Abstract
Emerging and re-emerging viruses are a global health concern. Genome sequencing as an approach for monitoring circulating viruses is currently hampered by complex and expensive methods. Untargeted, metagenomic nanopore sequencing can provide genomic information to identify pathogens, prepare for or even prevent outbreaks.
SMART (Switching Mechanism at the 5' end of RNA Template) is a popular method for RNA-Seq but most current methods rely on oligo-dT priming to target polyadenylated mRNA molecules. We have developed two random primed SMART-Seq approaches, 'SMART-9N', and a version compatible with barcoded PCR primers available from
Oxford Nanopore Technologies, ‘Rapid SMART-9N’, for the detection, characterization, and whole-genome sequencing of RNA viruses. The methods were developed using viral isolates, clinical samples, and compared to a gold-standard amplicon-based method. From a Zika virus isolate the SMART-9N approach recovered 10kb of the 10.8kb RNA genome in a single nanopore read. We also obtained full genome coverage at a high depth coverage using the Rapid SMART-9N, which takes only 10 minutes and costs up to 45% less than other methods. We found the limits of detection of these methods to be 6e00 focus forming units (FFU)/mL with 99.02% and 87.58% genome coverage for SMART-9N and Rapid SMART-9N respectively. Yellow fever virus plasma samples and SARS-CoV-2 nasopharyngeal samples previously confirmed by RT-qPCR with a broad range of Ct-values were selected for validation. Both methods produced greater genome coverage when compared to the multiplex PCR approach and we obtained the longest single read of this study (18.5 kb) with a SARS-CoV-2 clinical sample, 60% of the virus genome using the Rapid SMART-9N method. This work demonstrates that SMART-9N and Rapid SMART-9N are sensitive, low input, and long-read compatible alternatives for RNA virus detection and genome sequencing and Rapid SMART-9N improves the cost, time, and complexity of laboratory work.

**Keywords**

RNA virus, metagenomic, nanopore sequencing, genomic surveillance, diagnostic, ZIKV, YFV, SARS-CoV-2
Corresponding authors: Nicholas J. Loman (n.j.loman@bham.ac.uk), Joshua Quick (j.quick@bham.ac.uk)

Author roles: Claro IM: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Ramundo MS: Conceptualization, Data Curation, Formal Analysis, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Coletti TM: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization; da Silva CAM: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization; Valenca IN: Formal Analysis, Investigation, Visualization, Writing; Candido DS: Formal Analysis, Investigation, Visualization, Writing – Review & Editing; Sales FCS: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization; Manuli ER: Investigation, Methodology, Validation, Visualization; de Jesus JG: Investigation, Methodology, Validation, Visualization; de Paula A: Methodology, Validation, Visualization; Felix AC: Methodology, Validation, Visualization; Andrade PdS: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization; Pinho MC: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization; Souza WM: Writing – Original Draft Preparation, Writing – Review & Editing; Amorim MR: Methodology, Resources, Validation, Visualization; Proenca-Modena JL: Investigation, Methodology, Resources, Writing – Review & Editing; Kallas EG: Investigation, Methodology, Resources, Validation, Visualization; Levi JE: Investigation, Methodology, Resources, Validation, Visualization; Faria NR: Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Sabino EC: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Loman NJ: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Quick J: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by a Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MRC MR/S0195/1 and FAPESP 18/14389-0) (https://caddecentre.org); FAPESP/Newton Funds (2017/08012-9), and support from FAPESP for I.M.C (2018/17176-8); M.S.R. is supported by the Faculdade de Medicina da Universidade de São Paulo (FMUSP); T.M.C. (2019/07544-2), C.A.M.S. (2019/21301-5), F.C.S.S. (2018/25468-9), J.G.J. (2018/17176-8, 2019/12000-1, 18/14389-0) and E.C.S (2018/14389-0) are supported by FAPESP. P.S.A. is supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (88887.596940/2021-00). W.M.S is supported by FAPESP (2017/13981-0, 2019/24251-9), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (408338/2018-0), and the National Institutes of Health (AI120942). D.S.C is supported by the Clarendon Fund and by the Department of Zoology, University of Oxford. M.R.A is supported by CAPES (88887.356527/2019-00). N.R.F. is supported by a Wellcome Trust Royal Society Sir Henry Dale Fellowship (204311). N.J.L. and J.Q. are supported by the Wellcome ARTIC network (206298). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Claro IM et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Claro IM, Ramundo MS, Coletti TM et al. Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing [version 1; peer review: awaiting peer review] Wellcome Open Research 2021, 6:241 https://doi.org/10.12688/wellcomeopenres.17170.1

First published: 20 Sep 2021, 6:241 https://doi.org/10.12688/wellcomeopenres.17170.1
Introduction
RNA viruses are responsible for causing a broad range of human and veterinary diseases. In recent decades, RNA viruses have been a major cause of emerging and re-emerging infections, including Zika virus (ZIKV), Dengue virus (DENV), Human Immunodeficiency Virus (HIV), Ebola virus (EBOV), yellow fever virus (YFV), and recently, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The resulting epidemics and pandemics have caused high morbidity, mortality, and economic costs.

