Borne Encephalitis virus), and on the depletion of human ribosomal RNAs 12S, 16S, 18S, 28S, and transcripts ATP6 (ATP synthase membrane subunit 6) and ND4 (NADH dehydrogenase subunit 4) in metagenomic sequencing libraries of TURBO + DASH-depleted human serum samples. A 50 °C, 20 min RNaseH incubation following RT was also evaluated as a GC-independent way to open up the cDNA for second strand synthesis by degrading the RNA in the RT hybrids

experiments, we simply attached ONT's PCR barcode sequences to the solB primer to keep the length of the overhang as short as possible. Multiple combinations of barcode and solB sequence were tested as primers and only those with similar PCR yields to solB primer amplification were retained. However, we found most reads (median 72.9%, n = 2) to end up as unclassified after demultiplexing due to the barcodes being truncated at the 3´end. Opting for ultramer synthesis of the primers, but primarily the addition of three nucleotides at the 5´ end of each primer, resulted in levels of unclassified reads (median 21.2%, IQR 18.6—25.0, n = 9) comparable to native barcoding (median 21.3%, IQR 19.9–22.9, n = 9). As sequencing happens from 5´ to 3´ end, the DNA strand might slip through the pore if it detaches prematurely from the motor protein resulting in truncated 3´ barcode sequence. Although ONT's PCR primers have seven nucleotides upstream of the barcode sequence, these flanking sequences are not required for successful sequencing since the adapter sequence is long enough to cover the distance between the pore and motor protein. However, since adapter ligation is not 100% efficient, a significant portion of dsDNA molecules might have adapter sequence ligated on only one end. If that were the case, flanking sequences upstream of the barcode would be able to rescue the sequencing of those reads (Fig. 6).

Next we set out to test whether performing SISPA PCR with large overhangs does not significantly impact virus detection. In a sample spiked with DENV2 (at a low concentration) and PDV (at a high concentration), the percentage of viral nucleotides in the library increased when

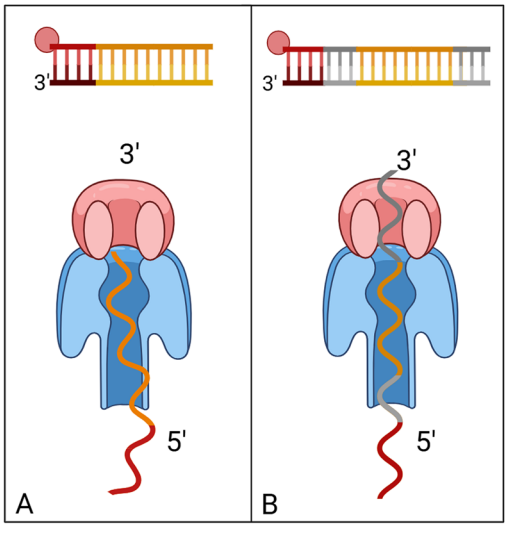


Fig. 6 **A)** Pore slippage might occur when Oxford Nanopore adapter protein and sequence (red) are ligated on only one side of the dsDNA molecule (yellow) resulting in uncontrolled ssDNA translocation. **B)** Providing extra base pairs (grey) that flank the sequence of interest (yellow), keeps the ssDNA molecule attached to the motor protein thereby preventing pore slippage

sequenced with custom PCR barcodes as opposed to ligated barcodes (Fig. 7A). Concurrently, the detection of certain host transcripts such as ND1, ND2, ND6, etc. decreased while for others, such as UBC (Ubiquitin C) and ACTB, it increased. This seemed to be related to the mean read length of the transcript in the library, biasing towards transcripts with larger mean read lengths. We therefore repeated these experiments by PCR barcoding a previously sequenced sample for which the read lengths of each target were known. Although diluting the sample prior to PCR might have slightly affected these input read lengths, a strong correlation (Spearman r = 0.877 $p < 0.0001$) could be observed between the read length and the factor by which detection decreased compared to ligation barcoding. During single primer amplification, hairpins can be formed between the complementary ends of each DNA molecule and this will prevent amplification of especially shorter molecules [16]. PCR barcoding probably shifts this length threshold by increasing the length of the complementary sequence and hence the temperature at which stable hairpins can be formed. Although losing short host RNAs is ideal, shorter viral genomes or viral segments potentially generate shorter DNA molecules after second strand synthesis which would then be at risk of being lost. When spiking RVF, we did indeed see small reductions for the 1689 bp S and 3877 bp M segment (respectively 0.9 × and 0.93 × the detection obtained by ligation barcoding) but an increase in detection for the 6333 bp L segment (1.14×) (Supp. Figure 3A). When spiking Influenza B, which has 8 segments from 2369 bp down to 1097 bp, we observed either an increase or a decrease for half of its segments ranging from 0.82 to 1.20× (Supp. Figure 3B). Finally, we found that the use of PCR barcodes comes at the expense of a 5 × reduced PCR yield (Fig. 7B). but that it also results in 1.2 × longer viral sequence reads most likely due to the hairpin formation (Fig. 7C). Longer sequence reads are preferred as they are more specific and hence facilitate downstream genome assembly. Overall, our observations support the use of PCR barcoding over ligation-based barcoding.

Optimizing DASH-based host depletion

After second strand synthesis with Klenow, the reactions undergo a beads cleanup prior to DASH to remove the buffers and enzymes and concentrate the sample. This cleanup is laborious, increases the risk of RNase contamination into the DASH reactions and most importantly, can be a source of pre-PCR DNA contamination between samples. We therefore tested whether it is possible to perform DASH directly in a 50 µl reaction on unwashed second strand reactions after heat inactivation of the Klenow enzyme. We found that when extra albumin is added to these reactions, direct DASH

