














METHOD ARTICLE

REVISED Rapid viral metagenomics using SMART-9N

amplification and nanopore sequencing [version 2; peer review: 2 approved]

Ingra M. Claro ¹⁻⁴, Mariana S. Ramundo ³, Thais M. Coletti ³,
Camila A. M. da Silva ³, Ian N. Valenca ³, Darlan S. Candido^{2,3,5},
Flavia C. S. Sales ^{1,3}, Erika R. Manuli³, Jaqueline G. de Jesus ^{2,3},
Anderson de Paula³, Alvina Clara Felix³, Pamela dos Santos Andrade^{3,6},
Mariana C. Pinho³, William M. Souza ⁷, Mariene R. Amorim⁸,
José Luiz Proenca-Modena^{8,9}, Esper G. Kallas¹, José Eduardo Levi^{3,10},
Nuno Rodrigues Faria ^{2,3,5}, Ester C. Sabino ³, Nicholas J. Loman ⁴,
Joshua Quick⁴

¹Faculdade de Medicina da Universidade de São Paulo, Sao Paulo, 05403-000, Brazil

²MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, SW7 2AZ, UK

³Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, Sao Paulo, 05403-000, Brazil

⁴School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

⁵Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

⁶Faculdade de Saúde Pública da Universidade de São Paulo, Sao Paulo, 01246-904, Brazil

⁷World Reference Center for Emerging Viruses and Arboviruses and Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX, 77555, USA

⁸Laboratory of Emerging Viruses, Department of Genetics, Microbiology, and Immunology, Institute of Biology, University of Campinas, Campinas, 13083-862, Brazil

⁹Experimental Medicine Research Cluster, University of Campinas, Campinas, 13083-862, Brazil

¹⁰DASA, Sao Paulo, 06455-010, Brazil

v2 First published: 20 Sep 2021, 6:241
<https://doi.org/10.12688/wellcomeopenres.17170.1>








Latest published: 24 Apr 2023, 6:241
<https://doi.org/10.12688/wellcomeopenres.17170.2>

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 approach 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, a sequencing agnostic approach 'SMART-9N' and a version compatible rapid

Open Peer Review

Approval Status  

	1	2
version 2		
(revision)		
24 Apr 2023	view	view
		
version 1		
20 Sep 2021	view	view
1. Edward Cunningham-Oakes  , University		


adapters available from Oxford Nanopore Technologies 'Rapid SMART-9N'. 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 6 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

of Liverpool, Liverpool, UK

2. **Mirela D'arc Ferreira da Costa** , Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Any reports and responses or comments on the article can be found at the end of the article.

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; **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:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Sabino EC:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, 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 (MR/S0195/1 and FAPESP 18/14389-0) (<https://caddecentre.org>); FAPESP/Newton Funds (FAPESP 2017/08012-9), Bill & Melinda Gates Foundation (INV-034540 and INV-034652), and I.M.C (FAPESP 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: © 2023 Claro IM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted 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 2; peer review: 2 approved]** Wellcome Open Research 2023, 6:241 <https://doi.org/10.12688/wellcomeopenres.17170.2>

First published: 20 Sep 2021, 6:241 <https://doi.org/10.12688/wellcomeopenres.17170.1>

REVISED Amendments from Version 1

We thank the reviewers for the consideration of revising the manuscript.

In this version, the article has been updated to incorporate all the suggestions from the reviewers. In particular, we have tried to improve the image quality of [Figure 1](#), and we have changed the [Figure 2](#) legend and x-axis in panel B to FFU/ml. Additionally, the text has been made clearer, and we add to the methods and discuss in more detail the use of an untargeted analysis for potential viral classification as requested by both reviewers.

Any further responses from the reviewers can be found at the end of the article

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 costs¹.

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^{2,3}. 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 outbreaks⁴.

New applications have been driven by technological advances in sequencing. The first examples of real-time genomic surveillance^{5,6} 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, chikungunya virus (CHIKV), DENV, and YFV outbreaks in Brazil⁷⁻⁹, 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 surveillance¹⁰⁻¹². However, viral metagenomic sequencing

directly from clinical samples suffers from poor sensitivity, especially in samples with a low abundance of viral genomic material relative to host-derived nucleic acid¹³⁻¹⁵. Nanopore metagenomic sequencing has already shown promise by Kafetzopoulou *et al.* (2018) who reported metagenomic sequencing of Lassa virus (LASV), DENV, and CHIKV samples¹⁶, and by Lewandowski *et al.* (2019) who sequenced the Influenza virus from respiratory samples¹⁷. In both of these studies the approach used was SISPA¹⁸ which generates double-tagged cDNA during second-strand synthesis rather than by the SMART mechanism.

In this study, we describe a high-sensitivity, low input, SMART (Switching Mechanism at the 5' end of RNA Template) approach for nanopore metagenomics of RNA viruses from isolated samples or from clinical samples. The SMART approach was originally described in 2001¹⁹, using oligo-dT priming to target polyadenylated mRNA molecules. We adapted this method to random priming for cDNA synthesis followed by PCR amplification (SMART-9N), 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 a ZIKV isolate titrated down to 6 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 method²⁰, demonstrating improvements in sequencing sensitivity, coverage, depth, cost, and complexity of both SMART-9N and Rapid SMART-9N, enabling enhanced pathogen detection for both diagnosis and surveillance of RNA viruses.

Methods

Sample collection

ZIKV isolate strain BeH815744 (GenBank Accession No. KU365780) was propagated into Vero cells (CCL-81; ATCC, 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 1% streptomycin, to prevent bacterial growth. 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)²¹ and quantified by focus-forming units (FFU) assay in Vero cells²². This sample was used to assess the performance of all three methods: multiplex PCR, SMART-9N, and Rapid SMART-9N. The metagenomic approaches, SMART-9N, and Rapid SMART-9N, were 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-qPCR²³ collected between January 11 and May 10, 2018, with a ct-value cut-off of ≤ 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 Ct-values between 4.6 and 33 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-qPCR²⁴, 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 Diseases Institute “Emilio Ribas” for the YFV study as part of the Efficacy of Sofosbuvir as a treatment for Yellow Fever study, protocol number CAAE 82673018.6.1001.0068. For the SARS-CoV-2 study, ethical approval was by the national ethical review board (Comissão Nacional de Ética em Pesquisa), protocol number CAAE 30127020.0.0000.0068.

