Journal of Clinical Virology Plus 2 (2022) 100120

Contents lists available at ScienceDirect

Journal of Clinical Virology Plus

journal homepage: www.elsevier.com/locate/jcvp

Improvements in metagenomic virus detection by simple pretreatment methods

Anna S. Fomsgaard[a,b,*], Morten Rasmussen[a], Katja Spiess[a], Anders Fomsgaard[a], Graham J. Belsham[b], Jannik Fonager[a]

[a] Department of Virus & Microbiological Special Diagnostics, Statens Serum Institut, 5 Artillerivej, DK-2300 Copenhagen, Denmark
[b] Department of Veterinary and Animal Sciences, University of Copenhagen, 4 Stigboejlen, 1870 Frederiksberg, Denmark

ARTICLE INFO

Keywords:
SARS-CoV-2
HPV
Metagenomics
Point of care detection
Nanopore sequencing

ABSTRACT

Early detection of pathogens at the point of care helps reduce the threats to human and animal health from emerging pathogens. Initially, the disease-causing agent will be unknown and needs to be identified; this often requires specific laboratory facilities. Here we describe the development of an unbiased detection assay for RNA and DNA viruses using metagenomic Nanopore sequencing and simple methods that can be transferred into a field setting. Human clinical samples containing the RNA virus SARS-CoV-2 or the DNA viruses human papillomavirus (HPV) and molluscum contagiosum virus (MCV) were used as a test of concept. Firstly, the virus detection potential was optimized by investigating different pretreatments for reducing non-viral nucleic acid components. DNase I pretreatment followed by filtration increased the proportion of SARS-CoV-2 sequenced reads > 500-fold compared with no pretreatments. This was sufficient to achieve virus detection with high confidence and allowed variant identification. Next, we tested individual SARS-CoV-2 samples with various viral loads (measured as CT-values determined by RT-qPCR). Lastly, we tested the assay on clinical samples containing the DNA virus HPV and co-infection with MCV to show the assay's detection potential for DNA viruses.

This protocol is fast (same day results). We hope to apply this method in other settings for point of care detection of virus pathogens, thus eliminating the need for transport of infectious samples, cold storage and a specialized laboratory.

1. Introduction

Viruses continuously emerge or re-emerge, causing diseases in humans and animals, and sometimes in both, as seen with the zoonotic COVID-19-pandemic. In the stages of initial incidence, it is crucial to be able to detect the disease-causing agent rapidly because early detection can assist interventive measures in order to avoid an extensive outbreak.

Conventional diagnostic tests are often targeted towards a single or a panel of specific pathogens and can be relatively fast and powerful laboratory tools. These assays can be, e.g., PCR (conventional or real-time), serological such as ELISA assays or microarrays that can test for multiple pathogens simultaneously [16]. Although sensitive, these tests are only as good as the suspicion about the disease-causing agent because the primers and/or probes for the molecular assays are targeted towards specific viruses. This can be a problem since different pathogens can produce similar symptoms that can lead to multiple independent tests being required to provide an accurate diagnosis. It can also be time consuming to design and validate primers and/or probes for specific viruses and these can furthermore become obsolete if the agent changes in the targeted region, e.g. as observed for SARS-CoV-2 in the S-gene [23].

Metagenomic next-generation sequencing (mNGS) has the ability to identify pathogens in a hypothesis-free manner compared to targeted strategies. All nucleic acid (NA) in a sample is sequenced and the reads generated can be assigned to different organisms in the given sample, resulting in both host and pathogen being detected simultaneously [20]. There are multiple platforms for mNGS such as Illumina, Ion Torrent, Pacific Biosciences and Oxford Nanopore. These platforms have different properties and, as a user, there are several factors to consider when choosing the platform that is best for a particular purpose, for example read-length, base-calling error rate, sequencing time, price, hands-on steps, etc. Oxford Nanopore sequencing offers long-read generating real-time sequencing of multiple samples on the small portable MinION device. Multiple library preparation kits are available, some with as little as 10 minutes preparation time prior to sequencing. Nanopore technology has been compared extensively with the Illumina technology that produces short reads and has higher accuracy but requires more hands-on steps, is immobile and does not provide real-time analysis. mNGS is not

* Corresponding author.
E-mail address: ANFD@ssi.dk (A.S. Fomsgaard).

https://doi.org/10.1016/j.jcvp.2022.100120
Received 20 March 2022; Received in revised form 26 September 2022; Accepted 24 October 2022

without challenges, e.g. the proportion of the viral genomes can be very small compared to the host-genome, bacterial genomes, plant genomes and other nucleic acids in a given clinical sample [20]. This "viral needle in a metagenomic haystack" has proven troublesome. In attempts to overcome this problem, two main strategies have been used, either virus enrichment or host depletion. Virus NA enrichment may involve some degree of positive virus selection, e.g. probe capture or poly-(A)-selection. However, these methods are targeted, to at least some degree, since the probes are virus specific and not all viral RNAs have poly-(A) tails. In contrast, host depletion involves removing as much non-viral NA as possible. Host depletion can include pretreatment of samples by low-speed centrifugation, ultracentrifugation, filtration, polyethylene glycol precipitation or, depending on the pathogens of interest, the level of DNA or RNA genomes can also be reduced, e.g. using particular nucleases [1,2,6,7,10–13, 19]. Pretreatments have been extensively used but we wanted to investigate which very simple pretreatment steps could make virus detection in a metagenomics assay more efficient and we excluded other well-known pretreatments if they could not be made field deployable as easily.

Here, we present a protocol that includes simple to use sample pretreatments, NA extraction, isothermal random-amplification followed by non-targeted sequencing using the Oxford Nanopore Technology system for real-time identification of viral NA from different clinical swab samples.