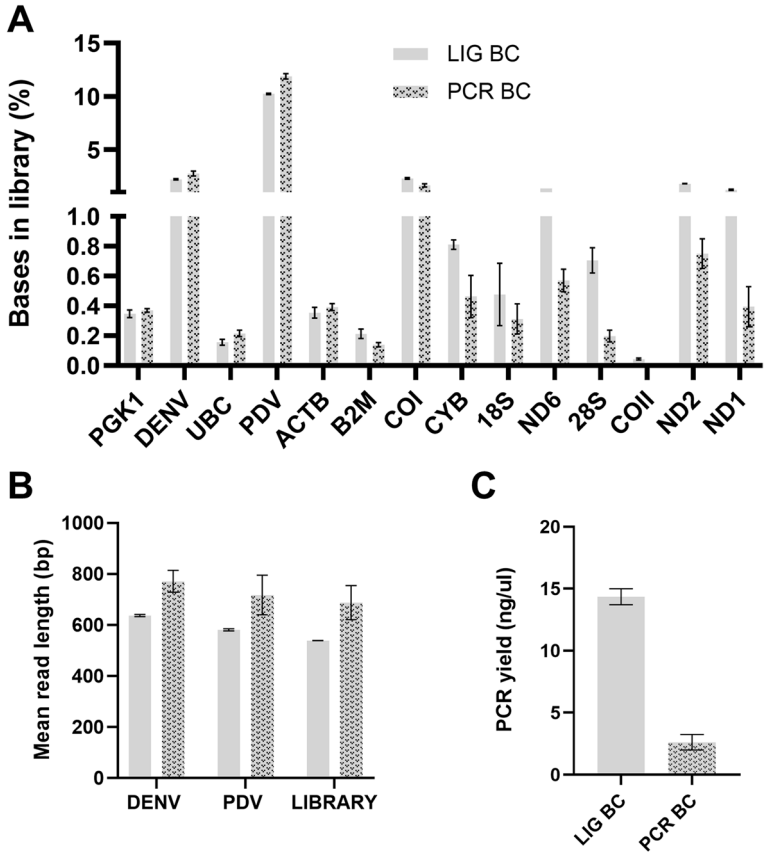


Fig. 7 **A)** The effect of attaching sample barcodes through post-PCR TA ligation (LIG BC) versus incorporating barcodes through primer overhangs during PCR (PCR BC) on the detection of jointly spiked PDV (Phocine distemper virus) and DENV (Dengue virus serotype 2), non-depleted cytoplasmic transcripts PGK1 (Phosphoglycerate kinase 1), UBC (Ubiquitin C), ACTB (Actin beta), B2M (β2 microglobulin), mitochondrial transcripts ND1/ND2/ND6 (NADH dehydrogenase subunit 1/2/6), COI/COII (Cytochrome c oxidase subunit I/II), and CYB (cytochrome B), in metagenomic sequencing libraries of TURBO + DASH-depleted human serum samples. Transcripts were ordered decreasingly from left to right according to their mean read length in the library using ligation barcoding. **B)** the effect of PCR barcodes on the mean read length for each virus and of the whole library. **C)** The effect of PCR barcodes on PCR yield

shows similar viral sensitivity though at the expense of a 3.7 × and 3.6 × decrease in respectively 18S and 28S depletion (Fig. 8 A left). 16S levels also increased several fold when performing direct DASH but are not depicted since its low abundance meant we could not reliably compare with standard DASH levels. Due to the larger reaction volumes, direct DASH also consumes 2.5 × more sgRNAs, Cas9, and proteinase K of which the costs aren't fully compensated for by saving on one beads cleanup. On the other hand, not washing leads to a desirable increase in read lengths (Fig. 8A right).

Because of the loss of depletion when performing direct DASH, we tested if we could improve depletion by varying the sgRNA and Cas9 concentrations. A 3 × Cas9 concentration increased depletion significantly while seemingly also increasing the detection of PDV and non-depleted transcripts (Supp. Figure 4) whereas decreasing the sgRNA concentration exerted no effect (data not shown). This suggests that, relative to the amount of sgRNAs, we were not adding enough Cas9. When an unwashed 50 µL was compared against a washed 20 µL reaction, with a 3 × Cas9 concentration, the loss in depletion we observed earlier was less severe at 2.8 × for 18S and 2.6 × for 28S, again with no effect on PDV levels (Fig. 8B). Even when we spiked PDV at lower levels than 18S, we observed no significant impact of the decreased depletion levels on virus detection (data not shown).

Thus, as performing DASH directly without cleanup looked promising, we then tested if the Klenow reaction could be performed in 30 µL instead of 40 µL. This would allow for a 40 µL instead of a 50 µL DASH reaction and hence save on sgRNAs, Cas9, beads, and proteinase K. Interestingly, performing smaller Klenow reactions resulted in higher virus and human transcript detection and better depletion though probably causing a small decrease in read lengths (Supp. Figure 5). This effect seemed to be mostly independent from DASH as a washed 20 µL reaction compared to an unwashed 40 µL reaction -both performed after a small Klenow reaction- (Fig. 8C), yielded similar results as observed between a washed 20 µL and unwashed 50 µL reaction (Fig. 8B),

