



OPEN ACCESS

EDITED BY

Fuqing Wu,
The University of Texas Health Science Center
at Houston, United States

REVIEWED BY

Erika Cione,
University of Calabria, Italy
David W. Ussery,
University of Arkansas for Medical Sciences,
United States

*CORRESPONDENCE

Jiannis Ragoussis,
✉ ioannis.ragoussis@mcgill.ca

RECEIVED 24 October 2024

ACCEPTED 11 July 2025

PUBLISHED 15 August 2025

CITATION

Bayega A, Reiling SJ, Liu JL, Dubuc I, Gravel A,
Flamand L and Ragoussis J (2025) A Benchmark
of methods for SARS-CoV-2 whole genome
sequencing and development of a more
sensitive method.
Front. Genet. 16:1516791.
doi: 10.3389/fgene.2025.1516791

COPYRIGHT

© 2025 Bayega, Reiling, Liu, Dubuc, Gravel,
Flamand and Ragoussis. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

A Benchmark of methods for SARS-CoV-2 whole genome sequencing and development of a more sensitive method

Anthony Bayega^{1,2}, Sarah J. Reiling^{1,2}, Ju Ling Liu^{1,2},
Isabelle Dubuc³, Annie Gravel³, Louis Flamand^{3,4} and
Jiannis Ragoussis^{1,2,5*}

¹McGill Genome Centre, Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montreal, QC, Canada, ²Department of Human Genetics, McGill University, Montreal, QC, Canada, ³Division des maladies infectieuses et immunitaires, Centre de Recherche du Centre Hospitalier Universitaire de Québec - Université Laval, Quebec City, QC, Canada, ⁴Département de microbiologie-infectiologie et d'immunologie, Faculté de médecine, Université Laval, Quebec City, QC, Canada, ⁵Department of Bioengineering, McGill University, Montréal, QC, Canada

The raging COVID-19 pandemic caused by SARS-CoV-2 has so far claimed the lives of 7 million people and continues to infect many more. Further, virus evolution has caused mutations that have compromised public health interventions like vaccination regimes and monoclonal antibody and convalescent sera treatments. In response, unprecedented large-scale whole genome viral surveillance approaches have been devised to keep track of the evolution and transmission patterns of the virus within and across populations. Here, we aimed to compare efficiencies of SARS-CoV-2 whole genome sequencing approaches using synthetic SARS-CoV-2 genome and six cell culture SARS-CoV-2 variants titrated to represent samples at high, medium, and low viral load. We found that the ARTIC protocols performed best in terms of PCR amplicon yield returning 67% more amplicons than Entebbe protocol which was the second highest PCR amplicon yielding protocol. ARTIC v4.1 protocol yields were only slightly better than ARTIC v3. Despite yielding the lowest PCR amplicons, the SNAP protocol showed the highest genome completeness using a synthetic genome at high viral titre followed by ARTIC protocols. However, the ARTIC protocols showed highest genome completeness with cell culture SARS-CoV-2 variants across high, medium and low viral titres. ARTIC protocol also performed best in calling the correct lineage among cell culture SARS-CoV-2 variants across different viral titres. We also designed a new method termed ARTIC-Amp which leverages ARTIC protocol and performs a rolling circle amplification to increase yield of amplicons. In a proof-of-principle experiment, this method showed 100% coverage in all four targeted genes across three replicates unlike the ARTIC protocol missed one gene in two of the three replicates. Our results demonstrate the robustness of the ARTIC protocol and propose an improved method that could be useful for samples that routinely have limited SARS-CoV-2 RNA such as wastewater samples.

KEYWORDS

artic, nanopore sequencing, rolling circle amplification, COVID-19, SARS-CoV-2, wastewater, benchmarking

Introduction

As of October 2024, coronavirus disease 2019 (COVID-19) pandemic has resulted in at least 776 million infections and 7 million deaths worldwide since its inception in December 2019 WHO (World Health Organization), 2024. The causative agent of COVID-19 was quickly identified as a respiratory virus of Betacoronavirus genus and named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wu et al., 2020; Zhu et al., 2020). This enveloped virus is composed of a 29,903 bp positive single stranded RNA genome that encodes four structural (Spike, Membrane, Envelope, and Nucleocapsid) and at least 29 non-structural proteins (Lu et al., 2020; Wu et al., 2020). The NSP12 gene encodes an RNA-dependent RNA polymerase (RdRp) which is responsible, along with other viral and host co-factors, for replication of the viral genome once inside the cell. Despite the proofreading activity of the RdRp conferred by NSP14, the evolutionary rate of SARS-CoV-2 measured as inter-host temporal variation in consensus sequence is estimated at $1.5 \pm 0.5 \times 10^{-3}$ per site per year (Domingo et al., 2021). This viral evolution creates mutations that are of huge public health concern. Indeed, it was noted in November 2020 that new mutations were arising that eventually led to more transmissible and vaccine evading variants of SARS-CoV-2 (Davies et al., 2021; Thorne et al., 2022; Walker et al., 2021). This alerted scientists of the need to perform large-scale whole genome sequencing (WGS) as a surveillance method. Large scale WGS in turn elucidated the worldwide spread of variants and led to identification of new variants. Currently SARS-CoV-2 variants are of three classes: variants of concern (VOCs), variants of interest (VOIs), and variants under monitoring (VUMs). In 2023, VOCs included the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (BA.1) (Harvey et al., 2021; Mendiola-Pastrana et al., 2022). Viral surveillance is critical to track the evolution and transmission patterns of these VOC within and across communities and quickly identify new mutations and variants. Further, surveillance facilitates tracking of the impact of mutations on public health and public health interventions such as vaccinations, monoclonal and convalescent sera treatment, drugs, and personal hygiene and protection interventions like masks and number of metres required for social distancing.

Real-time quantitative PCR (RT-qPCR) is the gold standard method for diagnostics and surveillance with very high sensitivity. This quantitative method provides a cycle threshold metric (Ct value) that provides a linear inverse relationship with amount of viral RNA in a sample. The drawback of RT-qPCR, however, is that no sequence information is produced. Whole genome sequencing methods have thus been used for SARS-CoV-2 surveillance. A widely used SARS-CoV-2 WGS protocol was developed by an international workgroup comprising scientists from UK, Belgium, and USA called Advancing Real-Time Infection Control network (ARTIC). Primers used in the ARTIC protocol have undergone at least three major revisions in response to SARS-CoV-2 evolution giving rise to versions 3, 4, and 4.1. Several other SARS-CoV-2 whole genome sequencing protocols have been made available either by academic research labs or companies. Although benchmarking studies have been attempted including external quality proficiency testing (Liu et al., 2021; Plitnick et al., 2021), it is not clear how the different methods compare to each other and how they perform with low viral titre samples such as environmental samples.

Despite their low viral titre, environmental samples have shown promise as an alternative to clinical samples for early detection of

SARS-CoV-2 variants. Municipal wastewater surveillance, for example, has potential to provide passive population-scale surveillance and has been shown to identify new SARS-CoV-2 variants up to 3 weeks before clinical identification (Ahmed et al., 2021; Larsen and Wigginton, 2020; Smyth et al., 2022). One of the biggest drawbacks of environmental samples, particularly municipal wastewaters, is the very low amounts of SARS-CoV-2 with Ct values routinely above 35 (Jafferali et al., 2021). This can be attributed partly to low recovery rates of methods used to extract nucleic acids that can be as low as 0%–25% and presence of inhibitors of molecular assays in extracted wastewater samples (Jafferali et al., 2021). It is thus imperative to develop very sensitive methods to enable reliable early detection of SARS-CoV-2.

Here, we aimed to compare the performance of five SARS-CoV-2 whole genome sequencing protocols including the ARTIC versions 3 and 4.1, QIAseq DIRECT SARS-CoV-2 (Qiagen, Hilden, Germany), Swift Normalase Amplicon Panel for SARS-CoV-2 additional genome completeness (SNAP, Swift Bioscience), Midnight protocol (Freed et al., 2020), and Entebbe protocol (Cotten et al., 2020) (see Figure 1). After finding that all protocols performed poorly with low viral titre samples we sought to design a custom protocol. In the custom protocol which we termed ARTIC-Amp (for amplified ARTIC), we leveraged the ARTIC v4.1 protocol which we found to perform best by using amplicons generated from the ARTIC v4.1 protocol to perform another round of isothermal rolling circle amplification as applied recently (Volden et al., 2018). In a proof-of-principle experiment (see Figure 1), the ARTIC-Amp protocol showed improved performance over regular ARTIC protocol.