During the course of this project, the newly emerged SARS-CoV-2 had been identified in China during late 2019 and developed into a pandemic [21,22]. The purpose of this project was to improve non-targeted point of incidence testing. The COVID-19 pandemic provided clinical samples containing a RNA virus and thus SARS-CoV-2 samples were used as test of concept to assess the ability of different pretreatments to detect and identify viral pathogens. After choosing the best pretreatments, we tested individual samples containing SARS-CoV-2 or human papillomavirus (HPV) for RNA and DNA virus detection, respectively.

2. Method

2.1. Clinical samples

All SARS-CoV-2 samples in this study were oropharyngeal swabs in PBS collected between August and September 2021. The samples were processed and diagnosed using the RT-qPCR assay described by Corman et al. [3] on the same day as they were collected at the Test Center Denmark (TCDK), Statens Serum Institut (SSI). All HPV samples in this study were cervical swabs in PBS. These samples had been frozen at -80˚C and were thawed for this study because it was not possible to obtain fresh non-frozen HPV samples at the time.

2.1.1. Pooled samples

To test the effect of different pretreatments on indistinguishable sample materials, equal volumes of 16 fresh non-frozen SARS-CoV-2 real-time PCR positive samples (median CT-value of 22 IQR=7.7 [3]) were pooled and aliquoted into four portions. Each aliquot received a different pretreatment: no pretreatment, DNase I treated, filtration or DNase I treatment followed by filtration as described in 2.2.1 and 2.2.2. The sample material was then processed through the rest of the protocol from 2.2.3 (see also Supplementary Figure 1).

2.1.2. Individual samples

Fourteen SARS-CoV-2 RT-qPCR positive samples covering a broad range of SARS-CoV-2 input material (CT-values 18 to 36 determined using the assay described by Corman et al. [3]) and a negative clinical sample (with CT > 38) were collected for pretreatment analysis. All positive samples had been typed to be of the Delta variant B.1.617 using the variant-PCR surveillance program from TCDK screening for the L452R substitution in the spike protein coding sequence but with the absence of the deletion H69-70 or the N501Y and the E484K substitutions that have been used to type other variants.

Ten cervical swabs in PBS, were included in the study to test the ability of the method to detect DNA viruses. Eight of the samples were HPV positive, one sample was also positive for molluscum contagiosum virus subtype 1 (MCV-1) and two were negative clinical samples. The samples had been received in 2021 from clinicians to aid diagnosis of patients using the in-house microarray described by Rosenstierne et al. [16].

2.1.3. Ethics

Exemption for review by the ethical committee system and informed consent was given by the Committee on Biomedical Research Ethics - Capital region in accordance with Danish law on assay development projects. Samples were anonymized prior to analysis.

2.2. Pretreatment and extraction

2.2.1. DNase treatment

For DNase treatment, we used the Zymo Research DNase I Set (Zymo Research) essentially according to the manufacturer's recommendations. A down-scaled sample volume of 300 µL was mixed with 37.5 µL DNase I and 37.5 µL DNase I buffer and incubated for 15 minutes at room temperature. Due to the subsequent filtration step (2.2.2.), small sample volumes were diluted to 1 ml with PBS to avoid the sample being lost in the filter's dead-volume. Only the undiluted sample was DNase treated, then after 15 minutes, PBS was added to 1 ml and then filtered directly into the MagNA Pure Lysis and Binding-buffer (MPLB-buffer) (Roche Life Sciences) for NA extraction (2.2.3).

2.2.2. Filtration

For filtration treatment, we used a 5 ml syringe attached to a 0.22 µM Minisart® Syringe Filter (Sartorius Stedim Biotech, France) to process 1 ml of sample. An additional 1 ml of air in the syringe ensured that the liquid otherwise lost in the filter's dead-volume was recovered.

2.2.3. Nucleic acid recovery

NA extraction was performed using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Life Sciences) by adding 1 ml sample directly into 1 ml of MPLB-buffer following the field extraction method of Rosenstierne et al. [15]. Briefly, sample and MPLB-buffer were mixed for lysis, before the magnetic glass particles (MGPs) were added and the contents mixed by gently flipping the tube. Using a small magnet, the MGPs were captured within the tube and washed three times with kit-included buffers in order to release unbound nucleic acids. The MPLB-buffer contains guanidine iso-thiocyanate and detergent; which efficiently inactivates virus and proteins including the added DNase I and thereby avoids the need for heat or EDTA treatment.

2.3. Double-stranded cDNA synthesis and random amplification

For random amplification, we used Whole Transcriptome Amplification (WTA) for SARS-CoV-2 samples and Whole Genome Amplification (WGA) for HPV samples from the REPLI-g Cell WGA & WTA Kit (Qiagen, Hilden, Germany). The kit includes five steps including cell lysis, genomic DNA removal for WTA, reverse transcription or genomic DNA repair for WTA and WGA, respectively, ligation of small gDNA or cDNA fragments and random amplification with the phi29 DNA polymerase. We made three modifications of the protocol: firstly, we omitted the cell lysis step because the samples had been lysed with MPLB-buffer. Secondly, the 10 µL input material required in the REPLI-g protocol remained bound to the MGPs from the NA extraction step as recommended by Rostenstierne et al. [15]. Thirdly, instead of using the oligo dT primers provided in the kit for the reverse-transcription step, we used 20 µM 5′-phosphorylated random hexamers (P-N6) as described earlier [16].

Table 1
Sample pretreatments enhance viral RNA sequence detection efficiency (data after 20 hours of sequencing for quality trimmed reads).