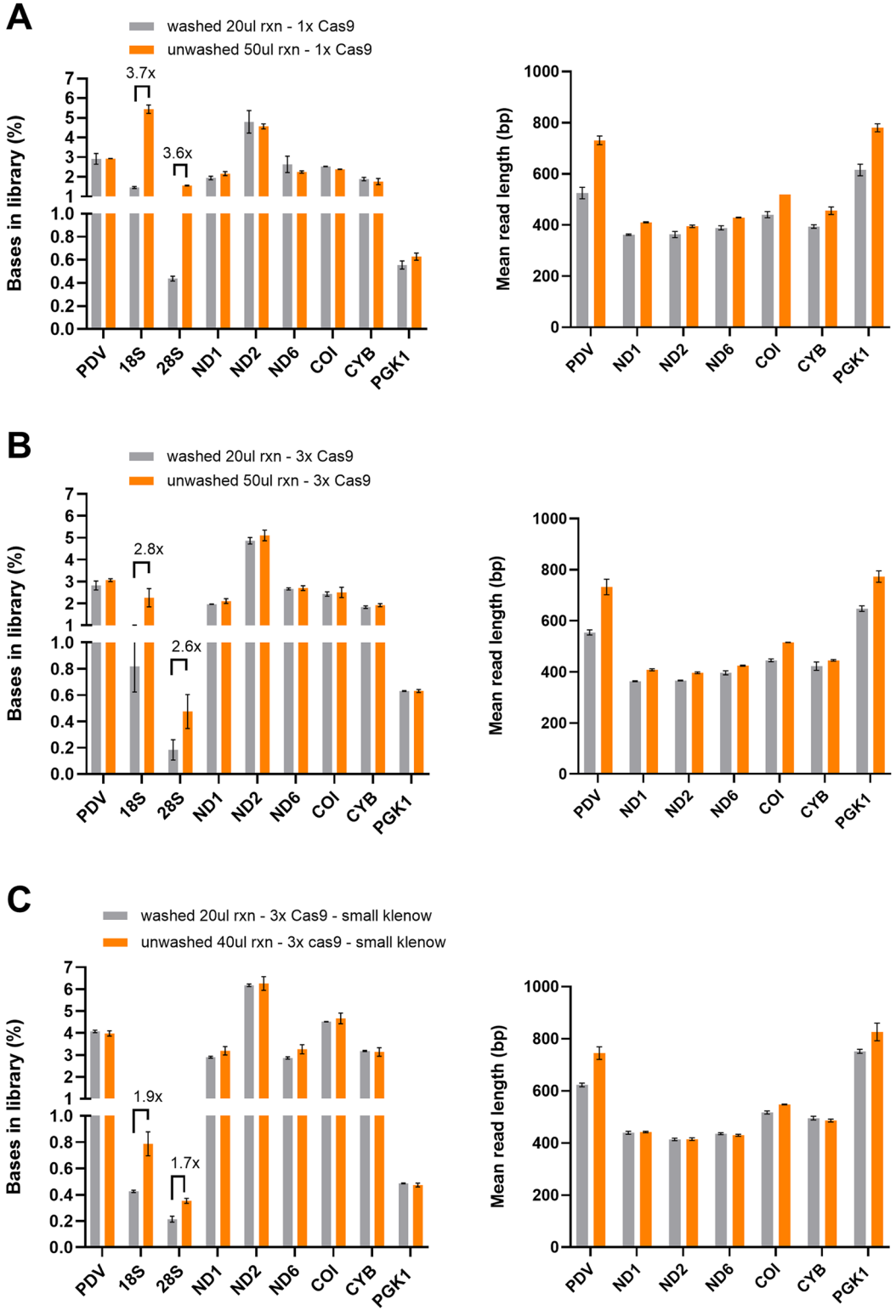


Fig. 8 **A)** The effect of performing DASH directly on unwashed Klenow reactions in 50 µL (unwashed 50 µL rxn) as opposed to performing DASH after a beads cleanup in 20 µL (washed 20 µL rxn), on the detection (left) and read length (right) of spiked PDV (Phocine distemper virus), of non-depleted transcripts PGK1 (phosphoglycerate kinase 1), ND1/ND2/ND6 (NADH dehydrogenase subunit 1/2/6), COI (cytochrome c oxidase subunit I), and CYB (cytochrome B), and of depleted ribosomal 18S and 28S RNA in metagenomic sequencing libraries of TURBO + DASH-depleted human serum samples. **B)** Increasing the Cas9 concentration lowers the loss of depletion observed when using direct DASH. **C)** Performing Klenow reactions in smaller volumes further reduces this loss of depletion

maintaining the loss of depletion albeit this time again a bit smaller at 1.9× (18S) and 1.7× (28S).

Finally, in order to further reduce labor and pre-PCR contamination risk, we tested whether Cas9 inactivation by proteinase K followed by heat inactivation at 95 °C are redundant as beads cleanup is performed immediately after DASH. Figure 9A shows that these steps can potentially be omitted as no huge differences in PDV detection were observed. However, depletion levels of 18S, 28S, and 16S decreased respectively 3.3x, 3.1x, and 2.9x, whereas detection of most non-depleted transcripts (ND1, ND6, COI, PGK1, UBC) also decreased except for COII and B2M. Importantly, when omitting proteinase K digestion, also significant reductions in average read lengths were observed (Fig. 9B).

Viral enrichment by varying beads ratios

After TURBO DNase treatment, a beads clean-up is performed by adding RNase free RNAClean XP beads at a recommended 1.8 × the sample volume. Lowering this ratio would in theory result in longer RNA being bound and as such result in substantial target enrichment since at this stage, contaminating host RNAs are often shorter than viral genomes. However, the exact ratios for size selection with these beads are unknown. We therefore tested the effect of different beads ratios (0.6×, 1×,

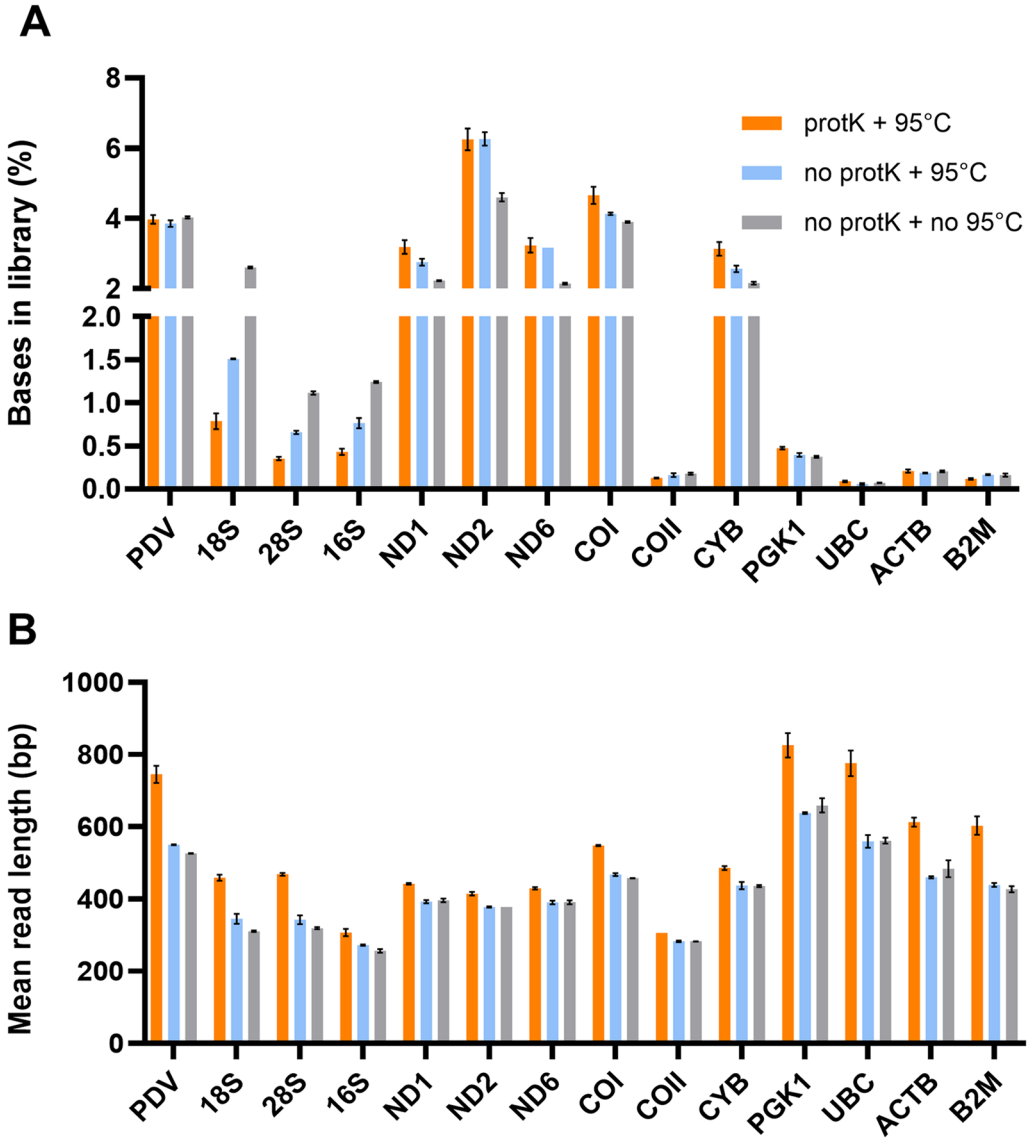


Fig. 9 The effect of omitting Cas9 inactivation by proteinase K (no protK) and subsequent heat inactivation at 95 °C (no 95 °C) on the detection of spiked PDV (Phocine distemper virus), non-depleted transcripts PGK1 (phosphoglycerate kinase 1), ND1/ND2/ND6 (NADH dehydrogenase subunit 1/2/6), COI (cytochrome c oxidase subunit I), CYB (cytochrome B), ACTB (actin beta), and B2M (β2 microglobulin), and on the depletion of ribosomal RNAs 18S, 28S, and 16S in TURBO + DASH-depleted human serum samples. DASH was performed in a 40 µL volume directly on the second strand Klenow reactions