Materials and methods

Sample preparation: synthetic SARS-CoV-2 genome

We purchased the Twist Synthetic SARS-CoV-2 RNA Control 1 (102019, Twist Bioscience, CA, USA). This control consists of six non-overlapping 5 kb ssRNA fragments covering 99.9% of the viral genome (GenBank ID MT007544.1, GISAID NAME Australia/VIC01/2020) and is reconstituted at 1,000,000 copies per microliter. We serially diluted the control in water to generate solutions at 1×10^6 , 1×10^5 , 1×10^4 , 1×10^3 , and 1×10^2 copies/mL. These were stored at -80°C .

Sample preparation: cell culture SARS-CoV-2 virus

We obtained SARS-CoV-2 virus samples as follows: B.1 and B.1.1.7 were obtained from the Laboratoire de Santé Publique du Québec (LSPQ), B.1.351 and P.1 were obtained from BEI Resources (VA, USA), B.1.617.2 was obtained from Canada's National Microbiology Laboratory (NML), and BA.1 was obtained from British Columbia Centre for Disease Control (BCCDC, Canada). The samples were cultured in Vero cells at a multiplicity of infection of 0.002 for 4–5 days. The supernatants were collected and centrifuged to remove cells and debris. Total RNA was extracted using the Bead Mill Tissue RNA purification kit (26-010B, OMNI International, GA, USA). Briefly (see

TABLE 1 Cell culture wildtype SARS-CoV-2 and its variants used in this study and their RT-qPCR cycle threshold (Ct) values and digital droplet PCR (ddPCR) concentrations (ND, not determined).

No.	WHO label	PANGO lineage	First detection	Ct value	ddPCR Conc'n (copies/ μ L)
1	Wild-type	B.1	WU	15.21	2.1×10^5
2	Alpha	B.1.1.7	UK	12.06	9.5×10^5
3	Beta	B.1.351	South Africa	11.19	2.8×10^6
4	Gamma	P.1	Brazil	10.41	5.8×10^6
5	Delta	B.1.617.2	India	15.46	1.3×10^5
6	Omicron	BA.1	South Africa	12.62	ND

full protocol in [Supplementary Material](#)), 300 μ L of supernatant were mixed with ceramic beads, 300 μ L of RLB buffer, 10 μ L of antifoam reagent, and 12 μ L of β -mercaptoethanol. Following homogenization, the solution was centrifuged, reconstituted with one volume of 70% ethanol and vortexed. The solution was then transferred to an Omni RNA column, where the RNA was hybridized to the column, washed, and eluted in DEPC-treated water. RNAs were reverse transcribed to cDNA using SuperScript[™] IV VILO[™] mastermix, as per manufacturer's recommendations (ThermoFisher Scientific). Quantitative real-time PCR (RT-qPCR) was then performed to determine cycle threshold values (Ct, [Table 1](#)). SARS-CoV-2 Envelope (E) gene forward (5'-ACAGGTACGTTAATAGTTAATAGCGT-3') and reverse (5'-ATATTGCAGCAGTACGCACACA-3') primers, along with the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories Ltd.) were used. RNAs were also diluted at 1:100 and quantified by digital droplet PCR (ddPCR) using the One-Step RT-ddPCR Advanced kit for Probes (Bio-Rad Laboratories Ltd.) following manufacturer's recommendations. Primers used for ddPCR included the SARS-CoV-2 E primers used for qPCR experiments and the SARS-CoV-2 probe (5'-ACACTAGCCATCCTTACTGCGCTTCG-3'). All samples were normalized to 1,000 copies/ μ L and then further log-serial diluted to generate new samples at 100, 10, 1, and 0.1 copies/ μ L. Eleven microliters of each sample was then used as input template for reverse transcription. For the main experiments, we used samples at 100, 1, and 0.1 copies/ μ L to simulate clinical samples at high viral load, medium viral load, and low viral load, respectively.

ARTIC protocol

We used the ARTIC workflow as described by the authors [<https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuk6w> (Tyson et al., 2020)]. Briefly, 11 μ L of sample was reverse transcribed using LunaScript RT SuperMix (NEB, MA, USA) in a 20 μ L reaction. We used the Q5 Hot Start High-Fidelity 2X Master Mix (NEB, MA, USA) to prepare two separate pools of PCR mixes corresponding to the two primer pools in the ARTIC workflow. We used both ARTIC version 3 and version 4. 1 primers. Five micro litres of cDNA mix was directly added to 20 μ L of PCR mix and amplified for 36 cycles. Amplicons were purified using SPRI paramagnetic beads (Beckman Coulter, IN, USA) and quantified using either Qubit fluorometer (Thermo Fischer Scientific) or Quant-iT PicoGreen dsDNA Assay Kits (Thermo Fischer Scientific). Samples were barcoded using Native

Barcoding Expansion Kit 96 (EXP-NBD196, Oxford Nanopore Technologies, UK). To multiplex samples, we used 150 ng of purified amplicons or 15 μ L of sample for samples with less than 10 ng/ μ L. The multiplexed library was purified once more followed by sequencing on the PromethION (R9.4.1 flow cells, Oxford Nanopore Technologies, UK).

Swift Normalase Amplicon Panel (SNAP) protocol

The Swift Normalase Amplicon Panel (SNAP, Swift Bioscience) for SARS-CoV-2 Additional Genome Coverage was used according to manufacturer instructions, except where mentioned otherwise in this manuscript. Briefly, cDNA was generated as described for ARTIC protocol and 10 μ L of cDNA added directly into each PCR tube followed by 20 μ L of a master mix composed of 2 μ L Reagent G1 (primer set), 3 μ L Reagent G2, and 15 μ L Enzyme G3. We used the Low Viral Load Input Recommendations and thus amplified the cDNA for a total of 28 cycles, except for Omicron (BA.1) samples. Amplicons were purified, quantified, pooled, and sequenced as described for the ARTIC protocol.

Qiaseq protocol

The QIAseq DIRECT SARS-CoV-2 (Qiagen, Hilden, German) protocol, herein referred to as Qiaseq was used according to manufacturer instructions, except for the cDNA synthesis step which was performed as described for the ARTIC protocols. We followed the guidelines for "Samples with broad/unknown range Ct value" and thus performed 29 PCR cycles. Amplicons were purified, quantified, pooled, and sequenced as described for the ARTIC protocol.

Midnight protocol

The Midnight protocol was carried out as described by its authors [<https://www.protocols.io/view/ncov-2019-sequencing-protocol-rapid-barcoding-1200-bgggjtww> (Freed et al., 2020)]. cDNA synthesis, amplification, purification, pooling, and sequencing were carried out as described for the ARTIC protocol except that during PCR amplification we used 2.5 μ L of cDNA for

each of the two 1,200 bp primer sets and PCR extension temperature was 65 °C as recommended by the Midnight protocol developers.

Entebbe protocol

The Entebbe protocol (Cotten et al., 2020) was carried out with some changes to the instructions of the developers. Firstly, cDNA synthesis was carried out as described for the ARTIC protocol that uses random hexamers rather than gene-specific primers as instructed in the Entebbe protocol. cDNA amplification, purification, pooling, and sequencing were carried out as described for the ARTIC protocol except that PCR amplification was carried out following the conditions recommended by Entebbe protocol developers (Cotten et al., 2020).

Noteworthy, we used LunaScript RT SuperMix to generate enough cDNA in one batch from each sample that would be used to evaluate all protocols and thus minimise batch effects. The cDNA was used as input and as specified for each protocol: 5 µL for each of the two pools for ARTIC, Entebbe, and custom protocols, 2.5 µL for each of the two pools of Midnight protocol, and 10 µL for the single pool of SNAP protocol. PCR amplification was done as specified for each protocol: 36 cycles for ARTIC, Midnight, Entebbe, and custom protocol, 29 for Qiaseq protocol, and 28 cycles for SNAP (except for BA.1 where samples were amplified for 36 cycles).