To date, our ability to manage these outbreaks is hampered by the challenge of making a definitive clinical diagnosis as many of these viruses are often clinically indistinguishable from those caused by co-circulating viruses and some bacterial pathogens. Diagnostic tests can be limited by low specificity, in the case of serological tests, or require a priori knowledge of the viruses to be targeted in the case of RT-PCR (reverse transcription-polymerase chain reaction). For these reasons, acute febrile illness often remains undiagnosed, leading to a failure of epidemiological surveillance. Rapid genomic surveillance systems are essential to identify emerging viruses, detect and monitor viral diversity, and be able to prepare for or even prevent new outbreaks.

New applications have been driven by technological advances in sequencing. The first examples of real-time genomic surveillance were conducted using targeted amplicon sequencing on the MinION (Oxford Nanopore Technologies). These studies exploited the portability of nanopore sequencing to achieve a faster turnaround time by sequencing the samples close to where they were collected. While successful for the EBOV epidemic in West Africa, and ZIKV, CHIKV, DENV, and YFV outbreaks in Brazil, this approach is best when the outbreak strain is known but is less suited to diverse viral groups or virus discovery.

Viral metagenomics, the process of sequencing the total viral nucleic acid content in a sample (typically cDNA or DNA), allows the genomic characterization of known and novel viruses in an untargeted manner. This technique is particularly useful for diagnostic, clinical laboratories, and public health surveillance. However, viral metagenomic sequencing directly from clinical samples can result in poor sensitivity especially in samples with a low abundance of viral genomic material relative to host-derived nucleic acid. Metagenomic sequencing using nanopore sequencing has already shown promise by Kafetzopoulou et al. (2018) who reported metagenomic sequencing of LASV, DENV, and CHIKV samples, and by Lewandowski et al. (2019) who sequenced the Influenza virus from respiratory samples.

In this study, we describe a high-sensitivity, low input, SMART approach for nanopore metagenomics of RNA viruses from isolated samples or from clinical samples. The SMART approach was originally described in 2001, using oligo-dT priming to target polyadenylated mRNA molecules. We adapted this method to use random priming for cDNA synthesis followed by PCR amplification, and Rapid SMART-9N barcoded PCR primers are used in the PCR amplification enabling the addition of barcodes in a single step. SMART-9N recovered a high proportion of viral reads from ZIKV isolate titrated down to 6e00 FFU/mL of material input including 94.4% of the genome in a single read. The methods were validated in YFV and SARS-CoV-2 directly from plasma and residual nasopharyngeal samples, respectively. The performance of this assay was compared to a gold-standard multiplex PCR method, demonstrating improvements in sequencing sensitivity, coverage, depth, cost, and complexity of both SMART-9N and Rapid SMART-9N, enabling enhanced pathogen detection for both diagnostic and surveillance application of RNA viruses.

Methods
Sample collection
ZIKV isolate strain BeH815744 (GenBank Accession No. KU365780) was propagated into Vero cells (CCL-81; ATTC, Manassas, USA) with minimum essential medium (MEM) for 2 hours at 37°C and 5% CO₂. The supernatant was removed, and MEM supplemented with 2% fetal bovine serum, 1% penicillin, and streptomycin were added. The cells were incubated for 4 days until 70% of cytopathic effect. Subsequently, the cell culture supernatant was collected and viral replication was confirmed through real-time quantitative reverse transcription-PCR (RT-qPCR) and quantified by an FFU assay in Vero cells. This sample was used to assess the performance of all three methods: multiplex PCR, SMART-9N, and Rapid SMART-9N, and the metagenomics approaches, SMART-9N, and Rapid SMART-9N, was tested in different serial ten-fold MEM dilutions up to 1–1,000,000 to assess the limit of detection (Extended data: Tables S1 and S2).

For methodological validation, human clinical samples included:
- 41 plasma samples previously positive for YFV by RT-qPCR collected between January 11 and May 10, 2018, with a cut-off value of Ct-values ≤ 37 (Extended data: Table S2) obtained from Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP), São Paulo, Brazil. The samples were amplified by multiplex PCR and only those amplified and visible on agarose gel were sequenced. From these, samples with a range of Ct-values were selected for SMART-9N and Rapid SMART-9N sequencing (Extended data: Table S1).
- Ten residual nasopharyngeal samples previously positive for SARS-CoV-2 by RT-qPCR, collected between November 17, 2020, and January 05, 2021, with Ct-values ≤ 34, obtained from the Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, Brazil (Extended data: Table S2). These samples were selected for multiplex PCR and Rapid SMART-9N sequencing (Extended data: Table S1).