Nucleic acid extraction and RT-qPCR testing

For the ZIKV isolate, the viral RNA was isolated from 200 μ l of the culture supernatant material using the QIAamp Viral RNA Mini Kit (Cat No. 52906, Qiagen, Germany) according to the manufacturer’s instructions and eluting in 50 μ l of elution buffer. For the SARS-CoV-2 nasopharyngeal and YFV plasma samples, 500 μ l of the clinical samples were centrifuged for 5 minutes at 10,000 g. The viral RNA was extracted from 200 μ l of supernatant material using the NucliSENS Easy-Mag system (BioMerieux, UK) automated DNA/RNA extraction platform, and eluted in 50 μ l. Ct-values were previously determined for all samples by RT-qPCR for ZIKV²¹, YFV²³, and SARS-CoV-2^{23,24}.

44 μ l of 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 μ l 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^{9,20}. 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, YFV²⁰,

and with the ARTIC V3 multiplexed amplicon scheme for SARS-CoV-2²⁵ (<https://artic.network/ncov-2019>). PCR products were purified using a 1:1 ratio of AMPure XP beads (Cat No. A63881, Beckman Coulter, UK) and 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. A gel was prepared with the PCR products using the E-gel Agarose 2% (Cat No. G402002, Thermo Fisher Scientific, USA) on E-gel equipment (Thermo Fisher Scientific, USA) and the run was performed until the bands were distinguishable by transillumination. The samples that presented bands with the expected amplicon size (approximately 500 base-pairs (bp)) were selected for MinION sequencing.

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.

SMART-9N

SMART cDNA synthesis and PCR

For the SMART-9N cDNA synthesis, 10 μ l of the concentrated RNA, 1 μ l NEB bRT 9N (AAGCAGTGGTATCAACGCA-GAGTACNNNNNNNNN, 2 μ M) and 1 μ l 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 μ l SuperScript IV first-strand buffer, 1 μ l 0.1 M DTT, 1 μ l RNase OUT, 1 μ l NEB SSP (RNA oligo) (GCTAATCATTGCAAGCAGTGGTATCAACG-CAGAGTACATrGrGrG, 2 μ M), and 1 μ l SuperScript IV (Cat No. 18091200, Thermo Fisher Scientific, USA) were mixed with the 12 μ l 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 μ l Q5 reaction buffer (Cat No. M0491, New England BioLabs, USA), 0.5 μ l 10 μ M dNTP, 1 μ l NEB PCR (AAGCAGTGGTATCAACGAGAGT, 20 μ M), 15.75 μ l Nuclease-free water (NFW), 0.25 μ l Q5 DNA polymerase, and 2.5 μ l 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 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, 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.

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 9N (TTTTTCGTGCGCCGCTCAACNNNNNNNN, 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) (GCTAATCATTGCTTTTTCGTGCGCCGCTTCAACATrGrGrG, 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), 19.5 µl of NFW, 0.5 µl RLB (1-12) from SQK-RPB004 kit (ONT, UK) and 5 µl of cDNA. PCR cycling conditions were: 95°C for 45 sec, followed by 25–30 cycles of 95°C for 15 sec, 56°C for 15 sec, and 65°C for 5 min and a final step of 65°C for 10 min. For the PCR step, the RLB PCR (TTTTTCGTGCGCCGCTTCA, 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 200 Femtomolar (fM) in 10 µL of 10 millimolar (mM) Tris-HCl pH 8.0 with 50 millimolar (mM) NaCl. 1 µl RAP adapter was added to the library and incubated at room temperature for 5 min. 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 2.28.0²⁶ and converted to a sorted BAM file using SAMtools²⁷. NanoStat version 1.1.2²⁸ 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 1.19.05.28²⁹ was used for genome visualization, and to

compute the number of mapped reads, percentage of genome coverage, and coverage depth. Samtools stats and samtools depth²⁷ 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 Nanopolish 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](#), which comprises eukaryotic, bacterial, viral, and archaeal Refseq complete genomes. After classification, those classified as “Viruses” in the output reports, were analysed for each barcode individually. The manual downstream analyses consisted on mapping each FASTQ file to the respective potential FASTA of the virus of interest downloaded from NCBI. Tablet²⁹ was used to verify the genomes mapping pattern, and to exclude the possibility of genome chimera or false positive interpretation. A dsDNA virus genus Pa6virus, family Siphoviridae was identified in one YFV sample, and the pipeline described above was used to generate consensus sequences, using the reference sequence (NC_018838.1).

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 (cat no. SQK-NBD104, 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 (SQK-RBP004, 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 approach²⁰. Compared to multiplex PCR, the total time of hands-on laboratory work dropped 15% and 57% for the SMART-9N and Rapid SMART-9N respectively, and reagent costs were reduced by 40% and 45% ([Figure 1](#)).

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 sequencing²⁰. This sample had a Ct-value of 15.1 and an RNA titer of 6e7 FFU/mL (*Extended data*: Table S2). 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](#)).