Sample pretreatment	Number of BLAST hits assigned to SARS-CoV-2/Total number of BLAST hits (%)	Coverage of SARS-CoV-2 reference genome (%)	Fold increase in mean number of viral reads per 100,000 reads
No pretreatment	7/22 (31.8)	7.5	1
Filtration alone	34/51 (65.4)	32.8	3.3
DNase I alone	654/687 (95.3)	94.1	36.2
DNase I and filtration	10736/10747 (99.9)	99.3	525

2.4. Library preparation and nanopore sequencing

Between 4 and 12 individual samples were multiplexed using the Rapid Barcoding Kit (Oxford Nanopore Technology). The DNA concentrations of the samples were determined using a Qubit Fluorimeter (Life Technologies) and used to normalize the input material for sequencing libraries. Here, sample concentrations were standardized to contain 400 ng DNA (in 7.5 µL) when possible, incubated with 2.5 µL of the individual barcodes for 1 minute at 30°C and then 80°C, before 10 µL of multiplexed samples were incubated for another 5 minutes at room temperature with the rapid adaptor enzyme. Some samples were included despite not meeting the library input recommendations (400 ng). The protocol suggestion of a bead purification step when multiplexing four or more samples was omitted for faster library preparation with field use in mind. The libraries were loaded and sequenced on R9.4.1 flowcells on the MK1C device (Oxford Nanopore Technology) with default settings, fast base calling enabled and a run length of 1200 minutes (20 hours). To investigate the efficiency and speed of the protocol's ability to detect viral reads the first three hours of sequencing were extensively monitored. DNase/RNase free water was included as a negative control sample.

2.5. Data analysis

2.5.1. Trimming

Reads were de-multiplexed using MinKNOW (version 21.02.1) before being imported to Geneious Prime version 2021.2.2 (https://www.geneious.com) for data analysis. Using the quality trimming plugin tool BBduk2 (version 1.0), the raw reads were quality-trimmed with a minimum quality score of 7 at each end of each read as recommended for Oxford Nanopore reads by Geneious Prime and a read length above 150 nt.

2.5.2. Metagenomic analysis of reads

The trimmed reads were blasted using Virosaurus (version 98, 2020_4.2), a curated offline database made for clinical metagenomics analysis, which contains full-length genome and segment (for segmented viruses) sequences from all known viral pathogens of vertebrates [5]. HIV-1 was excluded to avoid the ethical dilemma resulting from randomized finds, which could not be acted upon. Only hits with an e-value at or below 1e-5, with a minimum length of 100 nt, a pairwise identity of ≥ 90% and ≥ 10 hits were considered valid.

2.5.3. Mapping of reads and virus typing

The reads from hits identified using BLAST were mapped to the relevant reference sequence from Virosaurus using MiniMap2 for long-read assembly [9] in Geneious Prime (version 2021.2.2.) using default parameters. For typing SARS-CoV-2 reads, a consensus sequence constructed manually from 15 randomly selected sequences of the then most commonly circulating Delta (B.1.617.2) variants in Denmark was used as the reference genome sequence for the reads obtained after a full sequence run of 20 hours. Consensus sequences were generated with majority vote SNP calling and incorporation into the consensus sequence and typed using the online COVID-19 lineage assigner, Pangolin (version 3.1.16, lineages version 2021-11-25 available at https://pangolin.cog-uk.io/). For typing reads for the DNA viruses, the HPV-type specific reference genomes identified in the Virosaurus database using BLAST were mapped to the relevant reference in Virosaurus and the MCV-1 consensus sequence was extracted from the Virosaurus genome reference alignment and confirmed using BLASTn NCBI (see Supplementary Table 1 for reference genomes' Genbank IDs).

2.5.4. Data analysis and visualization

Statistics and figures were performed using Microsoft Excel 2016 and GraphPad Prism version 8.3.0 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com).

For graphic protocol presentation, please see Supplementary Figure 1.

3. Results

3.1. Evaluation of pretreatments on pooled samples

For the purposes of developing the methodology, a pool of clinical samples with known presence (or absence) of SARS-CoV-2 was used.

3.1.1. Effect of pretreatments

We initially compared three known sample pretreatment options: i) DNase I treatment, ii) filtration, iii) DNase I treatment followed by filtration and no pre-treatment (as control) on the 16 pooled samples described in 2.1.1. As shown in Table 1, adding two pretreatment steps to the metagenomic sequencing protocol increased the percentage of BLAST hits corresponding to SARS-CoV-2 from 31.8% (7 correct hits of 27 hits in all) without any treatment to 99.9 % % (10736 correct hits of 10747 hits in all) using just DNAse I followed by filtration. Furthermore, these pretreatments lead to a much faster accumulation of SARS-CoV-2 mapped reads (Fig. 1A and B). In contrast, without any pretreatments, too few SARS-CoV-2 BLAST hits to meet the quality requirement of >10 hits were obtained. With DNAse I or filtration as individual pretreatments alone, SARS-CoV-2 hits also met the quality requirements (Table 1) but BLAST hits for Orthohepevirus A (accession no.: MF444119) were also observed (data not shown).

3.1.2. Assembling reads to reference genomes to spot false-positive BLAST hits

The sequence reads obtained following each pretreatment of the samples were assembled on the genomes that met the BLAST hits requirements as described in Section 2.5.2. and visually inspected. As indicated above, matches to Orthohepevirus A sequences also met the quality requirements for samples pretreated with the DNase I alone or filtration alone. However, upon inspection, it was apparent that all the reads produced following these single pretreatments tiled at just two regions of 220 nt and 124 nt in the reference genome, respectively (genome coverage = 1.7% using either pretreatment) and were, therefore, disregarded as false-positive matches. This contrasted with the reads that mapped to the SARS-CoV-2 reference genome, which were evenly spread across the entire genome and were, therefore, regarded as true positive matches. Following the combination of DNase I and filtration pretreatments, no