Participants or their legal representatives provided signed informed consent. The ethical overview was provided by the institutional review boards at HC-USP and the Infectious
For the rapid Nucleic acid extraction and RT-qPCR testing
Samples were centrifuged for 5 minutes at 10,000 xg. For the ZIKV isolate, the viral RNA was isolated from 200 μl of the recovered material using the QuickAmp Viral RNA Mini Kit (Cat No. 52906, Qiagen, Germany) according to the manufacturer’s instructions and eluting in 50 μl of elution buffer. For the SARS-CoV-2 nasopharyngeal and YFV plasma samples, the viral RNA was extracted from 200 μl of material using the NucliSENS EasyMag system (BioMerieux, UK) automated DNA/RNA extraction platform, and eluted in 50 μl. Ct-values were previously determined for all samples by RT-qPCR for ZIKV[20], YFV[21], and SARS-CoV-2[22].

The extracted RNA was treated using TURBO DNase (Cat No. AM2239, Thermo Fisher Scientific, USA) at 37°C for 30 min to remove residual DNA before being cleaned up and concentrated to 10 μl using Zymo RNA clean & concentrator-5 (Cat No. R1016, Zymo Research, USA).

Multiplex tiling PCR
cDNA synthesis and Multiplex PCR. Samples were submitted to whole-genome sequencing using a gold-standard multiplex PCR amplicon sequencing approach[23]. Briefly, the cDNA was produced from RNA-positive samples using random hexamers (Cat No. N8080127, Thermo Fisher Scientific, USA) and ProtoScript II reverse transcriptase (Cat No. E6560, New England BioLabs, USA) according to the manufacturer’s instructions. The cDNA was then amplified with the multiplex PCR assay previously standardized for ZIKV, YFV[19], and with the ARTIC V3 multiplexed amplicon scheme for SARS-CoV-2[24] (https://artic.network/ncov-2019). PCR products were purified using a 1:1 ratio of AMPure XP beads (Cat No. A63881, Beckman Coulter, USA) and quantified using fluorometry with the Qubit dsDNA High Sensitivity Assay (Cat No. Q32854, Life Technologies, USA) both according to manufacturer’s instructions. The PCR products were purified using a 1:1 ratio of AMPure XP beads (Cat No. A63881, Beckman Coulter, USA) and quantified using fluorometry with the Qubit dsDNA High Sensitivity assay (Cat No. Q32854, Life Technologies, USA) on the Qubit 3.0 instrument (Life Technologies, USA) both according to manufacturer’s instructions.

Nanopore library preparation and sequencing, MinION libraries were prepared using 50 ng of each amplified cDNA, barcoded, and pooled in an equimolar fashion using the EXP-NBD104 (1–12) and EXP-NBD114 (13–24) Native Barcoding Kits (ONT, UK). Sequencing libraries were generated using the SQK-LSK109 Kit. 50 ng of the final libraries were loaded onto FLO-MIN106 flow cells on the MinION device (ONT, UK) and sequenced using MinKNOW 1.15.1 with the standard 48-hour run script.

SMART-9N
SMART cDNA synthesis and PCR. For the SMART-9N cDNA synthesis, 10 μl of the concentrated RNA, 1 μl NEB-RT tagged 9N (AAGCAGTGGTATCAAGCAGAGTACNNNNNNNNN, 2μM) and 1 μl 10 mM deoxyribonucleotide triphosphate (dNTPs) mix (Cat No. N0447L, New England Biolabs, USA) were mixed and incubated for 5 min at 65°C, then cooled on ice. 4 μl SuperScript IV first-strand buffer, 1 μl 0.1 M DTT, 1 μl RNase OUT, 1 μl NEB-SSP (RNA oligo) (GCTAACATTCGACGAGTGGTATCAGCGAGATCAGGTCACATGrGrG, 2 μM), and 1 μl SuperScript IV (Cat No. 18091200, Thermo Fisher Scientific, USA) were mixed with the 12 μl annealed RNA before incubation for 90 min at 42°C followed by 10 min at 70°C. These double-tagged cDNA products were amplified using 5 μl Q5 reaction buffer (Cat No. M0491, New England BioLabs, USA), 0.5 μl 10 μM dNTP 1 μl NEB-PCR (AAGCAGTGGTATCAGCGAGATGTCAGTGGTATCAACGCAGAGTACN, 20 μM), 15,75 μl Nuclease-free water (NFW), 0.25 μl Q5 DNA polymerase, and 2.5 μl of cDNA. PCR cycling conditions were: 98°C for 45 sec, followed by 30 cycles of 98°C for 15 sec, 62°C for 15 sec, and 65°C for 5 min and a final step of 65°C for 10 min. The products were purified using a 1:1 ratio of AMPure XP beads (Cat No. A63881, Beckman Coulter, UK) and quantified using fluorometry with the Qubit dsDNA High Sensitivity assay (Cat No. Q32854, Life Technologies, USA) on the Qubit 3.0 instrument (Life Technologies, USA) both according to manufacturer’s instructions.