1.5×, 2×) on the detection of SARS-CoV-2 (30 kb), PDV (15.5 kb) and RVF virus (segmented RNA genome, L: 6.5 kb, M: 3.9 kb, S: 1.7 kb). We found that a beads ratio of 0.6 × indeed results in a 1.3 × higher detection of longer viruses such as SARS-CoV-2 and PDV but at the expense of detecting the shorter RVF RNAs with segment S of RVF being 4 × less detected (Supp. Figure 6). However, the spiked RVF levels were rather low for the obtained depth, and this could have impacted our observation. We therefore repeated this experiment with higher RVF levels while also substituting SARS-CoV-2 for Influenza B (IFB), a negative sense ssRNA virus consisting of 8 segments ranging from 2368 to 1096 bp in length. Figure 10 confirms our observation showing 1.4 × better detection of PDV at a 0.6× beads ratio but on average up to 2.3 × worse detection of RVF and IFB segments. Of note, RVF S seemed to be impacted the most with an 8.6 × reduction in detection whereas IFB segments 7 and 8, which are smaller than RVF S, showed only a 2 × reduction. Most likely the RVF sample used for spiking was degraded as it was a clinical sample from Africa obtained under field conditions in contrast to the Influenza B sample which was sampled in Belgium and stored under ideal conditions. This is further in line with a similar drop observed for RVF L and M whereas no drop could be observed for the smaller IFB segments 2 and 3.

Reducing assay costs

Finally, we evaluated whether some of the expensive enzymes in the original SISPA protocol could be replaced by cheaper alternatives. We found Maxima H Minus reverse transcriptase to be a drop-in replacement for Superscript IV at half the price with no significant impact on viral base percentages or on mean virus read lengths. The choice of RT enzyme is of major importance as evidenced by the less balanced performance of FastGene Scriptase II showing significantly lower percentages of spiked virus and the cytoplasmic transcripts, while detecting higher percentages of the mitochondrial transcripts, especially ND1 and ND2. Potentially these observations are related to the lower operating temperature (42 °C) of this enzyme compared to Superscript IV or Maxima H Minus (50 °C) (Fig. 11A).

Also the use of Klenow Fragment (3´– > 5´ exo-) as a cheaper alternative to Sequenase 2.0 showed slightly better results with regard to the percentage of spiked virus detected and similar viral read lengths (Fig. 11B). However in literature this enzyme is used for one hour as opposed to 16 min for Sequenase 2.0. In an attempt to shorten the protocol length, we tested a shorter Klenow incubation time of 30 min and found that this gave reasonable performance with no differences in read length (data not shown), although it did impact the levels of virus, transcripts, and rRNAs in different ways. However, 60 min resulted in higher levels for 10 out of 16 RNAs that we evaluated and lower levels of rRNAs (Supp. Figure 7).

Furthermore, for the SISPA PCR, we tested two cheaper polymerases (LongAmp Taq (2×) and TaKaRa Taq LA) and found that TaKaRa Taq LA is the best replacement for AccuTaq LA with no significant impact on PDV detection in contrast to LongAmp Taq. However, both TaKaRa and LongAmp appear to skew towards detection of shorter, more abundant transcripts such as

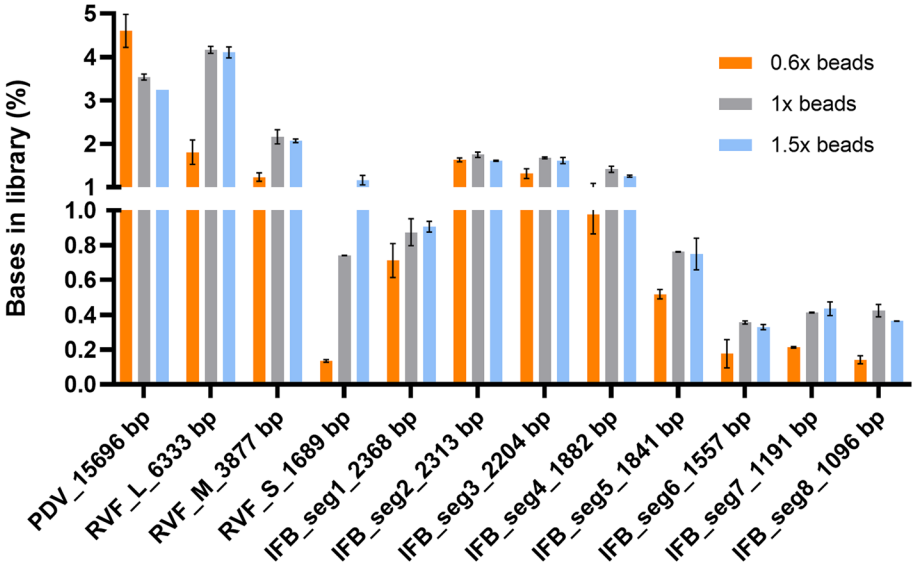


Fig. 10 The effect of using different beads ratios (0.6x, 1x, 1.5x), during cleanup of TURBO DNAse treated RNA, on the detection of jointly spiked IFB (Influenza B, 8 segments) virus, PDV (Phocine distemper virus), and RVF (Rift Valley Fever, 3 segments) virus in metagenomic sequencing libraries of TURBO + DASH-depleted human serum samples. Genome sizes are depicted in base pair (bp)