ARTIC-Amp protocol

We adapted the R2C2 protocol developed by Volden et al. (2018) to create the custom protocol we termed ARTIC-Amp for amplified ARTIC v4.1 (see [Supplementary Material](#) for the full protocol). In a proof-of-principle experiment, four primers among the ARTIC version 4.1 primers were selected as shown in [Table 2](#). For each of the left primers we added the adapter sequence “AATGATACGGCGACCACCGAGATCTACAC” and for each of the right primers we added the adapter “AAGCAGTGGTATCAACGCAGAGT”. We also obtained the Lambda_Splint_F “ACTCTGCGTTGATACCACTGCTTAAAGGGATATTTTCGATCGCTTG” and Lambda_Splint_R “ATCTCGGTGGTCGCCGTATCATTTGAGGCTGATGAGTTCCATATTTG” primers to amplify a 330 bp fragment of the Lambda phage genome (see [Supplementary Material](#) for the full protocol). We used the Lambda phage shipped in the SQK-LSK108 kit (ONT, UK) as template. In order to compare the ARTIC v4.1 protocol to our custom protocol we followed the ARTIC protocol for cDNA synthesis, amplification, and sequencing using BA.1 (Omicron) and B.1.617.2 (Delta) cell culture variants at 100, 1, and 0.1 viral particles/µL. For the custom protocol, purified PCR amplicons were taken through the R2C2 protocol prior to sequencing. Briefly, to circularize the amplicons a 12 µL reaction composed of 2x NEBuilder HiFi DNA Assembly Master Mix (NEB, USA), PCR amplicons, and Lambda splint was prepared. The PCR amplicons and Lambda splint were mixed in a molar ratio of 1:3 up to a total of 200 ng of PCR amplicons. This reaction was incubated at 55 °C for 60 min. Non-circularized molecules were removed by exonuclease followed by overnight rolling circle amplification using Phi29 polymerase. T7 Endonuclease was used to de-branch the DNA.

The debranched DNA was purified and sequenced as described for the ARTIC protocol. An overview of the work-flow is shown in [Figure 1](#).

Data analysis

Basecalling and read processing

All samples were sequenced and basecalled on PromethION with the following parameters; MiniKNOW version: 21.05.20, flow cell: FLO-PRO002, library kit: SQK-LSK109, basecalling type: High accuracy, barcoding kit: EXP-NBD193, Guppy version: 5.0.17. The resulting reads were processed using pychopper (<https://github.com/nanoporetech/pychopper>) to trim Nanopore sequencing adapters and split chimeric reads containing adapters within them. Processed reads were supplied to the ARTIC pipeline to perform read filtering, primer trimming, amplicon coverage normalisation, variant calling, and consensus building via the Medaka workflow. The specific lineage for each consensus sequence was assigned using the Phylogenetic Assignment of Named Global Outbreak (PANGO) package (Rambaut et al., 2020). To produce and plot statistics from the alignment files, we used the SAMtools (Li et al., 2009) packages ampliconstats and plot-ampliconstats (<http://www.htslib.org/doc/samtools-ampliconstats.html>), respectively.

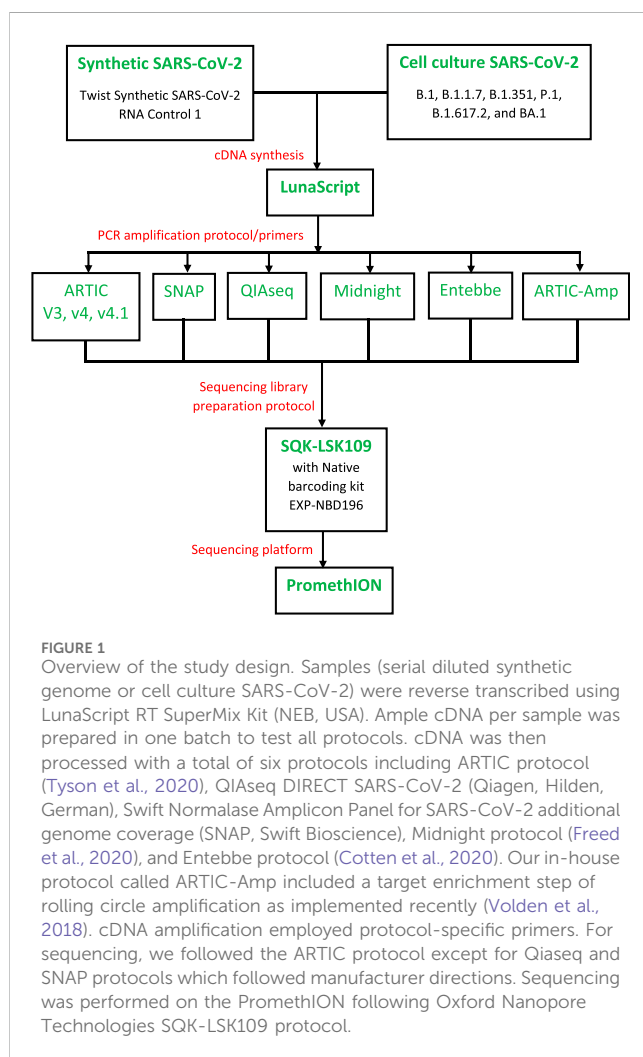
Results

Short amplicon protocols show highest genome completeness on synthetic genome

We obtained the Twist Synthetic SARS-CoV-2 RNA Control 1 whose commercial concentration is 1,000,000 particles per microlitre. We performed log serial dilution to obtain samples at 1×10^6 , 1×10^5 , 1×10^4 , 1×10^3 , and 1×10^2 particles per millilitre. We processed these samples using five different protocols in triplicate: ARTIC version 3, ARTIC v4, Midnight, SNAP, and Entebbe. ARTIC v3, ARTIC v4, and SNAP protocols are short amplicon protocols yielding amplicons of approximately 450, 450, and 350 bp, respectively while Midnight and Entebbe are long amplicon protocols yielding reads of ~1,200 and 1,800 bp, respectively ([Supplementary Figure S1](#)). We observed a higher amount of PCR amplicons using the ARTIC V4 protocol compared to all other protocols. For examples, in the highest viral titre samples with 1×10^6 particles/mL we obtained a mean of 162 ng/µL with ARTIC v4 compared to 137, 107, 70, and 8.4 ng/µL with Entebbe, ARTIC v3, Midnight, and SNAP protocols, respectively ([Supplementary Figure S2](#)). We then sub-sampled 40,000 reads and used the ARTIC pipeline to reconstruct the genome. We assessed genome completeness measured as the percentage of non “N” bases in the reconstructed genome of the whole genome ([Figure 2](#)). The SNAP protocol showed the highest genome completeness at all tested viral particle concentrations. The SNAP protocol showed an average of 96.45% coverage across the genome. This was followed by ARTIC V3, ARTIC V4, Midnight, and Entebbe. The Midnight protocol was only tested at 1×10^5 particles/mL.

TABLE 2 Primers used in the proof-of-principle experiment with our in-house ARTIC-Amp protocol.

No.	Reference	Start	End	Name	Pool	Sequence	Gene
1	MN908947.3	28,512	28,536	SARS-CoV-2_96_LEFT	2	AGATGACCAAAATTGGCTACTACCG	N
	MN908947.3	28,893	28,914	SARS-CoV-2_96_RIGHT	2	CCATTGCCAGCCATTCTAGCA	
2	MN908947.3	27,447	27,473	SARS-CoV-2_92_LEFT	2	CACTACCAAGAGTGTGTAGAGGTAC	ORF7a
	MN908947.3	27,826	27,855	SARS-CoV-2_92_RIGHT	2	GTTCAAGTGAGAACCAAAAGATAATAAGC	
3	MN908947.3	944	966	SARS-CoV-2_4_LEFT	2	GTGTATACTGCTGCCGTGAACA	ORF1ab
	MN908947.3	1,337	1,362	SARS-CoV-2_4_RIGHT	2	ACAACAGCATTTTGGGGTAAGTAAC	
4	MN908947.3	21,865	21,889	SARS-CoV-2_73_LEFT	1	AGAGGCTGGATTTTGGTACTACT	Spike
	MN908947.3	22,247	22,274	SARS-CoV-2_73_RIGHT	1	ACCTAGTGATGTTAATACCTATTGGCA	



ARTIC protocols show highest PCR amplicon yield with cell culture SARS-CoV-2

We obtained cell culture wildtype SARS-CoV-2 (B.1) and five different variants namely, alpha, beta, gamma, delta, and omicron (B.1.1.7, B.1.351, P.1, B.1.617.2, and BA.1, respectively). The concentration of all samples was determined using digital droplet

PCR and then normalized to 1,000 particle/μL and finally log serial diluted to 0.1 particle/μL. Initially, we processed all samples using ARTIC v4 protocol and assessed the yield following PCR amplification (Supplementary Figure S3). With the exception of B.1.617.2 variant, samples showed similar yields with average post PCR amplicon concentrations not statistically different (One-way ANOVA $p = 0.386$ for 100 particles/μL, $p = 0.687$ for 10 particles/μL, $p = 0.06$ for 1 particles/μL, and $p = 0.251$ for 0.1 particles/μL). We then selected the 100, 1, and 0.1 viral particles/μL dilutions to represent samples with high, medium, and low viral loads, respectively. These dilutions were selected since they yielded an average of 115, 40, and 10 ng/μL (Supplementary Figure S3) which is comparable to the first quartile (84 ng/μL), second quartile (31 ng/μL) and third quartile (8.7 ng/μL) yields observed across our clinical samples, and thus, could simulate samples at high, medium, and low viral titres, respectively.