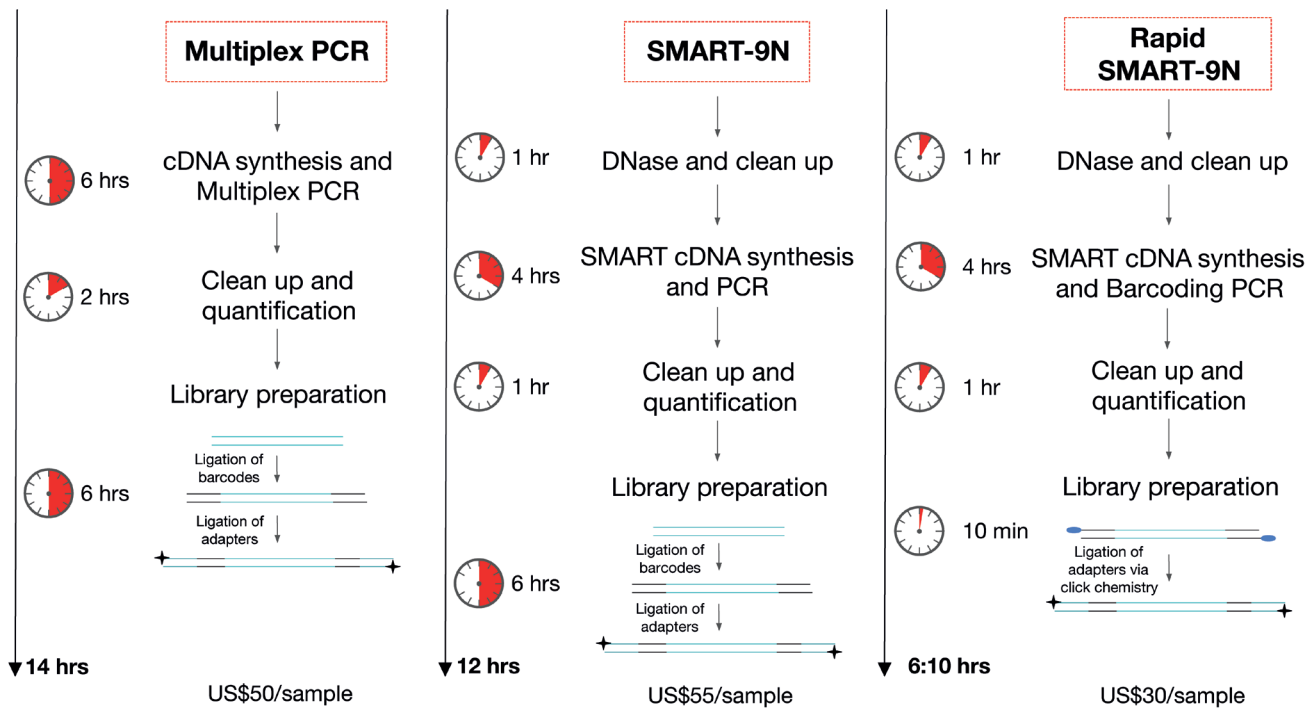


Figure 1. Comparison between the steps and cost of the workflows – tiling amplicon sequencing – Multiplex PCR, SMART-9N, and Rapid SMART-9N. Abbreviations: SSP: Strand Switching Primer. US\$, American dollar.

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 to 1.5e10 genome copies per mL of plasma³⁰. 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). The sequenced YFV samples (n=21) had a median Ct-values = 25.57, between 5 and 37 generated in one barcoded ONT library. The percentage of mapped reads ranged from 1.71% to 97.47%, with an average depth between 72.5x to 3370x, and the majority samples with genome coverage around 99.82% being the lowest 78.11% (*Extended data*: Table S3; [Figure 3](#)). 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 1.3e2 to 2.4e5 genome copies per mL. The percentage of mapped reads ranged from 94.51% to 97.27%, with an average depth of 821.77x to 1570x, and genome coverage median

of 98.8%, ranging from 95.90% to 99.92% (*Extended data*: Table S3; [Figure 4](#)).

SMART-9N and Rapid SMART-9N of ZIKV isolated-culture samples and limit of detection

For ZIKV, the titrated isolate RNA was diluted with serial ten-fold dilutions, up to 1:1,000,000 corresponding to 6e7 to 6 FFU/mL, and subjected to SMART-9N (*Extended data*: Table S4). The test resulted in a median of 99.7% genomic coverage for the tested dilutions with the lowest 99.02% for the 1:1,000,000 dilution. The percentage of genome coverage at 20x was 90.7% with 6 FFU/mL up to 99.73% with 6e7 FFU/mL of material input. The coverage depth was up to 10010x, and with 6 FFU/mL of material was 154.25x, compatible with single-cell assays. The average of mapped reads ranged from 56.29% to 0.52%. The median N50 was 1.7kb and when the reads were individually analysed, the test obtained complete ZIKV genome coverage in a single read (approximately 11kb longest read) ([Figure 2](#)).