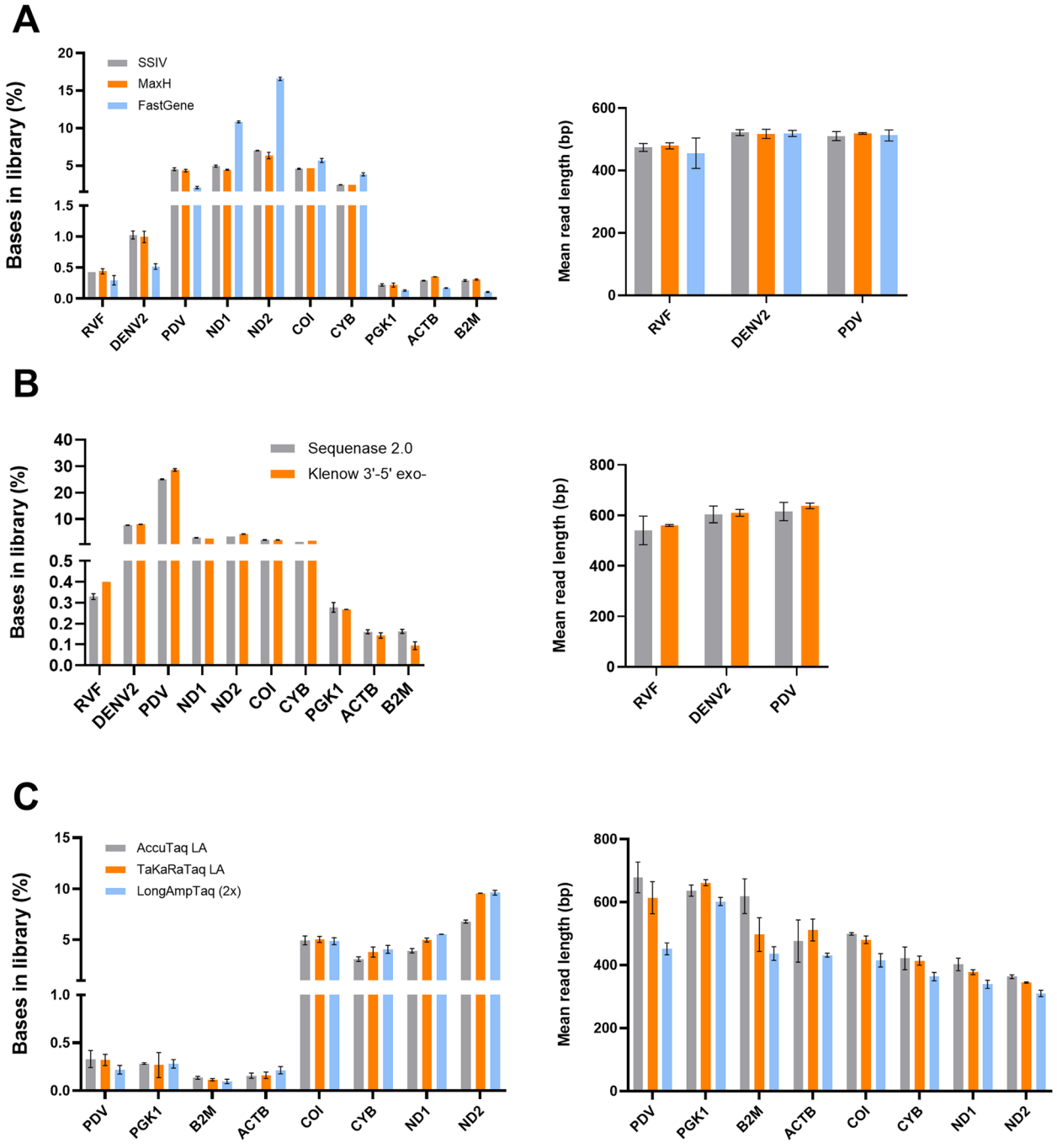


Fig. 11 **A)** The effect of different reverse transcriptase enzymes (Superscript IV, Maxima H Minus, FastGene Scriptase II), **B)** of different second strand synthesis enzymes (Klenow 3´–5´ exo- and Sequenase 2.0), and **C)** of different polymerases (AccuTaq LA, LongAmpTaq (2x), and TaKaRa Taq LA) on the detection (left) and mean read length (right) of jointly spiked RVF (Rift Valley Fever virus, 3 segments), PDV (Phocine distemper virus), DENV2 (Dengue virus serotype 2) and non-depleted transcripts COI (cytochrome c oxidase subunit I), CYB (cytochrome B), ND1/2 (NADH dehydrogenase subunit 1/2), ACTB (actin beta), B2M (β2 microglobulin), and PGK1 (phosphoglycerate kinase 1) in metagenomic sequencing libraries of TURBO + DASH-depleted human serum samples

ND1, ND2, and CYB. With regard to read lengths, LongAmp generated significantly shorter reads whereas TaKaRa decreased read length only slightly (Fig. 11C).

We then tested out Macherey Nagel's (MN) NucleoSpin virus kit, which comes at half the price of Qiagen's viral RNA mini kit, includes proteinase K, and is faster due to the columns accepting higher loading volumes. We first evaluated this by performing RT-qPCR on the extracted nucleic acids for the RNA viruses PDV and DENV, for the rRNAs 18S and 16S, and for the DNA virus mpox.

As expected, Qiagen extracted less mpox DNA then MN (Δ0.74 Ct), likely due to the addition of proteinase K digestion. However, Qiagen did significantly extract more PDV (Δ1.9 Ct) and DENV (Δ1.7 Ct) but also more 18S (Δ1.7 Ct) 16S and (Δ0.6 Ct) (Supp. Figure 8). However, when we increased the amount of lysis buffer and ethanol during the MN extraction, higher PDV, DENV, but also 18S, and 16S yields were obtained reducing the difference between the two kits to respectively 0.6, 0.7, 1.3, and 0.29 Cts (Supp. Figure 8). This would suggest that although MN extracts less virus in absolute terms, it might extract more relative to the amount of unwanted rRNA. Indeed, when we compared sequence data instead of qPCR data, the MN kit seemed to outperform Qiagen's kit showing higher virus and transcript levels, although with the exception of PGK1 and ACTB and with increased -but still acceptable- levels of 16S rRNA (Fig. 12).

Finally, the introduction of custom PCR barcodes resulted in significant cost reductions as it requires less of the NEBNext® Ultra™ II End Repair/dA-Tailing Module and due to lower purchasing costs when ordering from an oligonucleotide manufacturing company directly.

Benchmarking the updated protocol SISPA 2.0

Based on the results above, a new SISPA 2.0 protocol was developed as described in the methods and in Fig. 1. A detailed cost analysis for running 24 samples per minION flow cell, revealed that SISPA 2.0 (59 EUR/sample) was 2.5 × cheaper than the original SISPA combined with Qiagen's FastSelect for rRNA depletion (146 EUR/sample) (Supp. Table 7). We then benchmarked the protocol's sensitivity by sequencing serial dilutions in serum of an HIV stock with known copy number. Figure 13A shows that, at roughly 0.6 Gb per sample (equivalent to 30 samples per flow cell), SISPA 2.0 can recover near full-genome HIV sequence with a minimum depth of 20 × down to a dilution of 5000 copies/ml. Detection sensitivity of SISPA 2.0 is in the qPCR range with HIV being consistently detected at the lowest concentration of 100 copies/ml. In comparison, the original SISPA with or without FastSelect depletion is respectively 7.7 × and 32 × less sensitive (Fig. 13B). We further compared SISPA 2.0 to SMART-9N, a well-known protocol in the field of viral metagenomics. This protocol is a simpler variant of SISPA where single-stranded cDNA molecules are tagged on both ends during RT through the use of a template switching oligo [17]. As this excludes the use of DASH, which acts on ds cDNA, we combined SMART-9N with Qiagen's FastSelect to enable a fair comparison with SISPA 2.0. We found that SISPA 2.0 has a 2.7 × higher detection sensitivity (Fig. 13C).