Samples were processed with six different protocols: ARTIC v3, ARTIC v4.1, QIAseq, SNAP, Midnight, and Entebbe. We compared the yield of amplicons following PCR amplification across the different protocols, variants, and virus titres. The ARTIC protocols had the highest yield with version 4.1 having a slightly higher yield than version 3 except for BA.1 variant (Supplementary Figure S4). The SNAP and QIAseq protocols had the least yields (Table 3). We sequenced all samples on the PromethION and subsampled 100, 1,000, 5,000, 10,000, 20,000, 40,000, 80,000, 100,000, and 200,000 reads from each sample and used the ARTIC pipeline to process these reads and reconstruct the genome. From each reconstructed genome, we assessed the percentage genome completeness, ability to call the correct lineage, and primer drop out.

ARTIC protocols show best overall genome completeness with cell culture SARS-CoV-2

We measured genome completeness as the percentage of the genome without “N”s or gaps. Overall, ARTIC protocols performed the best (Figure 3). For example, among the high viral titre samples (100 particles/μL), the ARTIC protocols had the fewest number of N’s followed by Entebbe protocol followed by Midnight protocol. The ARTIC v4.1 protocol required only 40,000 reads to reach maximum genome completeness of 99.58% across all variants except for B.1.351 which required 80,000 reads and only reached a maximum of 99.57% genome completeness. Interestingly, among

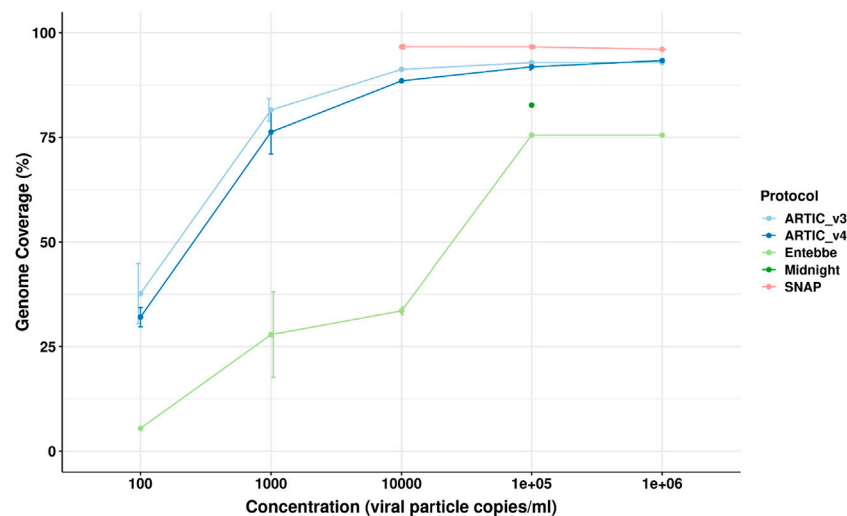


FIGURE 2

Genome completeness assessment using synthetic SARS-CoV-2 genome. The Twist Synthetic SARS-CoV-2 RNA Control 1 (102019, Twist Bioscience, CA, USA) representing wildtype SARS-CoV-2 was used as template at different log-serial diluted concentrations. cDNA used for all protocols was prepared according to the ARTIC protocol while PCR amplification followed protocol specifications. We used ARTIC v4 protocol here instead of v4.1. All data was analysed using ARTIC pipeline. Genome completeness was calculated as the percentage of the fully reconstructed genome (without gaps represented by "N"s). Errors bars represent standard deviation from the mean of three technical replicates.

TABLE 3 Average post-PCR amplicon yield per protocol. The mean PCR amplicon yield (ng) for the wildtype and five SARS-CoV-2 variants tested was computed for each viral titre level.

	Sample viral titre		
	100 particles/ μ L	1 particles/ μ L	0.1 particles/ μ L
ARTIC_v4.1	109	37	14.8
ARTIC_v3	105	31.2	9.35
Entebbe	65.2	9.58	4.2
Midnight	56.7	8.19	1.81
SNAP	6.32	3.21	1.7
Qiaseq	9.1	0.23	0.113

high viral titre samples, the ARTIC v4.1 protocol performed worst with the wildtype virus where the maximum genome completeness reached was 98.75% even with 734,000 reads.

The Entebbe protocol performed best in terms of number of reads needed to reach protocol-specific maximum genome completeness. Among high viral titre samples, only 5,000 reads were required to reach the maximum coverage obtained with this protocol, 98.33%, except for BA.1 which required 40,000 reads to reach this coverage.

The ARTIC v3 reached maximum genome completeness of 99.60% for B.1.1.7 using 20,000 reads. The variants P.1 and B.1.617.2 attained 99.60% genome completeness with 731,000 and 875,000 reads, respectively. ARTIC V3 protocol performed worst with BA.1 variant which only reached 98.66% genome completeness with 655,000 reads.

The Midnight protocol attained 99.36% genome completeness for P.1, B.1.617.2, B.1.351, and B.1.1.7 which required at least 40,000, 10,000, 10,000, and 5,000 reads, respectively. The Midnight protocol performed worst with wildtype B.1 and BA.1 which only reached 96.03% and 96.60% genome completeness with 158,000 and 233,000 reads, respectively.

The SNAP protocol reached 99.66% genome completeness with B.1.1.7, B.1.617.2, and B.1.351 variants all of which required at least 200,000 reads. With 469,000 reads, SNAP reached 99.75% genome completeness with B.1.617.2. This was the highest genome completeness of all protocols assessed. For B.1, P.1, and BA.1, the maximum genome completeness reached was 99.5, 99.28, and 97.70% using 23,000, 180,000, and 200,000 reads respectively.

The Qiaseq protocol reached a maximum of 98.04% genome completeness with the P.1 variant at 614,000 reads. Among the other samples, the maximum genome completeness reached for B.1, BA.1, B.1.1.7, B.1.351, and B.1.617.2 were 97.90, 96.11, 97.76, 96.17, and 97.14% with 0.6, 0.5, 0.8, 0.5, and 1.4 million reads, respectively.

Among medium viral load samples (1 particles/ μ L), we did not have enough reads to compare all protocols. Nevertheless, the ARTIC v4.1 performed best in comparison to ARTIC V3, Midnight, and Entebbe where we had comparable reads of up to 10,000 reads (Supplementary Figure S5). At 10,000 reads, ARTIC V3, Midnight, and Entebbe showed an average of 78.99, 71.58, and 71.28% genome completeness compared to 84.49% genome completeness attained with the ARTIC v4.1 protocol.

Among low viral titre samples (0.1 particles/ μ L), we only had enough reads from ARTIC v3 and ARTIC v4.1 protocols. ARTIC v4.1 showed 0.3% higher genome completeness across all lineages at 100,000 reads compared to ARTIC V3 (74.51% versus 74.15%, respectively, Supplementary Figure S6).