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

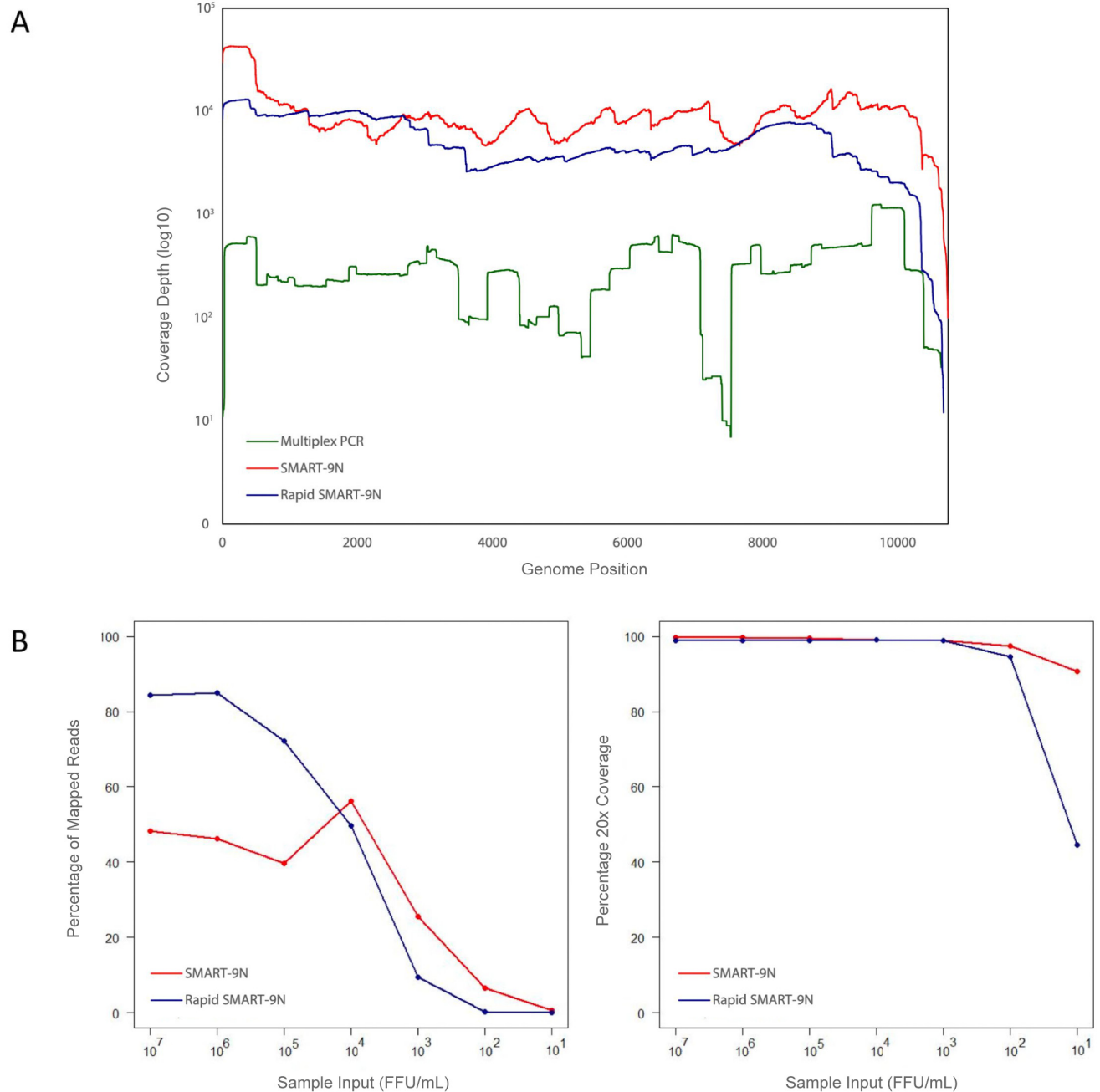


Figure 2. Comparison of multiplex PCR, SMART-9N, and Rapid SMART-9N approaches. **A)** Genome coverage of ZIKV genome for reference material as to coverage of reads mapped to the genome reference position comparing the multiplex PCR, SMART-9N, and Rapid SMART-9N approaches. **B)** Limit of detection 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.

the 6 FFU/mL dilution test. The median of the percentage of genome coverage at 20x was 90.73% and the N50 was 2.27kb (Figure 2).

The method was performed using 1 μ l and 0.5 μ l RLB barcodes from the SQK-RPB004 kit (ONT) with 6×10^7 FFU/mL of material input. The test resulted in 99.7% genomic

coverage for both 1 μ l and 0.5 μ l, and N50 of 1.84kb and 2.11kb respectively (Extended data: Table S5).

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,

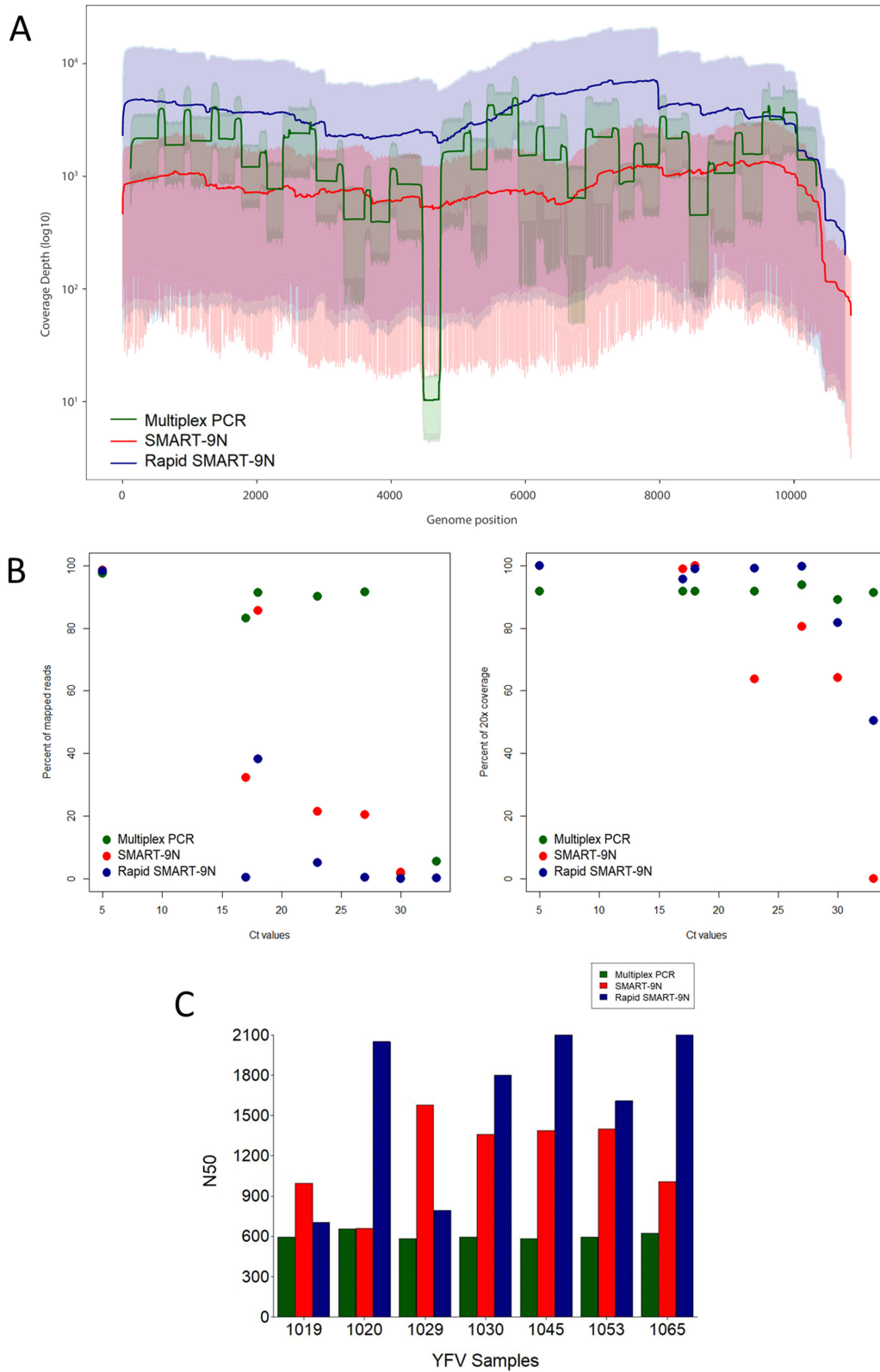


Figure 3. Comparison of multiplex PCR, SMART-9N, and Rapid SMART-9N results for YFV clinical samples. A) Average genome coverage depth and 95% of reads mapped to the genome reference position. **B)** Proportion of reads mapping to the reference genome across a range of Ct-values (left) and percentage 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).