Rapid SMART-9N
SMART cDNA synthesis and barcoded PCR. For the rapid SMART-9N cDNA synthesis, 10 μl of the concentrated RNA, 1 μl RLB RT tagged 9N (TCTTTTCGCTGGGCGGTCTCAACCNNNNNN, 2 μM) and 1 μl 10 mM dNTPs (Cat No. N0447L, New England BioLabs, USA) were mixed and incubated for 5 min at 65°C, then cooled on ice. 4 μl SuperScript IV First-strand Buffer, 1 μL 0.1 M DTT, 1 μl RNase OUT, 1 μl RLB TSO (RNA oligo) (GCTAACATTCGCTTTTCGCTGGGCGGTCTCAACCNNNNNN, 2 μM), and 1 μl SuperScript IV (Cat No. 18091200, Thermo Fisher Scientific, USA) was mixed with the 12 μl annealed RNA before incubation for 90 min at 42°C followed by 10 min at 70°C. These double-tagged cDNA products were amplified using 25 μl LongAmp Taq 2X master mix (Cat No. M0287, New England Biolabs, USA), 0.5 μl 10 μM dNTPs, 1 μl 0.1 M DTT, 1 μl RLB RT tagged 9N (TCTTTTCGCTGGGCGGTCTCAACCNNNNNN, 2 μM), and 1 μl SuperScript IV (Cat No. M0491, New England BioLabs, USA), 0.5 μl 10 μM dNTP 1 μl NEB-PCR (AAGCAGTGGTATCAGCGAGATGTCAGTGGTATCAACGCAGAGTACN, 20 μM), 15,75 μl Nuclease-free water (NFW), 0.25 μl Q5 DNA polymerase, and 2.5 μl of cDNA. PCR cycling conditions were: 95°C for 45 sec, followed by 25–30 cycles of 95°C for 15 sec, 65°C for 15 sec, and 65°C for 5 min and a final step of 65°C for 10 min. The products were purified using a 1:1 ratio of AMPure XP beads (Cat No. A63881, Beckman Coulter, USA) and quantified using fluorometry with the Qubit dsDNA High Sensitivity assay (Cat No. Q32854, Life Technologies, USA) on the Qubit 3.0 instrument (Life Technologies, USA) both according to manufacturer’s instructions.

Nanopore library preparation and sequencing. MinION libraries were prepared using a total input of 100 ng, barcoded, and pooled in an equimolar fashion using the EXP-NBD104 (1–12), and EXP-NBD114 (13–24) Native Barcoding Kits (ONT, UK). Sequencing libraries were generated using the SQK-LSK109 Kit (ONT, UK). 20 ng of the final libraries were loaded onto FLO-MIN106 flow cells on the MinION device (ONT, UK) and sequenced using MinKNOW 1.15.1 with the standard 48-hour run script.
for 10 min. For the PCR step, the RLB PCR (TTTTTTCGT-GCGCCGCTTCA, 20 μM) can be used as PCR control, changing the cycles conditions to 98°C for 45 sec, followed by 25–30 cycles of 98°C for 15 sec, 62°C for 15 sec, and 65°C for 5 min and a final step of 65°C for 10 min. The products were purified using a 1:1 ratio of AMPure XP beads (Cat No. A63881, Beckman Coulter, UK), quantified using fluorimetry with the Qubit dsDNA High Sensitivity Assay (Cat No. Q32854, Life Technologies, USA) on the Qubit 3.0 instrument (Life Technologies, USA) both according to manufacturer’s instructions.

**Nanopore library preparation and sequencing.** MinION libraries were prepared using 50 ng of each amplified cDNA, pooled in an equimolar fashion of 400 ng in 10 μl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. 1 μl RAP adapter was added to the library and incubated at room temperature for 5 min. 50 ng of the final libraries were loaded onto FLO-MIN106 flow cells on the MinION device (ONT, UK) and sequenced using MinKNOW 1.15.1 with the standard 48-hour run script.

**Bioinformatics workflow**

Raw FAST5 files were basecalled using Guppy software version 2.2.7 GPU basecaller (Oxford Nanopore Technologies), then demultiplexed and trimmed using Porechop v.0.3.2pre. The barcoded FASTQ files were aligned and mapped to the reference genome (GenBank accession no. JF912190 (YFV), KX893855.1 (ZIKV), and MN908947 (SARS-CoV-2)) using minimap2 version 2.28.0 and converted to a sorted BAM file using SAMtools version 2.6. NanoStat version 1.1.2 was used to compute the number of raw reads and minimum contig length to cover 50 percent of the genome (N50) of the aligned reads. Tablet 1.19.05.28 was used for genome visualization, and to compute the number of mapped reads, percentage of genome coverage, and coverage depth. samtools stats and samtools depth were used to calculate longest reads and genome coverage at 20x respectively. For the multiplex PCR analysis, length filtering, quality test, and primer trimming were performed for each barcode using artic guppyplex and variant calling and consensus sequences using artic minion Nanoplot version from ARTIC bioinformatics pipeline. For the SMART-9N and Rapid SMART-9N, called variants were detected with medaka_variants and the consensus sequence was built using medaka_consensus (ONT, UK).