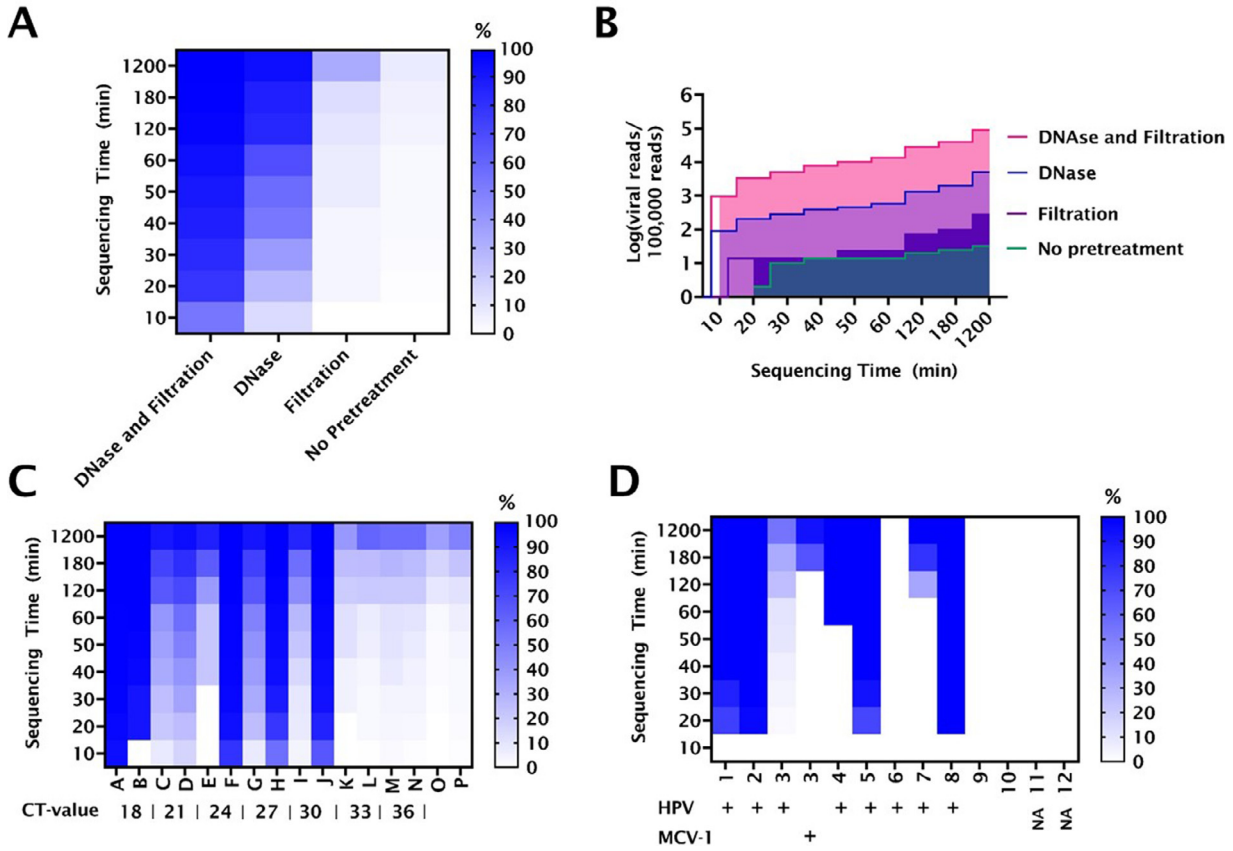

Fig. 1. Influence of virus genome content and sample pretreatment on the number of SARS-CoV-2 sequence reads **obtained. Pooled samples:** *A*) Coverage of the SARS-CoV-2 (NC045512) reference genome (%) for each evaluated sample pretreatment, *B*) Comparison of accumulated SARS-CoV-2 sequence reads over time for the indicated pretreatments. **Individual samples processed with pretreatment DNase I followed by filtration**: *C*) Coverage of the SARS-CoV-2 (NC045512) reference genome (%) for samples previously diagnosed using real-time RT-PCR. X axis: clinical samples A-N (divided into CT-range 18-36 in biological replicates shown as sub-X axis), clinical negative sample: O, water control: P. Left Y axis: Sequencing time; Right Y axis: Percentage coverage of the SARS-CoV-2 (Acc. No. NC045512) reference genome (%) achieved. *D*) Coverage of first detectable HPV reference (upper HPV-type for each sample in Table 3) from samples previously diagnosed using in-house microarray. X axis: clinical samples 1-8, 9-10 clinical negative swab and 11-12 being water controls (microarray results shown as sub-X axis (sample 3 co-infected with HPV and MCV-1). Left Y axis: Sequencing time; Right Y axis: Percentage coverage of the typed HPV or MCV-1 reference genomes (%) achieved.

virus other than SARS-CoV-2 was detected with BLAST hits requirements and only 10 minutes of sequencing time were needed before over 50% (53.3%) of the SARS-CoV-2 genome sequence was covered. For comparison, using DNase I pretreatment alone, it took 40 minutes of sequencing to obtain coverage of 52.5% of the reference genome.

3.2. Evaluation of pretreatment on individual samples

Using the same pretreatment conditions of DNase I treatment followed by filtration, 15 SARS-CoV-2 samples with various CT-values derived from TCDK were analyzed (Fig. 1C). All samples with a CT-value at or below 30 achieved a high reference sequence coverage (> 56.6-100%) after only 180 minutes and almost total coverage (85.8-100.0%) after 1200 minutes of sequence accumulation compared to 40.0-60.5% for samples with a CT-value between 33 and 36 (Fig. 1C, Table 2). However, it should be noted that even negative control samples (CT > 38) apparently produced some reads that mapped to the SARS-CoV-2 genome. The negative control reads mapped to a lower proportion (< 50%) of the reference genome and probably arose from a low level of index hopping or cross-contamination. However, this was orders of magnitude lower than that observed for the samples with a Pangolin lineage assignment. This is discussed further below (see 3.3 and 4.2). To evaluate the detection ability of the protocol towards DNA viruses, 8 samples with various HPV-types in each sample were analyzed. The mNGS protocol detected HPV in 7 out of 8 samples within 180 minutes of sequencing with almost full coverage (92.9-100.0%) of the genome after 1200 min sequencing (Table 3). Furthermore, it also correctly detected a co-infection of HPV-91 and MCV-1 in one of the samples (Fig. 1D, Table 3).