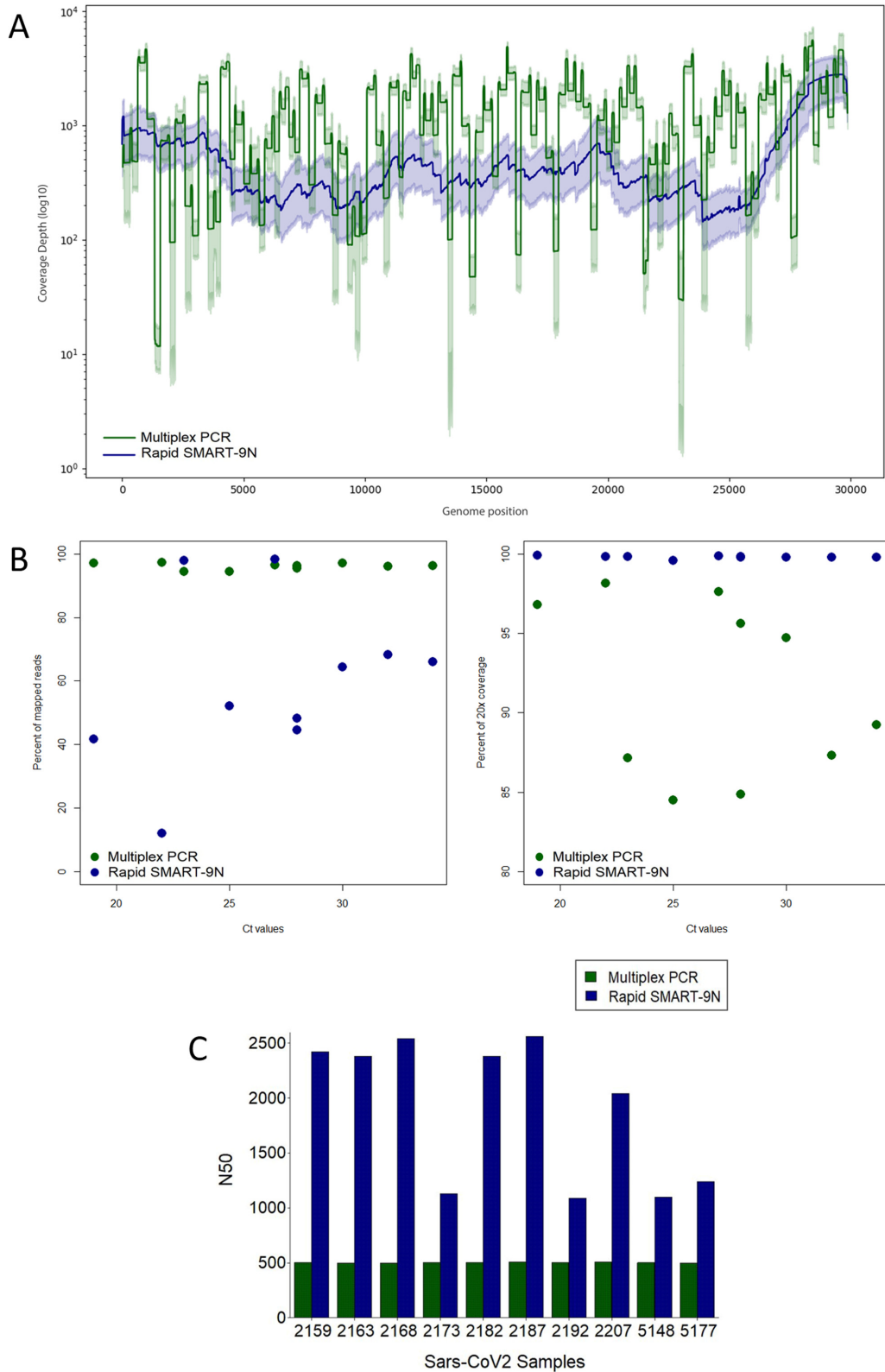


Figure 4. Comparison of multiplex PCR, and Rapid SMART-9N results for SARS-CoV-2 clinical samples. A) Average genome coverage depth and 95% of reads mapped to the genome reference position. **B)** Proportion of reads mapping to the reference genome across a range of Ct-values (left) and percentage of 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).

already sequenced with the multiplex PCR method, with Ct-values between 4.6 and 33 were selected (*Extended data: Table S6*). 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*). 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*). The average coverage depth revealed higher genome depth and better coverage pattern across the genome for the metagenomics methods when compared to the targeted 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*).

20-fold genome coverage across the Ct-values was compared between all methods, 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*).

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 1.58kbp for SMART-9N, 705bp to 2.16kb for Rapid SMART-9N. The median was 599.8bp, 1.6kbp, and 1.2kbp 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. Due to the emergence of SARS-CoV-2 during the validation of the protocols, we chose to test for SARS-CoV-2 only with the Rapid SMART-9N protocol, for being a faster and promising technique to be used in the course of the pandemic.

Reads mapping to reference virus genome (isolate: Wuhan-Hu-1, GenBank Accession No. MN908947) were present in all ten 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 10 samples and the lowest coverage depth of 97.51x (*Figure 4A*). When comparing each coverage depth across 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 for the multiplex PCR reactions 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 are 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 panels¹⁵. The etiology 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 history³¹, strain identification^{32,33}, and biology of new pathogens¹⁴. This is evidenced by the recent rapid and impactful metagenomic analysis of SARS-CoV-2 early in the pandemic^{34,35}.

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 method⁵⁻⁹.

For ZIKV isolated-culture, it was possible to obtain 99.02% of genome coverage with an input of 6 FFU/mL, an amount comparable to other single-cell methods available^{36,37} 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. 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 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^{16,17}. 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. The average N50 for the clinical samples using the SMART-9N was 1.2kb and for Rapid SMART-9N was 1.6kb, a difference we believe can be explained by the different tag sequence. When a single PCR primer is used any templates that self-anneal will not amplify resulting in an enrichment of longer products for SMART-9N. When looking at the average coverage depth and the CI of the metagenomics methods, we observe consistent amplification across the entire genome. Increased N50 provides higher confidence in individual read taxonomic assignment, improves mapping confidence, *de novo* assembly, and the ability to detect viral recombinations^{38,39}. 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 approach²⁰. Using this approach, our Rapid SMART-9N reduced the complexity, time, and cost from sample to sequence. The addition of the barcode during PCR decreased the library preparation time from 6 hours to 10 minutes, reducing 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 57% 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 from 72 to 144. 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.

The YFV and SARS-CoV-2 clinical samples were also analysed in an untargeted way, mapped to an available reference database and analysed manually in order to screen potential microbial contamination and/or co-infections. The methods allowed the identification of an unknown co-infection in a YFV clinical sample, a dsDNA virus genus Pa6virus, family Siphoviridae, had the full genome characterized. We showed that our non-targeted sequencing approach offers an opportunity for simultaneous testing for a wide range of potential pathogens, providing a faster route to identification followed by a potential specific treatment.