For detection of other viral RNA in the clinical samples, taxonomic classification was conducted using Kraken version 2.0.8-beta, using the MiniKraken2_v1_8GB Kraken 2 Database.

**Results**

In this study, we designed two methodologies, SMART-9N and Rapid SMART-9N. The SMART-9N approach is based on the NEBNext Single-cell/low-input RNA (cat No. E6420, New England BioLabs, USA) modified to use random priming and native barcoding library preparation (ONT, UK). The NEB method uses single-primer PCR amplification which we found we could perform using barcoded primers from the Rapid PCR Barcoding Kit (ONT, UK) if we modified the sequence of the RT and SSP oligos. This approach allows for amplification of RNA in the picogram input range (data not shown) making it ideal for low-input applications. We compared the complexity, costs, and time required of laboratory work to a previously standardized multiplex tilling PCR approach. Compared to multiplex PCR, the total time of hands-on laboratory work dropped 15% and 61% for the SMART-9N and Rapid SMART-9N respectively, and reagent costs were reduced by 40% and 45% (Figure 1).

**Figure 1.** Comparison between the steps and cost of the workflows – tiling amplicon sequencing – Multiplex PCR, SMART-9N, and Rapid SMART-9N. Abbreviations: SSP: Strand Switching Primer. US$, American dollar.
Multiplex PCR sequencing of ZIKV isolate and YFV and SARS-CoV-2 clinical samples

Initial testing was performed on a serial dilution of ZIKV isolate, which was subjected to the gold-standard multiplex PCR approach followed by MinION sequencing\(^{22}\). This sample had a Ct-value of 15.1 and an RNA titer of 6e07 FFU/mL (Extended data: Table S2\(^22\)). The percentage of mapped reads was 55.99% with an average depth of 326.98x covering 98.74% of the viral genome covered with at least 1 read (Extended data: Table S3; Figure 2\(^\text{22}\)).

The assay was also performed on 41 human clinical samples positive for YFV RNA by RT-qPCR from the 2018 YFV epidemic in São Paulo, Brazil. The median Ct-value was 27.74, ranging from 4.6 to 37 corresponding to 1.00e00 to 1.50e10 genome copies per mL of plasma\(^{22}\). After PCR product quantification and the E-gel agarose gel run, 21 samples presented specific bands distinguishable by transillumination and were selected to continue nanopore library preparation and sequencing (Extended data: Table S3\(^22\)). The sequenced YFV samples (n=21) had a median Ct-values = 25.57, between 5 and 37 generated in one 24 barcoded library. The percentage of mapped reads ranged from 1.71% to 97.47%, with an average depth between 72.50x to 3370x, and the majority samples with genome coverage around 99.82% being the lowest 78.11% (Extended data: Table S3; Figure 3\(^\text{22}\)). Genome regions with a depth of <20x coverage were represented with N characters.

For SARS-CoV-2 the assay was performed in 10 residual nasopharyngeal samples positive for SARS-CoV-2 RNA by RT-qPCR in April 2020 in São Paulo, Brazil. The median Ct-value was 26.9, ranging from 21.8 to 33.3 corresponding to 2.40e05 to 1.30e02 genome copies per mL. The percentage of mapped reads ranged from 94.51% to 97.27%, with an average depth between 821.77x to 3370x, and the majority samples with genome coverage greater than 99.9% ranging from 95.11% to 99.99% with Ct-values of 33 and 18 respectively, and a minimum average depth of 3.2x, and a maximum of 3480x (Figure 3A). The same samples were selected and subjected to the Rapid SMART-9N method (Extended data: Table S7\(^22\)). The highest mapped read percentages observed were 98.26% and 38.18% for Ct-values 4.6 and 17.4, respectively. A total of 86% of the samples presented genome coverage greater than 99.9% with the lowest of 94.28% with a Ct of 33, and the average depth ranged from 21.44x to 2530x (Figure 3A). We compared the coverage depth with different Ct-values samples across the relevant genome for each method (multiplex PCR, SMART-9N, and Rapid SMART-9N) (Extended data: Figure S1\(^22\)). The average coverage depth revealed higher genome depth and better coverage pattern across the genome for the metagenomics methods when compared to the multiplex PCR method.

SMART-9N and Rapid SMART-9N of YFV clinical samples

After validating the methods on ZIKV isolate we next applied them to clinical samples. Starting with the SMART-9N, seven representative human clinical samples positive for YFV RNA, already sequenced with the multiplex PCR method, with Ct-values between 4.6 and 33 were selected (Extended data: Table S6\(^22\)). A total of 86% of the samples presented genome coverage greater than 99.9% ranging from 95.11% to 99.99% with Ct-values of 33 and 18 respectively, and a minimum average depth of 3.2x, and a maximum of 3480x (Figure 3A). The same samples were selected and subjected to the Rapid SMART-9N method (Extended data: Table S7\(^22\)). The highest mapped read percentages observed were 98.26% and 38.18% for Ct-values 4.6 and 17.4, respectively. A total of 86% of the samples presented genome coverage greater than 99.9% with the lowest of 94.28% with a Ct of 33, and the average depth ranged from 21.44x to 2530x (Figure 3A). We compared the coverage depth with different Ct-values samples across the relevant genome for each method (multiplex PCR, SMART-9N, and Rapid SMART-9N) (Extended data: Figure S1\(^22\)). The average coverage depth revealed higher genome depth and better coverage pattern across the genome for the metagenomics methods when compared to the multiplex PCR method.