3.3. Typing of detected viruses

Consensus sequences were obtained for each of the SARS-CoV-2 samples and lineage typed as described in Section 2.5.3.. Pangolin identified various different Delta sub-lineages for the 10 samples with a CT-value below 33 (Table 2). For five of these samples, 100% of the reference sequence was covered. Samples with a lower level of viral RNA (CT of 33-36) could not be assigned to a lineage and reads only mapped to 50-60% of the genome which is insufficient for lineage assignment in Pangolin that requires at least 75% genome coverage for SARS-CoV-2-typing [14].

All the HPV positive samples except “Sample 7” contained multiple HPV-types (Table 3). It was not possible to find reads for all types that had been detected using the microarray. The HPV-types detected with the mNGS assay showed congruence with the results of the in-house microarray, with proper HPV-typing for at least one of the HPV-types present and no false-positive detection of HPV-types were observed (Table 3). For nine of these HPV-alignments, 100% of the reference sequence was covered. The mNGS assay detected the co-infection of two different viruses, HPV and MCV-1, in congruence with the in-house microarray.

Table 2
Quality trimmed read processing characteristics and lineage matching to SARS-CoV-2 Delta consensus sequence after 20 hours sequencing time of individual samples. Samples A-N: SARS-CoV-2 positive clinical samples. Sample O: SARS-CoV-2 negative clinical sample. Sample P: H2O negative control.

Sample name	SARS-CoV-2 CT-value	Mapped reads	Mapped viral reads pr. 100,000 reads generated	Fraction of reference covered (%)	Mean sequencing depth	Lineage using Pangolin
A	18	1,188,028	92,459	100.0	18216.8	AY.4
B	18	401,143	97,894	100.0	5018.7	AY.4
C	21	14,267	5,277	91.3	228.3	AY.4.5
D	21	18,084	13,096	95.1	283.4	AY.43
E	24	3,648	63,921	87.3	35.5	AY.42
F	24	445,092	84,910	100.0	6720.0	AY.29
G	27	8,543	2,043	92.1	88.8	AY.42
H	27	171,421	54,764	100.0	1974.0	AY.33
I	30	3,248	1,783	85.8	28.3	AY.42
J	30	367,216	68,210	100.0	5892.7	B.1.617.2
K	33	819	2,588	53.2	10.1	FAIL
L	33	550	233	56.1	7.4	FAIL
M	36	1,902	1,151	60.9	14.3	FAIL
N	36	1,130	1,073	57.6	15.3	FAIL
O	> 38	439	457	36.8	5.4	FAIL
P (H_2O)	No Ct	575	273	49.1	7.3	FAIL

Table 3
Quality trimmed read processing characteristics and type matching to reference sequence after 20 hours sequencing time for individual clinical samples containing DNA viruses. Samples 1-8: HPV positive clinical sample. Samples 9-10: HPV negative clinical sample. Sample 11-12: water negative control.

Sample name	Number of mapped reads	Mapped viral reads pr. 100,000 reads generated	Fraction of reference sequence covered (%)	Mean sequencing depth	Typing results for mNGS assay	In-house microarray detection results
1	6653	7708	100.0	206.7	HPV-91	HPV-33, HPV-43, HPV-91
	555	678	100.0	49.2	HPV-33	
2	6279	7034	100.0	425.2	HPV-91	HPV-6, HPV-43, HPV-51,
	5647	6350	100.0	394	HPV-6	HPV-91, HPV-108
3	100	101	92.9	9.8	HPV-91	HPV-59, HPV-87, HPV-91, MCV-1
	21423	18356	54.6	48.3	MCV-1	
4	1127	16322	100.0	99.6	HPV-91	HPV-43, HPV-66, HPV-91
5	8282	9379	100.0	222.7	HPV-27	HPV-2, HPV-27, HPV-57
6	ND	ND	ND	ND	ND	HPV-62, HPV-81
7	247	282	100	11.2	HPV-6	HPV-6
8	4552	3786	100	168.2	HPV-6	HPV-6, HPV-11, HPV-33,
	318	271	100	46.2	HPV-54	HPV-52, HPV-54
9	No virus detected	ND	ND	ND	No virus detected	No virus detected
10	No virus detected	ND	ND	ND	No virus detected	No virus detected
11	No virus detected	ND	ND	ND	No virus detected	NA
12	No virus detected	ND	ND	ND	No virus detected	NA

4. Discussion

Detection of emerging or re-emerging virus infections at the point of care is important to alert relevant health authorities to enable a fast response before substantial outbreaks develop. This can be challenging because a wide variety of viral agents can cause similar symptoms and novel viruses cannot be anticipated. mNGS allows non-biased investigation and detection of viral agents that could ease the need for highly specialized knowledge of symptoms and pathogens should this not be available at the point of care. The protocol described here has been shown to provide correct identification of the relevant pathogen (SARS-CoV-2, HPV, MCV-1).

4.1. Comparison to other metagenomic protocols for virus detection

One of the major challenges with unknown pathogen detection using mNGS is the low proportion of viral NA in a given clinical sample compared to the high levels of non-viral NA. This can consequently lead to the pathogen not being detected in an infected individual. A well-used strategy to overcome this problem is depletion of host sequences using nucleases while filtration treatments are also commonly used strategies alongside centrifugation [7]. The method described here is a further development from the in-house microarray methods described in Erlandsson et al. [4] and Rosenstierne et al. [16] but with field deployable pretreatments, NA extraction and mNGS using the Oxford Nanopore system. Others have also reported on virus detection and diagnosis using various pretreatments, and metagenomic assays.