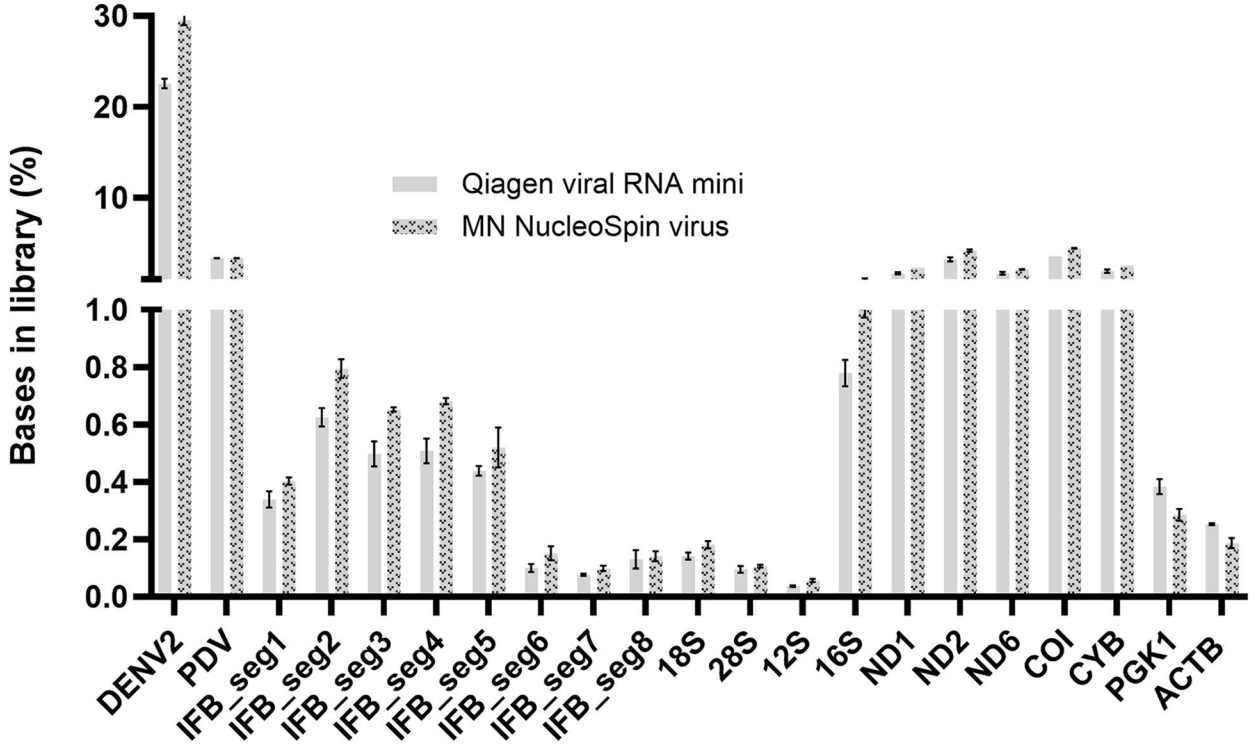


Fig. 12 The effect of different RNA extraction kits (Qiagen viral RNA mini kit vs. Macherey Nagel (MN) NucleoSpin virus) on the detection of jointly spiked IFB (Influenza B, 8 segments) virus, PDV (Phocine distemper virus), DENV2 (Dengue virus serotype 2), depleted ribosomal RNAs 18S, 28S, 12S, and 16S, and non-depleted transcripts COI (cytochrome c oxidase subunit I), CYB (cytochrome B), ND1/2 (NADH dehydrogenase subunit 1/2), ACTB (actin beta), and PGK1 (phosphoglycerate kinase 1) in metagenomic sequencing libraries of TURBO + DASH-depleted human serum samples

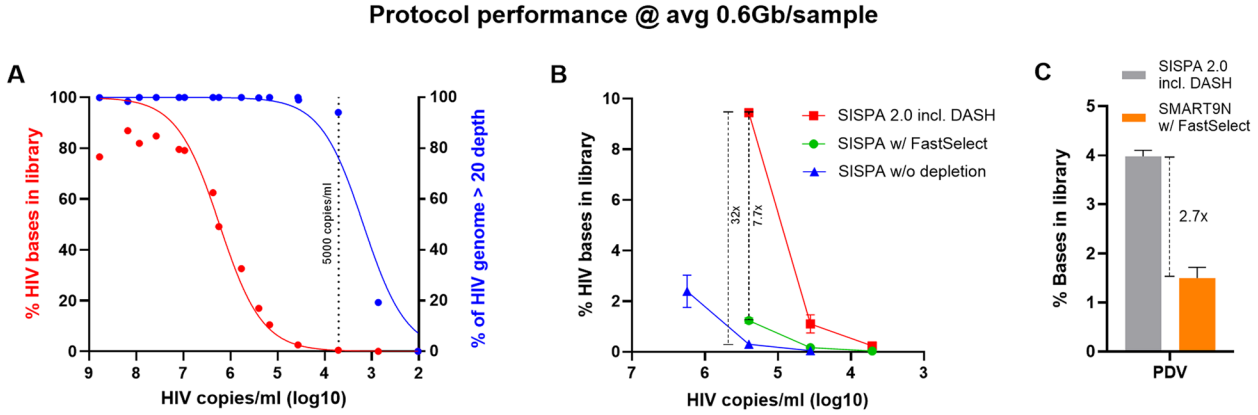


Fig. 13 Benchmarking SISPA 2.0 at a sequencing depth of approx. 0.6 Gb per sample **A**). The percentage of HIV bases in the sequencing library and the percentage of the HIV genome covered at a minimum depth of 20× for a dilution series of HIV in human serum. **B)** The percentage of HIV bases in the library compared between the old and new SISPA protocol. **C)** The percentage of PDV bases in the library between SISPA 2.0 and SMART-9N combined with FastSelect depletion

Discussion

Sequence independent single primer amplification (SISPA) has shown to be a highly effective pathogen metagenomics method although its sensitivity is limited due to the high degree of host contamination in many sample types. In this study, we adapted and integrated the CRISPR/Cas9-mediated DASH technique into a SISPA workflow to deplete human rRNAs as well as other abundant transcripts in serum samples, while leaving intact viral RNAs (with the exception of bacteriophages) and 16S rRNA of bacteria commonly detected in serum. Although, with the current panel, some off-target cleavage may occur in cytoplasmic rRNAs of eukaryotic serum pathogens, this should only impact detection sensitivity in samples with low pathogen load, as the remainder of their substantially larger transcriptomes would remain unaffected and would benefit from a sensitivity boost due to rRNA depletion.