All the seven sequenced samples with both methods were compared to the multiplex results. Despite the decrease in the proportion of mapped viral reads across the range of Ct-values (Figure 3B) with the SMART-9N and Rapid SMART-9N, we could obtain a comparable correlation (SMART-9N R=0.91, p=0.005; Rapid SMART-Metagenomics R=-0.86, p=0.012). The correlation showed a decreased proportion of viral reads as the Ct-values increased, with a considerable level of variation (0.3% to 98.6% with SMART-9N and 0.16% to 98.26% with Rapid SMART-9N method) between samples and methods (Extended data: Tables S6 and S7\(^22\)).

Comparable genome coverage 20-fold across the Ct-values between all methods is presented in Figure 3B. In the multiplex approach, the average of the genome coverage was 78.9% with a minimum of 35.01% for Ct 33 compared to 71.5% and 89.3% for SMART-9N and Rapid SMART-9N with a minimum of 0% and 50.5%, respectively (Extended data: Tables S6 and S7\(^22\)).

The same 1:1,000,000 dilution was used to test the Rapid SMART-9N approach. The lowest proportions of mapped reads observed were 0.06% and the highest 86.15%. The majority of samples returned a percentage of 99.87%, with 87.58% for the 6e00 FFU/mL dilution test. The median of the percentage of genome coverage at 20x was 90.73% and the N50 2.27kb (Figure 2).

The method was performed using 1 μl and 0.5 μl RLB barcodes from the SQK-RPB004 kit (ONT) with 6e07 of material input. The test resulted in 99.7% genomic coverage for both 1μl and 0.5 μl, and N50 of 1.84kb and 2.11kb respectively (Extended data: Table S5\(^22\)).
Figure 2. Comparison of multiplex PCR, SMART-9N, and Rapid SMART-9N approaches. A) Genome coverage of ZIKV genome for reference material as to coverage of reads mapped to the genome reference position comparing the multiplex PCR, SMART-9N, and Rapid SMART-9N approaches. B) Limit of detection of the SMART-9N and Rapid SMART-9N methods analyzing the proportions of reads mapping to the appropriate reference viral sequence across a range of sample input (FFU/mL) on the left plot and percentage of the reference genome sequenced to a minimum depth of 20-fold in the data generated across a range of sample input (FFU/mL) on the right plot.
Figure 3. Comparison of multiplex PCR, SMART-9N, and Rapid SMART-9N results for YFV clinical samples. A) Average genome coverage depth and 95% of reads mapped to the genome reference position. B) Proportion of reads mapping to the reference genome across a range of Ct-values (left) and percentage of the reference genome sequenced to a minimum depth of 20-fold across a range of Ct-values (right). C) N50 of each sample in bp. (n=7 samples).
Figure 4. Comparison of multiplex PCR, and Rapid SMART-9N results for SARS-CoV-2 clinical samples. A) Average genome coverage depth and 95% of reads mapped to the genome reference position. B) Proportion of reads mapping to the reference genome across a range of Ct-values (left) and percentage of the reference genome sequenced to a minimum depth of 20-fold across a range of Ct-values (right). C) N50 of each sample in bp. (n=10 samples).
For this subset of samples, we also compared the N50 results from the approaches for each sample (Figure 3C). Here we found the range was 525bp to 660bp for multiplex PCR, 659bp to 1.58kb for SMART-9N, 705bp to 2.16kb for Rapid SMART-9N. The median was 599.8bp, 1.6kb, and 1.2kb for the multiplex PCR, SMART-9N, and Rapid SMART-9N respectively (Extended data: Tables S6 and S7). For the YFV clinical samples, the longest reads observed were 10.08kb and 9.12kb for the SMART-9N, and Rapid SMART-9N, respectively 93.33% and 84.44% of the entire viral genome.

Rapid SMART-9N of SARS-CoV-2 clinical samples

SARS-CoV-2 clinical samples were subjected to the Rapid SMART-9N approach. Reads mapping to reference virus genome (isolate: Wuhan-Hu-1, GenBank Accession No. MN908947) were present in all ten seven samples up to Ct-value 34 (total reads ranged from 6480 to 93,570 reads). The sequenced samples were compared to the multiplex results and did not show a significant correlation (R=0.49, p=0.15) between the proportion of viral reads with increasing Ct-value (12.15 - 98.22%) (Extended data: Table S8). The genome coverage was 100% in all 7 samples and the lowest coverage depth of 977.5x (Figure 4A). When comparing each coverage depth against different Ct-values samples for the multiplex PCR, and Rapid SMART-9n methods (Extended data: Figure S2), we could observe a concordant coverage depth and coverage pattern across the genome for both methods.