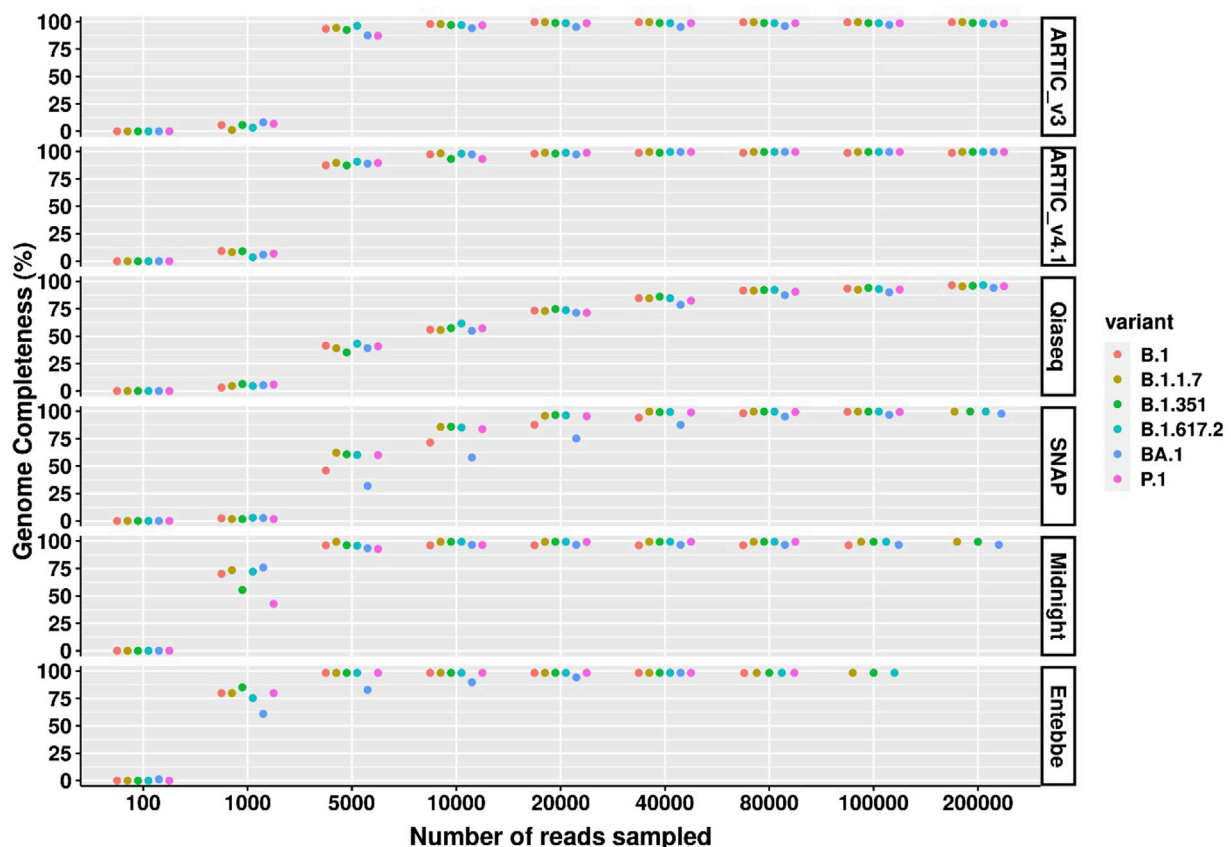


FIGURE 3
SARS-CoV-2 genome completeness comparison across high viral titre samples and protocols. Wildtype SARS-CoV-2 and five cell culture variants were processed for sequencing following six different protocols; ARTIC v3, ARTIC v4.1, Qiaseq, SNAP, Midnight, and Entebbe. Prepared cDNA libraries were sequenced on the PromethION and the data analysed using ARTIC pipeline to reconstruct the genomes using a set of randomly sub-sampled reads. At each set of sampled reads genome completeness was computed as the percentage of the fully reconstructed genome (without gaps represented by "N"s). Samples used here were the normalised and serial dilutions at 100 viral particles/ μ L.

Long amplicon protocols require fewest reads to call correct lineage

We also assessed the number of reads needed to call the correct lineage across all lineages tested. Among high viral load samples, the Entebbe protocol required the fewest number of reads to correctly call all lineages tested, 1,000 reads (Figure 4). This was followed by Midnight protocol which required 5,000 reads to correctly call all lineages. Surprisingly, ARTIC v3 only needed 40,000 reads to correctly call all lineages while the ARTIC v4.1 needed at least 200,000 reads to correctly call BA.1. The SNAP protocol failed to correctly call BA.1 even with 200,000 reads whereas the Qiaseq protocol showed highest sensitivity to B.1.17 and B.1.351 lineages only calling all lineages correctly with 200,000 reads.

Among medium viral load samples, the ARTIC v4.1 performed best, requiring only 5,000 reads to correctly call six lineages tested (Supplementary Figure S7). ARTIC v3 required at least 80,000 reads to call all lineages correctly. At 40,000 reads, the Midnight protocol failed to correctly call B.1 and B.1.617.2 while at 20,000 reads the Entebbe protocol failed to correctly call P.1 and B.1.617.2.

Among low viral load samples, the ARTIC v3 and ARTIC v4.1 could be compared directly at 100,000 reads (Supplementary Figure S8). At this level, ARTIC v4.1 performed better only failing to

return the correct lineage calls for B.1.351 and P.1 compared to ARTIC v3 which only returned correct calls for B.1 and BA.1. The Midnight protocol and Entebbe protocols, both of which are long-read protocols did not yield enough reads to evaluate them properly. Nevertheless, both protocols called BA.1 lineage correctly with only 1,000 reads.

Genome coverage and primer dropout

We assessed genome coverage for the different protocols across different lineages of SARS-CoV-2 after combining all reads to simulate a high viral load sample. Overall, the long amplicon protocols (Entebbe and Midnight) showed the least primer dropout with the Entebbe protocol performing better than Midnight protocol (Supplementary Figure S9; Supplementary Figure S10). The Entebbe protocol is the only protocol that did not have any primer fail to reach a minimum threshold of 20 \times coverage required to call a consensus sequence by the ARTIC pipeline (primer dropout). The Midnight protocol showed sensitivity to BA.1 and the wildtype virus with two primers failing to reach 20 \times coverage. The ARTIC v3 protocol had reduced performance against B.1.351 and BA.1 with one and four