The introduction of DASH into the SISPA protocol not only resulted in a large increase in sensitivity compared to the standard SISPA but also outperformed Qiagen's FastSelect, a commercial host depletion method, at a 5–10 × lower cost. More importantly, this approach is highly flexible as it can be tailored to any host species through the in-house design of sgRNAs that target the most abundant and unwanted RNAs depending on the sample type. In addition, performing DASH is not only more effective, but also much more reliable than using pre-extraction nucleases. The latter can not only degrade unprotected viral RNAs in clinical samples but would preclude detection of cell-free RNA from cellular pathogens, while leaving protected host nucleic acids intact that would otherwise be accessible post-extraction. Edridge et al. made a similar observation for DNA viruses in a DNA workflow [18]. Finally, for DASH to be effective, the use of TURBO DNase remains essential, suggesting an overwhelming presence of non-targeted, low abundant host DNA in serum samples as also suggested by others [18, 19].

We further show that to detect dsRNA viruses, it is absolutely essential to incorporate a 95 °C denaturation step prior to RT. This denaturation step is often overlooked and the lack thereof might explain why Buddle et al. [20], who compared different metagenomic sequencing approaches, untargeted ONT sequencing using the SMART-9N method, showed little to no detection capability for Orthoreovirus. Much more pronounced though is the effect of both pre- and post-RT temperatures on detection sensitivity. Although the underlying mechanism remains unclear, this seemed to be related to the removal of RNA from the RNA–DNA hybrids generated during RT, thereby facilitating primer annealing for second strand synthesis.

To enable sufficient throughput, samples for metagenomics are often processed in batch using tube strips or even 96-well plates. The original method involved the ligation of barcodes to the SISPA DNA so that reads obtained from different samples can be demultiplexed. However, the handling of highly amplified PCR products in batch, can lead to substantial intra-experimental contamination, especially if some samples contain a high concentration of virus. By incorporating the barcodes as overhangs during the SISPA PCR, we prevented post-PCR contamination in our workflow while also obtaining longer read lengths and additional virus enrichment. The latter is potentially caused by more stable hairpins formed during PCR as the extra barcode sequence at the 5′ will be complementary to the sequence on the 3′ end [16].

Although for most viruses viral enrichment can be obtained through lowering the bead ratio during RNA cleanup, we recommend against this when performing

agnostic viral metagenomics since lower ratios will result in a substantial loss of small viruses or viral segments (e.g. the 0.8–1.0 kb NS segment of Influenza). Nonetheless, this method could be useful in a diagnostic context when sequence data of smaller segments is not required or, alternatively, when performing metagenomics on samples with known pathogens of sufficient genome length, providing sample storage did not lead to RNA fragmentation.

As performing DASH is rather laborious, we further investigated simplifying the protocol. Ideally, a bead cleanup after second strand synthesis is avoided as the extra handling could be a source of pre-PCR contamination between samples. We found that when correcting the buffer with albumin and NaCl, DASH can be used directly on the second strand reactions as no effect could be observed on virus detection. Although a cleanup will always result in better depletion of rRNAs, by tripling the Cas9 concentration and performing smaller Klenow reactions, we managed to reduce the loss of depletion, with 18S levels remaining below 1%. The fact that tripling the Cas9 concentration significantly improved depletion whereas lowering the sgRNA concentration did not, isn't surprising. After all, during ribonucleoprotein complex formation each Cas9 molecule binds to one sgRNA, so minimally a 1:1 molar ratio would be required. In fact, in literature a 3 × molar excess of sgRNAs to Cas9 is often prescribed. Yet, in the protocol of Loi et al., on which we initially based ourselves, a 147× excess of sgRNAs was used, more than sufficient to bind to a 3 × increase in Cas9 molecules thus directly resulting in 3 × more catalytic ribonucleoprotein complex being added to the final reaction. Ideally, different Cas9:sgRNA ratios are empirically tested when designing a protocol. Furthermore, we found that, although possible, dropping proteinase K-based Cas9 inactivation, is not recommended as this negatively impacts depletion levels, read lengths, and detection levels of multiple non-depleted human transcripts.

Lastly, we attempted to lower the assay cost without compromising on sensitivity. This resulted in a switch from Superscript IV to Maxima H Minus, from sequenase 2.0 to Klenow 3´–5´ exo-, from 40 µL to 30 µL Klenow reactions, and from ONT's native barcoding to custom PCR barcoding. Although we find the PCR to yield slightly better results with AccuTaq LA polymerase, a switch to TaKaRa LA polymerase can be justified.

When benchmarking SISPA 2.0, we found our protocol to be significantly more sensitive, roughly 7.7 × more compared to FastSelect + SISPA, and 2.7 × more compared to FastSelect + SMART-9N. But at a detection sensitivity of ≤ 100 HIV copies/ml, SISPA 2.0 also compares very favourably to other published Nanopore protocols that report detection sensitivities ranging from 600 to 70 000 copies/ml though this is hard to compare due to the different definitions of detection, viruses, sample types, sequencing depths, and depletion methods used [20–23]. Moreover, unlike hybridization capture–based protocols that positively select for thousands of predefined viruses (e.g., Twist's Comprehensive Viral Research Panel or VirCapSeqVERT), our protocol represents a true untargeted metagenomic approach, relying on negative selection (i.e., host depletion). As a result, the protocol is well-suited for novel virus discovery because, due to sequence dissimilarity with human rRNAs, it is highly unlikely that a new virus or a variant evolving from known sequences would suddenly acquire dozens of off-target sites across its entire genome. Furthermore, in contrast to rapidly evolving pathogens, vertebrate hosts do not evolve on relevant timescales, eliminating the need for frequent panel updates.

Finally, our study had several limitations. First, due to the high costs of sequencing, we could not test each protocol modification across a serial dilution of spiked virus. However, for viral detection at higher prevalences (> 10%), this is less critical as genome coverage will be more than sufficient. We therefore focused on evaluating protocol changes at lower viral prevalences, around 1% of the library. A more important, second limitation of our study is the occurrence of possible interactions between protocol modifications. Although two changes might have a positive effect in isolation, their combination could yield different outcomes. As testing all possible combinations would make this research cost prohibitive, we built upon previous improvements and prioritized combinations where theoretical interactions could be suspected. Nevertheless, it remains possible that some individually tested protocol changes may have offset each other. For example, although overnight DASH incubation initially showed slightly better depletion than a 2.5 h incubation, this advantage disappeared when tripling the Cas9 concentration. Despite such nuances, final benchmarking incorporated all changes and demonstrated superior performance for the currently presented combination. A third limitation is that we did not originally design the presented DASH panel for cellular pathogen diagnostics, as serum -being cell-free- is not the most suitable sample type for this purpose. We are currently working on a revised, more sensitive human panel with broader applicability in cellular samples, designed to deplete human ribosomal and mitochondrial RNAs while preserving the 18S and 28S rRNAs of most eukaryotic pathogens. Lastly, following protocol development, a formal clinical validation of our assay will be required. This will include determining the 95% limit of detection, assay sensitivity and specificity, and positive and negative percent agreement, and will necessitate additional budget as well as a set of

conventionally diagnosed samples infected with a variety of pathogens for discrepancy analysis.