Allander et al. [1] reported that their pretreatment, similar to our pretreatment, increased the successful identification of two bovine parvoviruses. They filtered (0.22 μM) serum using microcentrifugation followed by a DNase I (Stratagene) treatment for 2 hours at 37°C. They reported their Sequence-independent, Single-Primer Amplification method to take approximately 9 hours incubation time excluding a PCR reaction with 40 cycles which would not be suitable for a fast point of incidence test whereas the DNase I (Zymo research), as used in our protocol, followed by syringe filtration takes approximately 15 min, the hand extraction method taking 5-10 min per sample and the whole transcriptome amplification using REPLI-g requires under 3 hours and 40

minutes incubation. Others have reported that nuclease treatment alone generated a higher proportion of viral reads than a passage through a 0.45 μm filter followed by nuclease treatment [17]. However, that study looked at virus in spiked serum and tissue samples using a protocol developed for metagenomic virus detection in a laboratory. That study used TURBO DNase, which has been used in several viral metagenomic studies [6–8]. The use of DNase I, as described here, halves the treatment time and avoids the incubation at 37°C to use room temperature instead. Furthermore, the use of Illumina sequencing, which can generate more reads of a higher quality, is not feasible for point of care use. Greninger et al. [6] demonstrated a metagenomic approach using the more portable Nanopore technology for RNA virus detection and this method has been used by others during disease outbreaks, e.g. for Lassa fever [8]. However, their method required a thermal cycler. By switching to isothermal reactions, as in our protocol, a mNGS protocol can be made less complex and more suitable for less specialized mobile laboratory systems and even field-testing. Furthermore, the library preparation for Nanopore sequencing used by Greninger et al. [6] includes many "hands-on" steps (e.g., three magnetic purification steps during library preparation) whereas the Rapid Barcoding kit used here has the advantage of not requiring purification steps plus quick fragmentation and adapter ligation. Using the similar Rapid Sequencing kit, the concentration normalization step can also be avoided when sequencing only single samples. The protocol presented in this study has been developed with the aim of excluding large non-mobile advanced equipment such as centrifuges, extraction robots, thermal cyclers and sequencing robots (e.g., the Illumina apparatus) in an attempt to obtain simpler and mobile metagenomic pathogen detection systems. Russell et al. [18] also performed metagenomic Venezuelan Equine Encephalitis Virus (VEEV) detection using field-deployable methods such as Nanopore technology. Despite the detection, the VEEV reads were relatively few (< 1%) and similar results were found for this study. Inclusion of the proper pretreatment steps investigated and presented in this study drastically improved the number of reads and highlights how essential these steps are for the whole protocol of virus detection using metagenomics.

4.2. Limitations and evaluation

Our protocol has shown to correctly detect the virus in a given clinical sample. However, this protocol has a variety of limitation that we will list and discuss below.

Like any diagnostic assay, the efficiency of mNGS as a diagnostic tool depends strongly on the amount of virus nucleic acid in the sample [17]. Thus, the higher viral loads at the start of the symptomatic infection may facilitate detection by mNGS because the amount of viral NA can be larger. For this protocol, the CT-values of the SARS-CoV-2 samples after NA extraction were higher (i.e. lower levels of RNA) than for the same samples when immediately tested at the TCDK the same day (data not shown). However, by removing free-floating DNA with DNase I and filtering away larger components like host-cells, bacteria and other potentially interfering contaminants, the proportion of SARS-CoV-2 RNA derived reads increased enough to be detected readily in the samples. The drop in apparent sensitivity might be explained by this protocol selecting only for extracellular viral RNA whereas the NA extraction methods used at the TCDK and the SSI allow for both intra- and extracellular viral RNA detection. The inclusion of the REPLI-g WTA and WGA step for random amplification was thus an attempt to outweigh this loss of sensitivity.

Other possible pitfalls, when using mNGS for pathogen detection, are the risks of false-negative and false-positive results and it is therefore important to have negative controls included such as water [20]. There are no accepted golden guidelines for proper mNGS protocols for virus detection yet. Therefore, defining a cut-off for a negative result (absence of pathogen) needs to be empirically determined. Here, some reads mapping to the SARS-CoV-2 genome were observed from negative swab samples and water controls in the reads accumulated after 1200 minutes (Table 2), albeit these reads mapped to a much lower proportion of the genome than for "true" positive samples. This could have been caused at one or several steps during the procedure, e.g., as cross-contamination during the sample processing, from the environment or from extraneous sources of DNA from reagents used in the workflow. We therefore used a SARS-CoV-2-specific one-step qPCR assay to monitor contaminants after the NA extraction and after the random amplification using the phi29 DNA polymerase. The reverse transcriptase was omitted from the post random amplification step to look only for ds-cDNA from SARS-CoV-2, which could have been generated following the REPLI-g WTA protocol and subsequently sequenced. In duplicate qPCR assays, no positive signals were detected for the water control nor for negative swab sample following the one-step qPCR assay but showed a clearly amplified signal (lower CT-values) for SARS-CoV-2 in clinical positive samples after random amplification (data not shown). A possible explanation seems to be that reads from other samples in the library have been misassigned to the negative controls during demultiplexing.