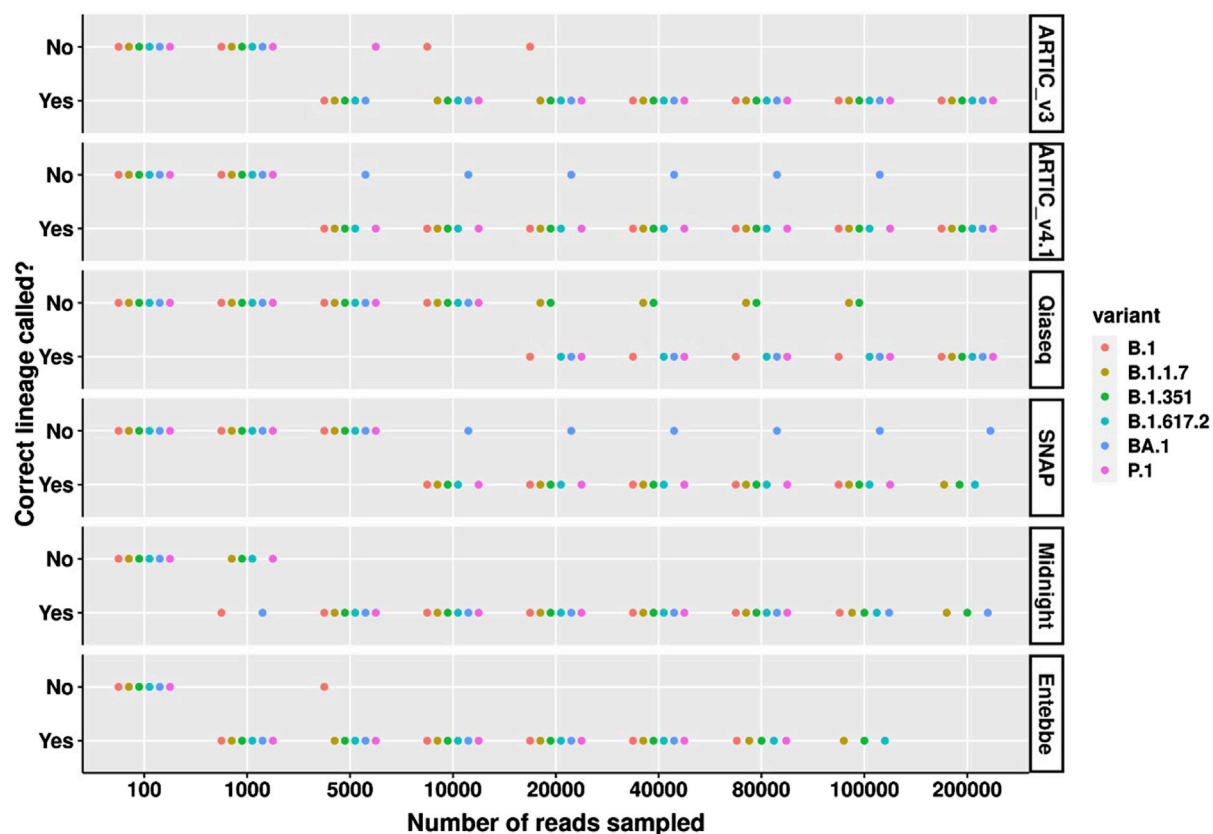


FIGURE 4

Correct SARS-CoV-2 lineage calling comparison across high viral titre variants and protocols. Wildtype SARS-CoV-2 and five cell culture variant samples were processed for sequencing following six different protocols; ARTIC v3, ARTIC v4.1, Qiaseq, SNAP, Midnight, and Entebbe. Prepared cDNA libraries were sequenced on the PromethION and the data analysed using ARTIC pipeline to reconstruct the genomes using a set of randomly subsampled reads. At each set of sampled reads the Pangolin pipeline (O'Toole et al., 2021) was used to assign Pango lineages (Rambaut et al., 2020) to the reconstructed genome. The figure shows whether the reconstructed genome allowed for the correct lineage to be assigned or not. Samples used here were the normalised and serial dilutions at 100 viral particles/ μ L.

primers, respectively failing to reach 20 \times coverage even when others reached 1,000 \times coverage (Supplementary Figure S11). The ARTIC v4.1 protocol only showed one primer failing to reach 20 \times coverage with the wildtype virus. For all other variants, deep sequencing to achieve at least 100 \times average depth remedied poor primer performance (Supplementary Figure S12). SNAP and Qiaseq protocols have overlapping primers in a single tube which makes it difficult to assign reads to a single primer. Overall, these protocols achieved similar depth as the rest of the protocols although many primers showed reduced sensitivity to all viruses used (Supplementary Figure S13; Supplementary Figure S14, respectively). The use of overlapping primer helps to compensate for this reduced sensitivity. For example, the SNAP protocol showed only one region with significant dropout on BA.1 and none were seen for the Qiaseq protocol. We also assessed evenness of genome coverage measured as percentage coefficient of variation across protocol, variants, and subsampled number of reads (Supplementary Figure S15). The ARTIC v4.1 protocol showed the lowest CoV averaging \sim 70% across all variants and subsampled reads, performing worst with BA.1 which averaged \sim 85%. This was followed by ARTIC v3 protocol which averaged 72% across all variants and subsampled reads. The long amplicon

protocols Midnight and Entebbe performed similarly with overall CoV average of 85% while the commercial protocols SNAP and Qiaseq performed worst with overall CoV of 86% and 103%, respectively.

Our custom protocol drastically improves over ARTIC protocol

Given the poor performance of all protocols on samples with medium and low viral load titres, we sought a method that would increase sensitivity of SARS-CoV-2 detection. We tweaked a recently published method called Rolling Circle Amplification to Concatemeric Consensus (R2C2) (Volden et al., 2018) and termed it ARTIC-Amp (for amplified ARTIC) as it leverages the ARTIC v4.1 protocol and complements it with rolling circle amplification. In a “proof of principle” experiment, we selected four primer-sets from the ARTIC v4.1 primers that target four genes namely, N, ORF7a, ORF1a, and Spike, respectively and added extra tag sequences to allow for circularisation via Gibson assembly. To this effect, we also amplified a 330 bp “splint” with Lambda virus DNA as the template using primers carrying the same tag sequences



FIGURE 5

Comparison of genome coverage between ARTIC v4.1 and ARTIC-Amp protocols. Four primer sets targeting regions in four SARS-CoV-2 genes: N, ORF7a, ORF1a, and Spike, respectively were used to prepare sequencing libraries either following the ARTIC v4.1 protocol or our in-house protocol termed ARTIC-Amp. The ARTIC-Amp protocol takes the final products of the ARTIC v4.1 protocol and circularises them via Gibson assembly followed by isothermal rolling circle amplification as described previously (Volden et al., 2018). Reads generated by the ARTIC-Amp protocol were processed using R2C2 pipeline (Volden et al., 2018). All reads were aligned to the genome and loaded on IGV (Robinson et al., 2011) for viewing. A screenshot is shown here. Red arrows indicate regions missed by ARTIC v4.1 protocol. No regions were missed by ARTIC-Amp protocol in this proof-of-principle experiment. Samples used here were the B.1.617.2 low viral titre (0.1 viral particles/ μ L) done in triplicate.

(Supplementary Figure S16A). Next, we optimised primer annealing and extension temperatures and found that unlike ARTIC v4.1 primers that had the highest yield at 64.3°C, the ARTIC-Amp modified primers had highest yield at 65.6°C (Supplementary Figure S16B). Using this optimised annealing temperature, we completed the ARTIC-Amp protocol and generated products with peak length of 8.9 kb (Supplementary Figure S16C). We compared the ARTIC v4.1 PCR amplicon yields and ARTIC-Amp final yields following rolling circle amplification (RCA) (Supplementary Figure S17). We noted that whereas the ARTIC v4.1 yields dropped with reducing viral titres, ARTIC-Amp yields remained relatively stable across viral titres except for BA.1 (BA.1_100) sample (Supplementary Figure S17). On average, the ARTIC-Amp protocol resulted in a 40 \times amplification of yields among the low viral titre samples. For example, for both B.1.617.2 and BA.1 low viral titre samples that were done in triplicate, the average amplicon yields increased from 1.0 and 1.3 to 40 and 44 ng/ μ L, for ARTIC v4.1 and ARTIC-Amp protocols, respectively (Supplementary Figure S18). Sequencing of these products on the PromethION resulted in products with median length of 805 bp (Supplementary Figure S19). The number of reads generated with the ARTIC-Amp was evenly distributed across samples compared to the reads generated from ARTIC v4.1 which corresponded to viral titre (Supplementary Figure S20).

We then assessed the number of times each original amplicon was amplified by rolling circle amplification of circularised product. The number of repeats identified in each sequence had a median of 1 and a mean of 2. The distribution of the number of repeats in each read is shown in Supplementary Figure S21. We looked at 12,394 sequences for which we could get a consensus following error correction. The alignment identity of these sequences

increased by 6%, from 91% without consensus correction to 97% following R2C2-mediated consensus correction (Supplementary Figure S23; Supplementary Figure S22). Finally, we compared genome coverage between ARTIC v4.1 and ARTIC-Amp in the B.1.617.2 low viral titre samples. We noted that whereas two targeted regions were missed in two out of the three low viral titre replicates for the ARTIC v4.1 protocol no region was missed in any of the three low viral titre replicates for the ARTIC-Amp protocol (Figure 5).

Discussion

The COVID-19 pandemic started in late 2019 and continues to rage throughout the world causing losses both in life and labour shortages due to disease burden. Several methods have emerged for high throughput whole genome sequencing of SARS-CoV-2 in clinical and wastewater samples. These methods can be categorised based on the target enrichment method used. The two most commonly employed target enrichment methods are target hybrid capture and PCR (also referred to as amplicon-based methods). In one study, hybrid capture-based methods were found to have higher accuracy of identifying “within-sample variation” compared to amplicon-based methods while amplicon-based methods had higher accuracy of identifying “between-sample variations” (Xiao et al., 2020). Perhaps the most widely applied amplicon-based method is one developed by the ARTIC network (<https://artic.network/ncov-2019>). The primers used in this method have undergone at least three changes to cope with SARS-CoV-2 evolution. In the current work we evaluated version 3 and 4.1 of the ARTIC primers. Further, we evaluated the Swift Normalase Amplicon Panel for SARS-CoV-2 (SNAP, Swift Bioscience), QIAseq DIRECT SARS-CoV-2 (QIAseq, Qiagen), Midnight, and

Entebbe protocols on post-PCR amplicon yield, genome completeness, correct lineage calling, and primer dropout.