Comparison of genome coverage 20-fold between multiplex PCR and Rapid SMART-9N across the viral titer range is shown in Figure 4B. The median revealed was 91.59%, minimum 84.49%, and the Rapid SMART-9N 99.79%, minimum of 99.57%. A comparison of the N50 in all the 10 samples was made resulting in a higher N50 of all samples to the Rapid SMART-9N approach, up to 2.56kb. The longest read was 18.48kb, the longest read obtained in this study, comprising approximately 62% of the SARS-CoV-2 genome (29,903 bp) (Figure 4C).

Detection of other RNA viruses in clinical samples and Kraken classification

To test the ability of our methods to detect other viruses in our samples, we assessed the taxonomic classification of reads using Kraken for all clinical samples. This allowed for the identification of a dsDNA virus genus Pa6virus, family Siphoviridae present in one YFV sample. After identification, the reads were mapped to the reference sequence (NC_018838.1) obtaining 197 reads of the virus, 84.78% of genome coverage with a maximum coverage depth of 32x, and identity of 81.4%. The consensus sequence was generated and bases with a depth of <10-fold were represented with N characters (github.com/CADDE-CENTRE/Rapid-RNA-SMART-Metagenomics). The proportion of unclassified, eukaryota, bacteria, archaea, and viruses reads, for each sample can be found in the Extended data, Table S9.

Discussion

A rapid, sensitive sequencing method for viral metagenomics is key to be able to identify the cause of unknown infections. Although PCR-based testing and amplicon-based sequencing methodologies are available and very sensitive, they are not suitable for the initial detection of emerging or re-emerging viruses due to the need for gene-specific primers/probes for diagnostic assays or primer panels. The aetiology of suspected infections in acute illness often remains undiagnosed. An untargeted sequencing method remains the best strategy for the identification of unknown viral infections, and the genome sequences provide information about the evolutionary history, strain identification, and biology of new pathogens. This is evidenced by the recent rapid and impactful metagenomic analysis of SARS-CoV-2 early in the pandemic.

In this study, we developed two viral metagenomic approaches, SMART-9N and Rapid SMART-9N as non-targeted metagenomics methods for detection and characterization of viral RNA. The two techniques demonstrated excellent specificity (100%) when tested in isolated and clinical samples that had been compared to a gold-standard multiplex PCR method and allowed the identification of an unknown co-infection in a YFV clinical sample. For ZIKV isolated-culture, it was possible to obtain 99.02% of genome coverage with an input of 6e00 FFU/mL, an amount comparable to other single-cell methods available for the SMART-9N approach. For the Rapid SMART-9N, 87.58% of the ZIKV genome was recovered for the same dilution of 1:1,000,000, equivalent to 6e00 FFU/mL. The sensitivity and high yield of viral sequences from clinical YFV and SARS-CoV-2 samples make it potentially feasible to directly perform metagenomic MinION whole-genome sequencing, even for higher Ct-values. Representative clinical samples with Ct-values between 4.6 and 33 for YFV, and between 21.8 and 33.3 for SARS-CoV-2 were selected to test the genome recovery for the viruses tested. Notably, the SMART-9N and Rapid SMART-9N methods were effective in directly genome sequencing clinical samples for both viruses tested since viral reads were detected in all samples, until in samples with 1.00e00 genome copy per mL.

Evaluating the read length during the validation, we observed that our approach generated very long reads when compared to other metagenomics approaches. In this study, we generated the whole ZIKV and YFV genome and approximately 60% of the SARS-CoV-2 genome in one single read. The N50 of the methods was up to 2.91kb with the isolated samples and up to 2.56kb with SARS-CoV-2 clinical samples and the Rapid SMART-9N approach. When looking at the average coverage depth and the CI of the metagenomics methods, we could observe a consistent amplification across the entire genomes due to this high N50 generated for the methods. Increased N50 provides higher confidence in individual read taxonomic assignment, improves mapping confidence, de novo assembly, and the ability to detect viral
recombinations. To our knowledge 18.5kb is the longest viral cDNA published to date produced by the Rapid SMART-9N method, this was likely due to the fact that LongAmp polymerase is used for barcoded primers as per ONT recommendations whereas Q5 polymerase was used for SMART-9N.

We also compared the complexity, costs, and time required of laboratory work to the multiplex tiling PCR approach. Using this approach, our Rapid SMART-9N reduced the complexity, time, and cost from sample to sequence. The possibility of barcode addition during PCR decreased the time of library preparation from 6 hours to 10 minutes, reducing the probability of laboratory error and the cost due to no longer needing enzymes for end-preparation and ligation which also rely on a cold chain making them inconvenient to use in the field. The total time of laboratory work dropped 15% and 61% for the SMART-9N and Rapid SMART-9N respectively when compared to the multiplex PCR. The costs when using Rapid SMART-9N dropped 45% and 40% compared to SMART-9N and multiplex PCR respectively. Using half of the volume for the rapid barcode primers doubles the number of samples that can be processed with the kit. This protocol has the potential to be further optimized and used in a lyophilized formulation with the elimination of any cold chain. These results demonstrate that the Rapid SMART-9N is an important approach in both the laboratory or field settings.