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 poor storage conditions. This is a limiting factor for the sensitivity of the approach that could result in low or no coverage of the infectious agent. While the DNase treatment dramatically improves sensitivity, more work is needed in depleting highly abundant rRNA species which are recovered as a result of the random priming. Lower sensitivity is seen in our study when comparing the number of viral reads from the ZIKV isolates to the YFV and SARS-CoV-2 clinical samples. The reduction in the number of viral reads as the Ct-value increases is due to the total level of non-viral host/background nucleic acid present and provides an upper limit for the approach above which amplicon sequencing is more useful. The difference we observed in N50 between SMART-9N and Rapid SMART-9N cannot be easily resolved so we recommend using SMART-9N for best representation and Rapid SMART-9N when speed is more important.

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

All raw files with the host reads depleted, and consensus sequences generated in this study can be found at <https://github.com/CADDE-CENTRE/Rapid-RNA-SMART-Metagenomics>

Repository: CADDE-CENTRE/Rapid-RNA-SMART-Metagenomics, DOI: [10.5281/zenodo.5391968](https://doi.org/10.5281/zenodo.5391968). License CC0.

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).

Extended data

Repository: CADDE-CENTRE/Rapid-RNA-SMART-Metagenomics/Supplementary_material_SMART_9N.pdf, DOI: [10.5281/zenodo.5391968](https://doi.org/10.5281/zenodo.5391968). License CC0.

This project contains the following extended data:

- 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=41), 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=21), 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).

Author contributions

Ingra M. Claro

Roles: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Mariana S. Ramundo

Roles: Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing

Thais Coletti

Roles: Methodology, Validation, Visualization

Camila A. M. da Silva

Roles: Methodology, Validation, Visualization

Ian N. Valenca

Roles: Methodology, Validation, Visualization

Darlan S. Candido

Roles: Methodology, Validation, Visualization

Flavia C. S. Sales

Roles: Methodology, Validation, Visualization, Writing – Original Draft Preparation

Erika R. Manuli

Roles: Methodology, Validation, Visualization

Jaqueline G. de Jesus

Roles: Methodology, Validation, Visualization

Anderson de Paula

Roles: Methodology, Validation, Visualization

Alvina Clara Felix

Roles: Methodology, Validation, Visualization

Pamela Andrade

Roles: Methodology, Validation, Visualization

Mariana Pinho

Roles: Methodology, Validation, Visualization

Esper G. Kallas

Roles: Methodology, Validation, Visualization

José Eduardo Levi

Roles: Methodology, Validation, Visualization

Mariene R. Amorim

Roles: Methodology, Validation, Visualization

William M. Souza

Roles: Writing – Original Draft Preparation, Writing – Review & Editing

José Luiz Proenca-Modena

Roles: Methodology, Validation, Visualization

Nuno Rodrigues Faria

Roles: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Ester C. Sabino

Roles: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Nicholas J. Loman

Roles: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Software, Resources, Validation, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Joshua Quick

Roles: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Resources, Validation, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

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 fundings.

References

- Devaux CA: **Emerging and re-emerging viruses: A global challenge illustrated by Chikungunya virus outbreaks.** *World J Virol.* 2012; **1**(1): 11–22. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Miller S, Naccache SN, Samayoa E, et al.: **Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid.** *Genome Res.* 2019; **29**(5): 831–842. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Washington JA: **Principles of Diagnosis.** Chapter 10. In: Baron S, editor. *Medical Microbiology.* 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996; 10. [PubMed Abstract](#)
- Wilson MR, Naccache SN, Samayoa E, et al.: **Actionable diagnosis of neuroleptospirosis by next-generation sequencing.** *N Engl J Med.* 2014; **370**(25): 2408–17. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quick J, Loman NJ, Duraffour S, et al.: **Real-time, portable genome sequencing for Ebola surveillance.** *Nature.* 2016; **530**(7589): 228–232. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faria NR, Quick J, Claro IM, et al.: **Establishment and cryptic transmission of Zika virus in Brazil and the Americas.** *Nature.* 2017; **546**(7658): 406–410. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Naveca FG, Claro I, Giovanetti M, et al.: **Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon.** *PLoS Negl Trop Dis.* 2019; **13**(3): e0007065. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- de Jesus JG, Dutra KR, Sales FCDS, et al.: **Genomic detection of a virus lineage replacement event of dengue virus serotype 2 in Brazil, 2019.** *Mem Inst Oswaldo Cruz.* 2020; **115**: e190423. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faria NR, Kraemer MUG, Hill SC, et al.: **Genomic and epidemiological monitoring of yellow fever virus transmission potential.** *Science.* 2018; **361**(6405): 894–899. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Houldcroft CJ, Beale MA, Breuer J: **Clinical and biological insights from viral genome sequencing.** *Nat Rev Microbiol.* 2017; **15**(3): 183–192. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Palacios G, Druce J, Du L, et al.: **A new arenavirus in a cluster of fatal transplant-associated diseases.** *N Engl J Med.* 2008; **358**(10): 991–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Nakamura S, Yang CS, Sakon N, et al.: **Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach.** *PLoS One.* 2009; **4**(1): e4219. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manso CF, Bibby DF, Mbisa JL: **Efficient and unbiased metagenomic recovery**