Initially, we used a synthetic genome diluted at different viral copy concentrations. The synthetic genome was composed of 5 kb non-overlapping fragments. The non-overlapping fragments, however, create a unique situation since SARS-CoV-2 positive samples would be expected to have either full genome and/or randomly sheared and overlapping fragments. We evaluated post-PCR concentration of amplicons because in our experience this generally correlates positively with genome completeness. We observed the highest amount of PCR amplicons using the ARTIC v4 protocol compared to all other protocols suggesting a higher PCR efficiency with this protocol. We utilized the ARTIC v4 primers for this experiment since the version 4.1, updated to address mutations in BA.1, was not yet available and would not be envisaged to improve yields of this wildtype synthetic virus. The SNAP protocol yielded the lowest concentrations of all protocols. Although this could probably be due to the lower number of PCR cycles recommended for this protocol compared to other protocols (28 versus 36 cycles, respectively), increasing PCR cycles to 36 in a later experiment did not increase yields. Nevertheless, the SNAP protocol showed the highest genome completeness at 1×10^4 , 1×10^5 , and 1×10^6 copies/mL where we had enough reads to compare across protocols. This is probably due to two factors. 1) The SNAP protocol has 344 primers which is by far the highest compared to 98 for ARTIC v3, 99 for ARTIC v4, 20 for Entebbe, and 29 for Midnight. 2) The synthetic genome having been composed of non-overlapping segments. The high number of primers in the SNAP protocol probably had better ability to cover the ends of the fragments than the other protocols could thus resulting in better genome completeness. The SNAP protocol, therefore, should perform well with highly degraded samples. However, the high number of primers also means higher sensitivity of the primers to mutations (Borcard et al., 2022; Kuchinski et al., 2022; Rosenthal et al., 2022). In fact, the SNAP primers underwent at least one round of changes in response to viral evolution. The ARTIC protocols were the second in achieving highest genome completeness with insignificant differences between version 3 and 4. The two long amplicon protocols, Entebbe and Midnight performed worst probably due to the lower number of primers. We only had enough reads to evaluate the Midnight protocol at 1×10^5 copies/mL where it performed better than Entebbe protocol.

Next, we used cell culture wildtype and variant SARS-CoV-2. The samples were quantified using digital droplet PCR (ddPCR) and then normalized in an attempt to have a comparable number of viral particles in each sample. Samples were then serial-diluted and then processed with ARTIC v4 protocol to determine the amplicon concentration. We have found that post-PCR amplicon concentration is indicative of genome completeness such that samples with 50 ng/ μ L and above have a positive correlation with genome completeness and yield the highest coverage. We then used three dilutions of cell culture SARS-CoV-2; 100 particles/ μ L to reflect samples with high viral titre, 1 particle/ μ L to reflect samples with medium viral titres, and 0.1 particle/ μ L to reflect samples with low viral titres. When processed with ARTIC v4.1 in triplicate, our normalisation was successful with no significant differences between yields for the wildtype and the four variants tested at 100 viral particles/ μ L and 0.1 particles/ μ L.

The B.1.617.2 variant did not seem to be as well normalised as the other samples which caused the samples at 1 particle/ μ L to be significantly different. The normalization however, worked well given that the post-PCR amplicon yield was comparable between the four ddPCR quantified samples and the Twist biosciences synthetic virus suggesting that they contained similar viral particles (Supplementary Figure S25). It is noteworthy that PCR amplicon yield has a limitation namely; PCR efficiency can be different when one set of primers is used on different lineages.

The main goal of whole genome SARS-CoV-2 sequencing is to obtain the whole sequence of the genome which provides the entirety of mutations acquired by the virus. Therefore, genome completeness, as determined by the percentage of the genome that is assembled without gaps represented as “N” is critical (Chiara et al., 2021). Genome completeness is positively correlated with sample viral titre such that samples with low viral titre have reduced genome completeness. We thus, evaluated genome completeness in samples with high (100 particles/ μ L), medium (1 particles/ μ L), and low viral titres (0.1 particles/ μ L). Among samples with high viral titres, the SNAP protocol achieved the highest genome completeness of 99.75% with 469,000 reads except for B.1.617.2 variant. The SNAP protocol performed worst with BA.1 variant. Noteworthy, we evaluated an earlier version of the SNAP protocol which did not include revisions to address virus evolution. The SNAP protocol thus, performed best with genome completeness but required very high sequencing depth. Choi et al. (2022) reported genome completeness of 99.78% with SNAP although no lineage information was reported. They further used ddPCR and reported that SNAP required a minimum of 10.5 copies/ μ L to achieve >95% genome completeness.

Among all protocols tested, it was the ARTIC v4.1 protocol that achieved highest coverage with the lowest number of reads. Further, the ARTIC v4.1 protocol performed relatively well across variants. The ARTIC protocol showed highest coverage across high and low viral titre samples in other studies (Charre et al., 2020; Nasir et al., 2020). The Qiaseq protocol achieved modest genome completeness attaining a maximum of only 98.04% which was the lowest among the protocols we tested. Noteworthy, we prepared the cDNA as described for ARTIC protocol and only followed the Qiaseq protocol for PCR amplification. Further, the Qiaseq and SNAP protocols were developed primarily with Illumina sequencing in mind although we did not see any technical barrier to sequencing libraries generated by these protocols with the PromethION.

Given high viral load titres, the long amplicon protocols showed fewest primer dropout suggesting that Entebbe and Midnight protocols might be less sensitive to changes in viral sequence. This is likely due to the lower number of primers employed in these protocols (20 and 29 primers, respectively). We find that the Entebbe primers are more suitable than Midnight primers in this case. However, we only tested the primers yielding 1,200 bp amplicons from the Midnight protocol and did not test their other counterparts which yield 1,500 and 2,000 bp amplicons (Freed et al., 2020). It is possible that these longer amplicon Midnight primers perform comparably to the Entebbe primers which yield 1,800 bp amplicons. Although we did not specifically assess genome coverage at medium and low viral loads, all evidence points towards the ARTIC protocol performing better than other protocols. In this work, the ARTIC protocols generated enough

libraries even from medium and low viral load samples. Further, Liu et al., reported highest genome coverage from ARTIC-based protocol among seven different protocols they tested (Liu et al., 2021). We further found that the ARTIC v4.1 protocol had the best evenness of genome coverage measured as coefficient of variation at 70%. This was also reported by Liu et al. who found that ARTIC-based protocol performed best out of the seven they tested on uniformity of genome coverage (Liu et al., 2021). Liu et al., used 1 million Illumina reads and found the ARTIC-based protocol to have a coefficient of variation of 100%. Using Nanopore reads, we find that ARTIC protocols have an average 70% which might be related to the length of reads. Therefore, overall, we find that the ARTIC v4.1 protocol outperforms other protocols in terms of low primer dropout and evenness of genome coverage.

Because of the relatively poor results obtained with medium and low viral titre samples across all protocols we sought to design a method that would improve sensitivity of amplification of SARS-CoV-2 and thus improve genome completeness and correct lineage calling among medium and low viral titre samples. We adapted a previously employed method called R2C2 (Volden et al., 2018). The authors had used this method to improve accuracy of the relatively high error-prone Nanopore reads. However, we envisaged that the isothermal multiple displacement amplification (MDA)-mediated rolling circle amplification (RCA) used in the R2C2 method would also work to increase abundance of amplicons. We called this method ARTIC-Amp (for amplified ARTIC) since it leverages the simplicity and robustness of the ARTIC protocol. Thus, samples were processed initially according to the ARTIC v4.1 protocol and then taken through another round of RCA overnight. In our proof-of-principle experiment that included only four primer pairs targeting four different SARS-CoV-2 genes, we noticed that indeed, there was an improvement of 40× in the amount of amplicons following overnight RCA. Further, this reaction seemed to proceed to its endpoint and thus resulted in similar amounts of material in each sample and thus normalised the amplicons eliminating the need to perform further normalisation for multiplexed samples. This normalisation further resulted in comparable number of reads once the samples were barcoded, multiplexed and sequenced on the same flow cell. We however, noticed that although we saw a peak of 8.9 kb in the RCA products, the sequenced reads had a peak at ~1.5 kb. Consequently, the circularization efficiency was very low with a median of one. It is likely that more stringent size selection will be beneficial to increase the read lengths. We also need to devise strategies for improving circularisation efficiency. This would help to increase consensus accuracy of reads. We achieved a 6% increase in read alignment rate, but this was for a small portion of reads that had enough repeats to create a consensus. Our main goal was to use the R2C2 method to increase sensitivity of detection. In our proof-of-principle experiment, this was achieved. As shown in Figure 5, we were able to recover all four target regions in the BA.1 low viral titre triplicate samples processed with ARTIC-Amp protocol yet two of these regions were missed in three of the same replicates processed with just ARTIC v4.1 protocol. Logistical and time reasons have hindered full exploration of this protocol but remains part of our ongoing work in the lab. We however,

strongly believe that this approach can dramatically improve sensitivity of detection and be very useful for samples that routinely have low viral titres such as wastewater samples. Further, this protocol only adds USD 21 per sample over the cost of generating ARTIC v4.1 amplicons. The calculated cost per sample of running our ARTIC v4.1 protocol is USD 46\$ and therefore, the additional \$21 of the ARTIC-Amp protocol would be 46% which we think is reasonable (see [Supplementary Material](#) for reagent costs). Further, since amplicon-based methods have been found to outperform capture-based methods particularly in challenging samples (Xiao et al., 2020), we believe that our approach which leverages the robust amplicon-based ARTIC protocol and further complements it with isothermal amplification should provide superior sensitivity.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