The aim of this study was to optimize and investigate the ability of the mNGS assays to detect a viral pathogen in a sample under simple field conditions. After successfully detecting virus in congruence with commonly used diagnostic assays we went a step further to test the mNGS assay's typing ability. In this study, the mNGS assay reliably generated Pangolin typing-grade sequences when the CT-value was 30 or below, whereas this could not be obtained with higher CT-values (Table 2). For the samples containing DNA viruses, 7 out of 8 samples were found to have HPV. Because of the methodology of HPV diagnosis using a microarray, no viral quantitative measure were available for the HPV types, thus the reasoning for not detecting HPV-62 or HPV-81 is not clear, however it could tentatively be suggested that the viral load might have been above the detection limit of this protocol as seen for SARS-CoV-2 samples with CT > 30. Regarding HPV-typing, this method was not able to find all the multiple HPV-types present in the same sample as identified by the in-house microarray. All, except one, had multiple HPV-types as identified by the microarray but not all of these types were detected for each sample using this mNGS assay. Some of the HPV-types identified in each sample were closest relatives e.g. HPV-43 and HPV-91 in sample 1, 2 and 4 or HPV-2, HPV-27 and HPV-57 in sample 5. It was noticed that when closely related HPV-types were both present in a sample, only one would be detected using this mNGS assay. It is worth noting that the REPLI-g workflow can produce chimeric reads due to the ligation step that is needed because of the long DNA strand preference of the phi29 in the subsequent step [18]. In combination with the long-read Nanopore sequencing method it can become troublesome to type specific alignments. However, the combination still allows for viral detection [18], which is the primary goal of this method. Should more extensive typing be relevant, we recommend to take and aliquot of raw sample material into MPLB-buffer. The MPLB-buffer can act as a storage buffer until the sample can reach a laboratory with proper equipment for high-throughput virus diagnosis and typing.

Lastly, although we used Virosaurus, which is an actively curated and broad database for viral detection, and also checked manually that it contained reference sequences for the virus species detected in our clinical specimens by other methods (i.e. RT-PCR and Microarray), we cannot rule-out that other virus species might have been present in some of the samples. This is especially an essential point to consider, when applying this method to samples with an unknown viral composition albeit in low concentration. The use of pretreatments like DNase I or filtration alone did provide some spurious BLAST hits that fitted within the initial quality requirements, especially for Orthohepevirus A, but were removed upon closer inspection of mapped reads to the Orthohepevirus reference (accession no.: MF444119). This showed that database searches can be misleading and that manual inspection by trained personal of the results may be necessary. However, this was not relevant for this study since 99.9% of the BLAST hits were to SARS-CoV-2, but this might change with different sample types or for other viruses. One approach could be to circumvent the BLAST step and assemble the qual-

ity trimmed reads directly onto all known virus genomes; however, this database would need to be built with careful selection and continuous updating.

4.3. Future investigations

Our field extraction method using MPLB-buffer was originally developed for whole blood samples and urine [15]. Within this study, the method has also proven useful for oropharyngeal, vesicular and cervical swab materials. For mNGS on tissue samples, the hand extraction protocol as described by Rosenstierne et al. [15] may need optimization but this remains to be investigated on other types of sample material.

5. Conclusions

The world is regularly experiencing emerging viruses infecting humans, most recently the pandemic zoonotic RNA virus SARS-CoV-2. For rapid detection in a point of care setting, the response-time to the relevant stakeholders needs to be short and equipment should be transportable. By introducing a 15-minute pretreatment, with DNase I treatment followed by a 0.22 μM filtration step, the field-deployable protocol with Nanopore sequencing technology greatly increased the probability of detecting RNA and DNA viruses in a clinical sample. The increase in the proportion of generated viral reads obtained in the subsequent Nanopore sequencing allowed for a simple BLAST identification of SARS-CoV-2. The reads mapped with even coverage across the genome to a reference sequence, which enabled identification of the virus clade and often even the lineage. This principle may be useful for prompt identification of virus pathogens in clinical samples.

Funding statement

This work was supported by the TELE-Vir project, the European Union's Horizon 2020 Research and Innovation program [grant no: 773830] and the Danish Statens Serum Institut.

Declaration of Competing Interest

The authors report no declaration of interest.

Acknowledgements

Thank you to Sofie Holdflod Nielsen and Arieh Cohen from "Test Center Danmark" for supply of relevant samples.

This work was done as part of the TELE-Vir project, supported by funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement No 773830: One Health European Joint Programme.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcvp.2022.100120.

References

[1] T. Allander, S.U. Emerson, R.E. Engle, R.H. Purcell, J. Bukh, A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species, Proc. Natl. Acad. Sci. 98 (20) (2001) 11609–11614.

[2] N. Conceição-Neto, M. Zeller, H. Lefrère, P. De Bruyn, L. Beller, W. Deboutte, C.K. Yinda, R. Lavigne, P. Maes, M.V. Ranst, E. Heylen, J. Matthijnssens, Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis, Sci. Rep. 5 (1) (2015) 16532, doi:10.1038/srep16532.

[3] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K. Chu, T. Bleicker, S. Brünink, J. Schneider, M.L. Schmidt, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Euro Surveill 25 (3) (2020) 2000045.

[4] L. Erlandsson, M.W. Rosenstierne, K. McLoughlin, C. Jaing, A. Fomsgaard, The microbial detection array combined with random Phi29-amplification used as a diagnostic tool for virus detection in clinical samples, PLoS One 6 (8) (2011) e22631.

[5] A. Gleizes, F. Laubscher, N. Guex, C. Iseli, T. Junier, S. Cordey, J. Fellay, I. Xenarios, K. Laurent, P. Le Mercier, Virosaurus, 2021 https://viralzone.expasy.org/8676 Last Accessed July 23 2021.

[6] A.L. Greninger, S.N. Naccache, S. Federman, G. Yu, P. Mbala, V. Bres, D. Stryke, J. Bouquet, S. Somasekar, J.M. Linnen, R. Dodd, P. Mulembakani, B.S. Schneider, J.-J. Muyembe-Tamfum, S.L. Stramer, C.Y. Chiu, Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis, Genome Med. 7 (1) (2015) 99, doi:10.1186/s13073-015-0220-9.