- of RNA virus genomes from human plasma samples. *Sci Rep.* 2017; **7**(1): 4173. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Gu W, Miller S, Chiu CY: **Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection.** *Annu Rev Pathol.* 2019; **14**: 319–338. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Frey KG, Herrera-Galeano JE, Redden CL, et al.: **Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood.** *BMC Genomics.* 2014; **15**: 96. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, et al.: **Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples.** *Euro Surveill.* 2018; **23**(50): 1800228. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 17. Lewandowski K, Xu Y, Pullan ST, et al.: **Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples.** *J Clin Microbiol.* 2019; **58**(1): e00963–19. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Reyes GR, Kim JP: **Sequence-independent, single-primer amplification (SISPA) of complex DNA populations.** *Mol Cell Probes.* 1991; **5**(6): 473–481. [PubMed Abstract](#) | [Publisher Full Text](#)
 19. Zhu YY, Machleder EM, Chenchik A, et al.: **Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction.** *Biotechniques.* 2001; **30**(4): 892–7. [PubMed Abstract](#) | [Publisher Full Text](#)
 20. Quick J, Grubaugh ND, Pullan ST, et al.: **Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples.** *Nat Protoc.* 2017; **12**(6): 1261–1276. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 21. Lanciotti RS, Kosoy OL, Laven JJ, et al.: **Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007.** *Emerg Infect Dis.* 2008; **14**(8): 1232–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 22. Silva-Filho JL, de Oliveira LG, Monteiro L, et al.: **Gas6 drives Zika virus-induced neurological complications in humans and congenital syndrome in immunocompetent mice.** *Brain Behav Immun.* 2021; **97**: 260–274. [PubMed Abstract](#) | [Publisher Full Text](#)
 23. Fischer C, Torres MC, Patel P, et al.: **Lineage-Specific Real-Time RT-PCR for Yellow Fever Virus Outbreak Surveillance, Brazil.** *Emerg Infect Dis.* 2017; **23**(11): 1867–71. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 24. Corman VM, Landt O, Kaiser M, et al.: **Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.** *Euro Surveill.* 2020; **25**(3): 2000045. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 25. Tyson JR, James P, Stoddart D, et al.: **Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore.** *bioRxiv.* [Preprint]. 2020; 2020.09.04.283077. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 27. Li H, Handsaker B, Wysoker A, et al.: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics.* 2009; **25**(16): 2078–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 28. De Coster W, D'Hert S, Schultz DT, et al.: **NanoPack: visualizing and processing long-read sequencing data.** *Bioinformatics.* 2018; **34**(15): 2666–2669. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 29. Milne I, Bayer M, Cardle L, et al.: **Tablet–next generation sequence assembly visualization.** *Bioinformatics.* 2010; **26**(3): 401–2. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 30. Kallas EG, D'Elia Zanella LGFAB, Moreira CHV, et al.: **Predictors of mortality in patients with yellow fever: an observational cohort study.** *Lancet Infect Dis.* 2019; **19**(7): 750–758. [PubMed Abstract](#) | [Publisher Full Text](#)
 31. Gire SK, Goba A, Andersen KG, et al.: **Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak.** *Science.* 2014; **345**(6202): 1369–72. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 32. Salipante SJ, SenGupta DJ, Cummings LA, et al.: **Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology.** *J Clin Microbiol.* 2015; **53**(4): 1072–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 33. Deurenberg RH, Bathoorn E, Chlebowicz MA, et al.: **Application of next generation sequencing in clinical microbiology and infection prevention.** *J Biotechnol.* 2017; **243**: 16–24. [PubMed Abstract](#) | [Publisher Full Text](#)
 34. Carbo EC, Sidorov IA, Zevenhoven-Dobbe JC, et al.: **Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness.** *J Clin Virol.* 2020; **131**: 104594. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 35. Zhou P, Yang XL, Wang XG, et al.: **A pneumonia outbreak associated with a new coronavirus of probable bat origin.** *Nature.* 2020; **579**(7798): 270–273. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. Nawy T: **Single-cell sequencing.** *Nat Methods.* 2014; **11**(1): 18. [PubMed Abstract](#) | [Publisher Full Text](#)
 37. Eberwine J, Sul JY, Bartfai T, et al.: **The promise of single-cell sequencing.** *Nat Methods.* 2014; **11**(1): 25–7. [PubMed Abstract](#) | [Publisher Full Text](#)
 38. Stephens Z, Wang C, Iyer RK, et al.: **Detection and visualization of complex structural variants from long reads.** *BMC Bioinformatics.* 2018; **19**(Suppl 2): 508. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Fuselli S, Baptista RP, Panziera A, et al.: **A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (*Rupicapra rupicapra*).** *Heredity (Edinb).* 2018; **121**(4): 293–303. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 16 May 2023

<https://doi.org/10.21956/wellcomeopenres.21114.r56455>

© 2023 Cunningham-Oakes E. This is an open access peer review report 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.

 **Edward Cunningham-Oakes** 
University of Liverpool, Liverpool, UK

The authors have responded to all revisions, and I have no further comments to make.

Competing Interests: No competing interests were disclosed.



Reviewer Expertise: Metagenomics, microbial genomics, bioinformatics, pathogen detection

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 04 May 2023

<https://doi.org/10.21956/wellcomeopenres.21114.r56454>

© 2023 D'arc Ferreira da Costa M. This is an open access peer review report 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.

 **Mirela D'arc Ferreira da Costa** 
Instituto de Biologia, Departamento de Genética, Laboratório de Diversidade e Doenças Virais,
Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

I confirm that the authors have reviewed all critical points cited in the first version. I have only two small suggestions to improve paper quality: 1. Standardize the way to indicate the FFU/mL (ex. 6×10^7 instead of $6e7$) everywhere (main text and tables); and 2. Several numbers in the Supplementary Material have the decimal place separated by commas and It is important to specify that "N50" and "Longest read" are in kilobases (kbs). Therefore, I believe the article is now refined and accurate enough for wide release. Finally, it is important to emphasize that the article

makes an important contribution to the field of metagenomics.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular biology, viral metagenomics and bioinformatics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 03 October 2022

<https://doi.org/10.21956/wellcomeopenres.18969.r52412>

© 2022 D'arc Ferreira da Costa M. This is an open access peer review report 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.



Mirela D'arc Ferreira da Costa

Instituto de Biologia, Departamento de Genética, Laboratório de Diversidade e Doenças Virais, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

The authors describes a improvement of the SMART approach published in 2001. They developed a rapid viral metagenomics test using initially a random priming for cDNA synthesis and then the SMART-9N barcoded amplification. I believe that it is a important work and should be indexed after a good grammar and concept revision. Some points that should be verified are listed below.

1. In "Introduction" section, it is not clear if the "random priming for cDNA synthesis followed by PCR amplification" refers to the SMART-9N methodology.
2. The ethical approval number for the YFV study is absent in the mean text.
3. The reagents concentration should be more clear in all steps. It is very important to allow replication of the method by others.
4. Some decimal values are incorrectly separated by commas (for example, 15,75 µl Nuclease-free water).
5. In "Bioinformatics workflow" section, it is necessary to introduce the citation of all softwares/tools used (for example, minimap2). Also, the duplicated brackets after SARS-CoV-2 should be removed.
6. It is not clear why the "Clean up and quantification" step has a dramatic time reduction, specially between SMART-9N and Rapid SMART-9N methodologies.
7. In "Rapid SMART-9N of SARS-CoV-2 clinical samples" section, the phrase "were present in all

ten seven samples" is not understandable.

8. It is not clear why the SMART-9N approach was not performed to the SARS-CoV-2 clinical samples.

9. The specific viral results obtained by Kraken classification should also be addressed in the Discussion/Limitations section.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular biology, viral metagenomics and bioinformatics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 August 2022

<https://doi.org/10.21956/wellcomeopenres.18969.r51813>

© 2022 Cunningham-Oakes E. This is an open access peer review report 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.