AB: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review and editing. SR: Conceptualization, Methodology, Writing – original draft, Writing – review and editing. JLL: Writing – review and editing, Data curation, Project administration. ID: Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing. AG: Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing. LF: Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing. JR: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Coronavirus Variants Rapid Response Network (CoVaRR-Net). CoVaRR-Net is funded by an operating grant from the Canadian Institutes of Health Research (CIHR) – Instituts de recherche en santé du Canada (FRN# 175622). JR is funded by a Genome Canada Genomics Technology Platform grant, the Canada Foundation for Innovation (CFI) and the CFI Leaders Opportunity Fund (32,557), Compute Canada Resource Allocation Project (WST-164-AB) and Genome Innovation Node (244819).

Acknowledgments

We would like to thank Qiagen and Swift Biosciences for free test kits they provided some of which were used in generating data for

the current work. We thank members of the McGill Genome Centre CovSeq project for their help with sequencing of samples.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

References

- Ahmed, W., Tschärke, B., Bertsch, P. M., Bibby, K., Bivins, A., Choi, P., et al. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: a temporal case study. *Sci. Total Environ.* 761, 144216. doi:10.1016/j.scitotenv.2020.144216
- Borcard, L., Gempeler, S., Terrazos Miani, M. A., Baumann, C., Grädel, C., Dijkman, R., et al. (2022). Investigating the extent of primer dropout in SARS-CoV-2 genome sequences during the early circulation of delta variants. *Front. Virology* 2. doi:10.3389/fviro.2022.840952
- Charre, C., Ginevra, C., Sabatier, M., Regue, H., Destras, G., Brun, S., et al. (2020). Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol.* 6 (2), veaa075. doi:10.1093/ve/veaa075
- Chiara, M., D'Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., et al. (2021). Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief. Bioinform.* 22 (2), 616–630. doi:10.1093/bib/bbaa297
- Choi, H., Hwang, M., Navarathna, D. H., Xu, J., Lukey, J., and Jinadatha, C. (2022). Performance of COVIDSeq and swift normalise amplicon SARS-CoV-2 panels for SARS-CoV-2 genome sequencing: practical guide and combining FASTQ strategy. *J. Clin. Microbiol.* 60 (4), e0002522. doi:10.1128/jcm.00025-22
- Cotten, M., Bugembe, D. L., Kaleebu, P., and Phan, M. V. T. (2020). Alternate primers for whole-genome SARS-CoV-2 sequencing. *bioRxiv* 2020.2010.2012, 335513. doi:10.1101/2020.10.12.335513
- Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., et al. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372 (6538), eabg3055. doi:10.1126/science.abg3055
- Domingo, E., García-Crespo, C., Lobo-Vega, R., and Perales, C. (2021). Mutation rates, mutation frequencies, and proofreading-repair activities in RNA virus genetics. *Viruses* 13 (9), 1882. doi:10.3390/v13091882
- Freed, N. E., Vlková, M., Faisal, M. B., and Silander, O. K. (2020). Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford nanopore rapid barcoding. *Biol. Methods Protoc.* 5 (1), bpaa014. doi:10.1093/biomethods/bpaa014
- Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., et al. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19 (7), 409–424. doi:10.1038/s41579-021-00573-0
- Jafferali, M. H., Khatami, K., Atasoy, M., Birgersson, M., Williams, C., and Cetecioglu, Z. (2021). Benchmarking virus concentration methods for quantification of SARS-CoV-2 in raw wastewater. *Sci. Total Environ.* 755 (Pt 1), 142939. doi:10.1016/j.scitotenv.2020.142939
- Kuchinski, K. S., Nguyen, J., Lee, T. D., Hickman, R., Jassem, A. N., Hoang, L. M. N., et al. (2022). Mutations in emerging variant of concern lineages disrupt genomic sequencing of SARS-CoV-2 clinical specimens. *Int. J. Infect. Dis.* 114, 51–54. doi:10.1016/j.ijid.2021.10.050
- Larsen, D. A., and Wigginton, K. R. (2020). Tracking COVID-19 with wastewater. *Nat. Biotechnol.* 38 (10), 1151–1153. doi:10.1038/s41587-020-0690-1
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Liu, T., Chen, Z., Chen, W., Chen, X., Hosseini, M., Yang, Z., et al. (2021). A benchmarking study of SARS-CoV-2 whole-genome sequencing protocols using COVID-19 patient samples. *iScience* 24 (8), 102892. doi:10.1016/j.isci.2021.102892
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395 (10224), 565–574. doi:10.1016/s0140-6736(20)30251-8
- Mendiola-Pastrana, I. R., López-Ortiz, E., Río de la Loza-Zamora, J. G., González, J., Gómez-García, A., and López-Ortiz, G. (2022). SARS-CoV-2 variants and clinical outcomes: a systematic review. *Life (Basel)* 12 (2), 170. doi:10.3390/life12020170
- Nasir, J. A., Kozak, R. A., Aftanas, P., Raphenya, A. R., Smith, K. M., Maguire, F., et al. (2020). A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses* 12 (8), 895. doi:10.3390/v12080895
- O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., et al. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* 7 (2), veab064. doi:10.1093/ve/veab064
- Plitnick, J., Griesemer, S., Lasek-Nesselquist, E., Singh, N., Lamson, D. M., and St George, K. (2021). Whole-genome sequencing of SARS-CoV-2: assessment of the ion torrent AmpliSeq panel and comparison with the illumina MiSeq ARTIC protocol. *J. Clin. Microbiol.* 59 (12), e0064921. doi:10.1128/jcm.00649-21
- Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5 (11), 1403–1407. doi:10.1038/s41564-020-0770-5
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29 (1), 24–26. doi:10.1038/nbt.1754
- Rosenthal, S. H., Gerasimova, A., Ruiz-Vega, R., Livingston, K., Kagan, R. M., Liu, Y., et al. (2022). Development and validation of a high throughput SARS-CoV-2 whole genome sequencing workflow in a clinical laboratory. *Sci. Rep.* 12 (1), 2054. doi:10.1038/s41598-022-06091-0
- Smyth, D. S., Trujillo, M., Gregory, D. A., Cheung, K., Gao, A., Graham, M., et al. (2022). Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat. Commun.* 13 (1), 635. doi:10.1038/s41467-022-28246-3
- Thorne, L. G., Bouhaddou, M., Reuschl, A. K., Zuliani-Alvarez, L., Polacco, B., Pelin, A., et al. (2022). Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* 602 (7897), 487–495. doi:10.1038/s41586-021-04352-y
- Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., et al. (2020). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*, 2020.2009.2004.283077. doi:10.1101/2020.09.04.283077
- Volden, R., Palmer, T., Byrne, A., Cole, C., Schmitz, R. J., Green, R. E., et al. (2018). Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc. Natl. Acad. Sci. U. S. A.* 115 (39), 9726–9731. doi:10.1073/pnas.1806447115
- Walker, A. S., Vihta, K. D., Gethings, O., Pritchard, E., Jones, J., House, T., et al. (2021). Tracking the emergence of SARS-CoV-2 alpha variant in the United Kingdom. *N. Engl. J. Med.* 385 (27), 2582–2585. doi:10.1056/NEJMc2103227
- WHO (World Health Organization) (2024). *WHO coronavirus disease (COVID-19) dashboard*. Available online at: <https://covid19.who.int/> (Accessed October 10, 2024).
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269. doi:10.1038/s41586-020-2008-3
- Xiao, M., Liu, X., Ji, J., Li, M., Li, J., Yang, L., et al. (2020). Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* 12 (1), 57. doi:10.1186/s13073-020-00751-4
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382 (8), 727–733. doi:10.1056/NEJMoa2001017

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1516791/full#supplementary-material>