[7] R.J. Hall, J. Wang, A.K. Todd, A.B. Bissielo, S. Yen, H. Strydom, N.E. Moore, X. Ren, Q.S. Huang, P.E. Carter, M. Peacey, Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery, J. Virol. Methods 195 (2014) 194–204, doi:10.1016/j.jviromet.2013.08.035.

[8] L.E. Kafetzopoulou, S.T. Pullan, P. Lemey, M.A. Suchard, D.U. Ehichioya, M. Pahlmann, A. Thielebein, J. Hinzmann, L. Oestereich, D.M. Wozniak, K. Efthymiadis, D. Schachten, F. Koenig, J. Matjeschk, S. Lorenzen, S. Lumley, Y. Ighodalo, D.I. Adomeh, T. Olokor, E. Omomoh, R. Omiunu, J. Agbukor, B. Ebo, J. Aiyepada, P. Ebhodaghe, B. Osiemi, S. Ehikhametalor, P. Akhilomen, M. Airende, R. Esumeh, E. Muoebonam, R. Giwa, A. Ekanem, G. Igenegbale, G. Odigie, G. Okonofua, R. Enigbe, J. Oyakhilome, E.O. Yerumoh, I. Odia, C. Aire, M. Okonofua, R. Atafo, E. Tobin, D. Asogun, N. Akpede, P.O. Okokhere, M.O. Rafiu, K.O. Iraoyah, C.O. Iruolagbe, P. Akhideno, C. Erameh, G. Akpede, E. Isibor, D. Naidoo, R. Hewson, J.A. Hiscox, R. Vipond, M.W. Carroll, C. Ihekweazu, P. Formenty, S. Okogbenin, E. Ogbaini-Emovon, S. Günther, S. Duraffour, Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak, Science 363 (6422) (2019) 74–77, doi:10.1126/science.aau9343.

[9] H. Li, Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics 34 (18) (2018) 3094–3100.

[10] C.B. Matranga, K.G. Andersen, S. Winnicki, M. Busby, A.D. Gladden, R. Tewhey, M. Stremlau, A. Berlin, S.K. Gire, E. England, L.M. Moses, T.S. Mikkelsen, I. Odia, P.E. Ehiane, O. Folarin, A. Goba, S.H. Kahn, D.S. Grant, A. Honko, L. Hensley, C. Happi, R.F. Garry, C.M. Malboeuf, B.W. Birren, A. Gnirke, J.Z. Levin, P.C. Sabeti, Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples, Genome Biol. 15 (11) (2014) 519, doi:10.1186/s13059-014-0519-7.

[11] A.M. Montmayeur, T.F.F. Ng, A. Schmidt, K. Zhao, L. Magaña, J. Iber, C.J. Castro, Q. Chen, E. Henderson, E. Ramos, High-throughput next-generation sequencing of polioviruses, J. Clin. Microbiol. 55 (2) (2017) 606–615.

[12] T.F.F. Ng, N.O. Kondov, X. Deng, A. Van Eenennaam, H.L. Neibergs, E. Delwart, A metagenomics and case-control study to identify viruses associated with bovine respiratory disease, J. Virol. 89 (10) (2015) 5340–5349.

[13] T.F.F. Ng, R. Marine, C. Wang, P. Simmonds, B. Kapusinszky, L. Bodhidatta, B.S. Oderinde, K.E. Wommack, E. Delwart, High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage, J. Virol. 86 (22) (2012) 12161–12175.

[14] Á. O'Toole, E. Scher, A. Underwood, B. Jackson, V. Hill, J.T. McCrone, R. Colquhoun, C. Ruis, K. Abu-Dahab, B. Taylor, Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool, Virus Evol. 7 (2) (2021) veab064.

[15] M.W. Rosenstierne, C.E. Jensen, A. Fomsgaard, Rapid, Safe, and Simple Manual Bedside Nucleic Acid Extraction for the Detection of Virus in Whole Blood Samples, J. Vis. Exp. 136 (2018) 58001, doi:10.3791/58001.

[16] M.W. Rosenstierne, K.S. McLoughlin, M.L. Olesen, A. Papa, S.N. Gardner, O. Engler, S. Plumet, A. Mirazimi, M. Weidmann, M. Niedrig, A. Fomsgaard, L. Erlandsson, The Microbial Detection Array for Detection of Emerging Viruses in Clinical Samples - A Useful Panmicrobial Diagnostic Tool, PLoS One 9 (6) (2014) e100813, doi:10.1371/journal.pone.0100813.

[17] T. Rosseel, O. Ozhelvaci, G. Freimanis, S. Van Borm, Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples, J. Virol. Methods 222 (2015) 72–80, doi:10.1016/j.jviromet.2015.05.010.

[18] J.A. Russell, B. Campos, J. Stone, E.M. Blosser, N. Burkett-Cadena, J.L. Jacobs, Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol, Sci. Rep. 8 (1) (2018) 1–12.

[19] V. Sauvage, M. Eloit, Viral metagenomics and blood safety, Transfus. Clin. Biol. 23 (1) (2016) 28–38, doi:10.1016/j.tracli.2015.12.002.

[20] P.J. Simner, S. Miller, K.C. Carroll, Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases, Clin. Infect. Dis. 66 (5) (2018) 778–788.

[21] WHOCoronavirus disease (COVID-19) outbreak, 2021 Link https://covid19.who.int/table Last Accessed December 29 2021.

[22] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 579 (7798) (2020) 265–269, doi:10.1038/s41586-020-2008-3.

[23] K. Ziegler, P. Steininger, R. Ziegler, J. Steinmann, K. Korn, A. Ensser, SARS-CoV-2 samples may escape detection because of a single point mutation in the N gene, Euro Surveill. 25 (39) (2020) 2001650, doi:10.2807/1560-7917.ES.2020.25.39.2001650.