Conclusion

The metagenomic approach presented here, is able to detect microbial RNA with PCR-range sensitivity in human serum samples but can be adapted -in-house- to any species and sample type. Due to its low cost, it can be leveraged globally in both human pathogen diagnostics as well as zoonotic disease research.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12268-4.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements

We thank Dr. Pieter Monsieurs for his assistance with bioinformatic issues as well as Prof. Peter Delputte, Marjan van Esbroeck, Dr. Anselme Shyaka, and Dr. Pascal Lutumba for the generous donations of Influenza B, Rotavirus, Rift Vally Fever virus, and clinical samples.

Authors' contributions

Conceptualization (PS, KA, FF, JM), Analysis (PS, EVV), Resources (KA, JM, PS, FF), Methodology (PS, EVV), Investigation (PS, EVV), Analysis (PS, EVV), Funding acquisition (PS, KA, JM, FF), Writing (PS, EVV), Supervision (KA), Review & edit (KA, JM, FF, PM).

Funding

PS is a member of the Institute of Tropical Medicine's Outbreak Research Team which is financially supported by the Department of Economy, Science and Innovation (EWI) of the Flemish government who also funded most of this work under 'ITM's Pump Priming Projects programme in addition to support from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health, USA, under Award Number U01AI151378. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data availability

All relevant data are in the manuscript and supporting information.

Declarations

Ethics approval and consent to participate

Clinical samples were collected in view of diagnosis and standard of care following the patient's informed consent. Sequencing was approved by the Institutional Ethics Committees of the University Cayetano Heredia (N° 211163), the Institute of Tropical Medicine (N° 1700), and the University Hospital of Antwerp (N°5651).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 July 2025 / Accepted: 22 October 2025
Published online: 12 February 2026

References

1. Global genomic surveillance strategy for pathogens with pandemic and epidemic potential, 2022–2032. Available from: https://www.who.int/publications/i/item/9789240046979. Cited 2025 Jun 6.
2. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20(6):341–55. https://doi.org/10.1038/s41576-019-0113-7.
3. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015. https://doi.org/10.1186/s13073-015-0220-9.
4. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, Crook A, Carter D, Osborne J, et al. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. Eurosurveillance. 2018;23(50). Available from: https://www.eurosurveillance.org/content/https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800228
5. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science. 2019;363(6422):74–7.
6. Conceição-Neto N, Yinda KC, Van Ranst M, Matthijnssens J. NetoVIR: Modular Approach to Customize Sample Preparation Procedures for Viral Metagenomics BT - The Human Virome: Methods and Protocols. In: Moya A, Pérez Brocal V, editors. New York, NY: Springer New York; 2018. p. 85–95. Available from: https://doi.org/10.1007/978-1-4939-8682-8_7
7. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, et al. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol. 2016;17(1):41. https://doi.org/10.1186/s13059-016-0904-5.
8. Loi DSC, Yu L, Wu AR. Effective ribosomal RNA depletion for single-cell total RNA-seq by scDASH. PeerJ. 2021;9:e10717.
9. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. 2016;44(W1):W272–6.
10. Johnson BW, Russell BJ, Lanciotti RS. Serotype-specific detection of dengue viruses in a fourplex real-time reverse transcriptase PCR assay. J Clin Microbiol. 2005;43(10):4977–83.
11. Panning M, Grywna K, Van Esbroeck M, Emmerich P, Drosten C. Chikungunya fever in travelers returning to Europe from the Indian Ocean region, 2006. Emerg Infect Dis. 2008;14(3):416–22.
12. Huits R, De Smet B, Grard G, Eggermont K, Minto-Bain C, Jess N, et al. Detection of Zika virus replication in human semen by reverse-transcription polymerase chain reaction targeting of antisense ribonucleic acid. J Infect Dis. 2020;222(2):319–23.
13. Al Amir Dache Z, Otandault A, Tanos R, Pastor B, Meddeb R, Sanchez C, et al. Blood contains circulating cell-free respiratory competent mitochondria. FASEB J. 2020;34(3):3616–30.
14. Liao PC, Bergamini C, Fato R, Pon LA, Pallotti F. Isolation of mitochondria from cells and tissues. Methods Cell Biol. 2020;155:3–31.
15. Enderle D, Spiel A, Coticchia CM, Berghoff E, Mueller R, Schlumpberger M, et al. Characterization of RNA from exosomes and other extracellular vesicles isolated by a novel spin column-based method. PLoS One. 2015;10(8):e0136133. https://doi.org/10.1371/journal.pone.0136133.
16. Bayega A, Oikonomopoulos S, Wang YC, Ragoussis J. Improved Nanopore full-length cDNA sequencing by PCR-suppression. Front Genet. 2022;13. Available from: https://www.frontiersin.org/journals/genetics/articles/https://doi.org/10.3389/fgene.2022.1031355
17. Claro IM, Ramundo MS, Coletti TM, da Silva CAM, Valenca IN, Candido DS, et al. Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing. Wellcome Open Res. 2021;6:241.
18. Edridge AWD, Deijs M, van Zeggeren IE, Kinsella CM, Jebbink MF, Bakker M, et al. Viral Metagenomics on Cerebrospinal Fluid. Genes (Basel). 2019 Apr;10(5).
19. Dynerman D, Lyden A, Quan J, Caldera S, McGeever A, Dimitrov B, et al. Designing and implementing programmable depletion in sequencing libraries with DASHit. bioRxiv. 2020;2020.01.12.891176. Available from: http://biorxiv.org/content/early/2020/01/13/2020.01.12.891176.abstract
20. Buddle S, Forrest L, Akinsuyi N, Martin Bernal LM, Brooks T, Venturini C, et al. Evaluating metagenomics and targeted approaches for diagnosis and surveillance of viruses. Genome Med. 2024;16(1):111. https://doi.org/10.1186/s13073-024-01380-x.
21. Alcolea-Medina A, Alder C, Snell LB, Charalampous T, Aydin A, Nebbia G, et al. Unified metagenomic method for rapid detection of microorganisms in

clinical samples. Commun Med. 2024;4(1):135. https://doi.org/10.1038/s43856-024-00554-3.
22. Lewandowski K, Xu Y, Pullan ST, Lumley SF, Foster D, Sanderson N, et al. Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. J Clin Microbiol. 2019. https://doi.org/10.1128/jcm.00963-19.
23. Gauthier NPG, Chan W, Locher K, Smailus D, Coope R, Charles M, et al. Validation of an Automated, End-to-End Metagenomic Sequencing Assay for Agnostic Detection of Respiratory Viruses. J Infect Dis. 2024;230(6):e1245–53. Available from: https://doi.org/10.1093/infdis/jiae226

Publisher's Note