Edward Cunningham-Oakes 

University of Liverpool, Liverpool, UK

The study presented by Claro *et al.*

Introduction

1. Minor point, but not essential: I would encourage a quick review of punctuation/grammar for some of the longer sentences in the introduction. Alternatively, I would break up some of these longer sentences to help the reader.
2. The abbreviation CHIKV is not introduced elsewhere as Chikungunya virus before abbreviation.
3. As per previous point, LASV is not introduced as Lassa mammarenavirus before use of the abbreviation.
4. Missing word (highlighted in bold): 'SMART-9N recovered a high proportion of viral reads from **a** ZIKV isolate.'
5. In the sentence "enabling enhanced pathogen detection for both diagnostic and surveillance application of RNA viruses", amend "diagnostic" to "diagnosis", and remove the word "application".
6. Depending on how this is published, it might be worth introducing "SMART" as (Switching Mechanism at the 5' end of RNA Template), as you have done in the abstract.

Methods

Sample collection

1. "The supernatant was removed, and MEM supplemented with 2% fetal bovine serum, 1% penicillin, and streptomycin were added" - it is not necessary to add "were added" to the end of this sentence, as you have already stated "supplemented".
2. What percentage of streptomycin was the fetal bovine serum supplemented with?
3. I assume penicillin and streptomycin were added to prevent bacterial growth, but it may be worth stating why your MEM was supplemented with them.
4. Please describe the term "FFU" in full, before providing the acronym.
5. Stylistic suggestion "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". I would split this sentence into two, after the first mention of "Rapid SMART-9N", correct "metagenomics approaches" to "metagenomic approaches", and correct "was tested" to "were tested".
6. "with a cut-off value of Ct-values ≤ 37 " -> "with a ct-value cut-off of"
7. "samples with a range of Ct-values" - can Ct-value ranges be stated in text? I know that a supplementary table is provided, but having a range in text would help readers with following the narrative without switching between supplementaries and the main text.

Multiplex PCR

1. Under the section "Multiplex tiling PCR", can the authors clarify what is meant by "specific bands"? Is there an amplicon size you are expecting on the gel for PCR products?

Bioinformatics workflow

1. The workflow presented is a robust and methodologically sound approach. However, as raised by the authors, the advantage of metagenomic sequencing is the genomic characterization of known and novel viruses in an untargeted manner, whilst this workflow uses mapping to a known reference genome, based on knowledge of what is in the clinical samples (e.g. SARS-CoV-2 nasopharyngeal swabs). It would probably be of wider interest (in the discussion?) as how you would see SMART sequencing being used if you were presented with a sample with no known clinical diagnostic. I assume this would be an extension of your method to detect other RNA viruses in these samples?
2. I also noted that you used results from Kraken2 to identify a co-infection in a YFV sample, and then you obtain a reference genome to assess coverage. This is not mentioned in your methods, though it is described in the results (I also assume this would be your approach for samples with no known pathogen based on clinical diagnostics?)
3. Minor typo "samtools stats and samtools depth were used to calculate longest reads and genome coverage at 20x respectively." - a capital is needed at the start of this sentence.

Results

1. "This approach allows for amplification of RNA in the picogram input range (data not shown)" - are there plans to include this data as a supplementary? It would be exciting to see, and also extremely relevant to clinical samples, which can be of varying concentration and quality.
2. Figure 1 - no amendments to make, but I wanted to comment that this was a very helpful visual.
3. *Multiplex PCR sequencing of ZIKV isolate and YFV and SARS-CoV-2 clinical samples*
4. Figure 2 - the figure legend (and manuscript text) refers to FFU/ml, but the x-axis in panel B shows PFU/ml
5. I note that detection limits are only assessed for SMART-9N and Rapid SMART-9N methods. Do we already know the limits of detection for multiplex PCR?
6. "The sequenced YFV samples (n=21) had a median Ct-values = 25.57, between 5 and 37 generated in one 24 barcoded library." - can the authors check this sentence, as I could not quite follow it?
7. Minor/optional point: I would keep the format ranges for genome copies consistent. In the section describing YFV, you go from low to high, (1.00e00 to 1.50e10), but here, you go from high to low (2.40e05 to 1.30e02). I would go from low to high.

8. Minor typo: with an average depth **of** between 821.77x to 1570x

SMART-9N and Rapid SMART-9N of ZIKV isolated-culture samples and limit of detection

1. "and submitted to SMART-9N" - subjected to?
2. "The coverage depth was up to 10010x, and 154.25x with 6e00 FFU/mL of material, compatible with single-cell assays." - do you mean ranged between? Up to implies one of these numbers is the upper limit, but you wouldn't usually report the other if you are describing the upper limit in this way.
3. "The median N50 was 1.7kb and when the readings" - should this be "reads?"
4. Should "ZIKV genome covers" be "ZIKV genome coverage?"
5. "and the N50 2.27kb" -> "and the N50 **was** 2.27kb"
6. "with 6e07 of material input" - please state units. Was this FFU/mL?

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

8. "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." - presumably, the multiplex PCR method was specific to YFV. Do the authors know have any thoughts as to why performance is worse with a pathogen-specific assay?
9. Minor typos/sentence structure issue: "Comparable genome coverage 20-fold across the Ct-values between all methods is presented in Figure 3B" - could the authors look at this sentence again, as it is a little bit hard to interpret.

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

1. Would the authors be able to clarify as to why SARS-CoV-2 were only subjected to Rapid SMART-9N?
2. "The median revealed was 91.59%, minimum 84.49%, and the Rapid SMART-9N 99.79%, minimum of 99.57%." I assume the first two values are for multiplex PCR, and the final two for Rapid SMART-9N, but this could be made clearer.

3. Detection of other RNA viruses in clinical samples and Kraken classification

4. "This allowed for the identification of a dsDNA virus genus Pa6virus, family *Siphoviridae* present in one YFV sample" - you mention (at the end of this section), the proportion of viral reads in other samples. Was there a reason that you chose to report on the Pa6virus in a YFV sample, as opposed to other identified viral reads in other samples? Was this the only evidence of putative co-infection? It may be worth noting your rationale for choosing this sample in-text.

Discussion

Please see my note in methods regarding bioinformatic workflow in samples with no known

clinical diagnostic result.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Metagenomics, microbial genomics, bioinformatics, pathogen detection

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