Limitations of the method
The overwhelming majority of reads are derived from the human host, mainly in clinical samples with high Ct-values (with a low relative abundance of viral genomic material) or in samples with degraded genetic material, due to incorrect transport or long-term storage. This is a limiting factor of the analytical sensitivity of the approach for infectious agent detection that could result in insufficient sequencing depth of the organism of interest. This disadvantage is shown in our study when comparing the viral mapped reads of the ZIKV isolates to the YFV and SARS-CoV-2 clinical samples. The lower proportion of viral reads as the increase of the Ct-values is likely because of the total level of non-viral host/background nucleic acid present due to variability between patients or in sample handling during collection.

Conclusion
Here we demonstrate a sensitivity workflow across viral isolate and clinical samples which takes advantage of long-read nanopore sequencing technology by generating long (up to 18.5 kb) cDNA amplification products for viral metagenomics. Therefore, our metagenomic sequencing approaches offer an opportunity for sensitive identification and characterization of RNA viruses directly from isolates or clinical samples with a range of viral loads. Also, the Rapid SMART-9N demonstrated a simple, low-cost, and faster method, promising for routine use in the research laboratory as well as in the field.

Data availability
Underlying data

Extended data

This project contains the following underlying data:
- SARS_CoV_2_CONSENSUS_SEQUENCES (SARS-CoV-2 consensus sequences (n=10) generated by multiplex PCR and Rapid SMART-9N methods).
- YFV_CONSENSUS_SEQUENCES (YFV consensus sequences (n=7) generated by multiplex PCR, SMART-9N, and Rapid SMART-9N methods).
- ZIKV_CONSENSUS_SEQUENCES (ZIKV reference consensus sequences (n=1) generated by multiplex PCR, SMART-9N, and Rapid SMART-9N methods).
- ZIKV_Multiplex_PCR_RAW_FILES (Raw data (fastq) of ZIKV, SARS-CoV-2, and YFV generated in this study).

This project contains the following extended data within the file ‘Supplementary_material_SMART_9N.pdf’:
- Table S1 - Description of samples collected and protocol realized to each sample.
- Table S2 - Description of samples positive for Zika virus (ZIKV) reference sample strain BeH815744 (n=1), yellow fever virus (YFV) (n=1), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (n=10) by real-time quantitative reverse transcription PCR with the corresponding sample types, Ct-values, estimated focus forming units (FFU) per milliliter or estimated genome copies per mL, and the virus reference size (nts).
- Table S3 - Summary of virus nanopore sequencing data using the tiling multiplex PCR approach of Zika virus reference strain BeH815744 (ZIKV) (n=1), yellow fever virus (YFV) (n=1), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (n=10) samples with the corresponding Ct-values.
- Table S4 - Sequencing summary and alignment statistics results for Zika virus (ZIKV) reference sample strain BeH815744 using the SMART-9N method during development (n=1 sample) according to the material input (FFU/mL).
- Table S5 - Sequencing summary and alignment statistics results for Zika virus reference sample strain BeH815744 using the Rapid SMART-9N method during development (n=1 sample) according to the material input (FFU/mL).
- Table S6 - Sequencing summary and alignment statistics results for yellow fever virus (YFV) plasma samples (n=7) using the SMART-9N protocol during method validation according to the Ct-values.
- Table S7 - Sequencing summary and alignment statistics results for yellow fever virus (YFV) plasma samples (n=7) using the Rapid SMART-9N protocol during method validation according to the Ct-values.
- Table S8 - Sequencing summary and alignment statistics results for Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) clinical samples (n=10) using the Rapid SMART-9N protocol during method validation according to the Ct-values.
- Table S9 - Proportion in the percentage of unclassified, Eukaryota, bacteria, archaea, and viruses reads, for each sample according to the Kraken classification distribution and metagenomics methodologies.
- Figure S1 - Comparison of genome coverage depth across the yellow fever virus (YFV) genome for different methods (i.e., multiplex PCR, SMART-9N, and Rapid SMART-9N) in all clinical samples tested with different Ct-values. YFV (n=7)
- Figure S2 - Comparison of genome coverage depth across the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus genome for different methods (i.e., multiplex PCR, and Rapid SMART-9N) in all clinical samples tested with different Ct-values. SARS-CoV-2 (n=10).

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Acknowledgments
We thank the CADDE (Brazil-UK Centre for Arbovirus Discovery, Diagnosis, Genomics, and Epidemiology), ARTIC Network team, and the FAPESP and MRC foundation for the funding.

References

20. Lanciotti RS, Kosoy OL, Laven jj, et al.: Genetic and serologic properties of