SARS-CoV-2 GENOME SEQUENCING METHODS DIFFER IN THEIR ABILITY TO DETECT VARIANTS FROM LOW VIRAL LOAD SAMPLES

C. Lam[1], K. Gray[1,2], M. Gall[2], R. Sadsad[1,3], A. Arnott[2], J. Johnson-Mackinnon[1,3], W. Fong[1], K. Basile[2,3], J. Kok[1,2], D. E. Dwyer[1,2,3] V. Sintchenko[1,2,3], R.J. Rockett[1,3]

[1]Centre for Infectious Diseases and Microbiology-Public Health, Westmead Hospital, Westmead, New South Wales, AUSTRALIA

[2]Centre for Infectious Diseases and Microbiology Laboratory Services, NSW Health Pathology, Institute for Clinical Pathology and Medical Research, Westmead, New South Wales, AUSTRALIA

[3]The University of Sydney, Sydney, New South Wales, AUSTRALIA

ABSTRACT (149 words)

SARS-CoV-2 genomic surveillance has been vital in understanding the spread of COVID-19, the emergence of viral escape mutants and variants of concern. However, low viral loads in clinical specimens affect variant calling for phylogenetic analyses and detection of low frequency variants, important in uncovering infection transmission chains. We systematically evaluated three widely adopted SARS-CoV-2 whole genome sequencing methods for their sensitivity, specificity, and ability to reliably detect low frequency variants. Our analyses highlight that the ARTIC v3 protocol consistently displays high sensitivity for generating complete genomes at low viral loads compared with the probe-based Illumina respiratory viral oligo panel, and a pooled long-amplicon method. We show substantial variability in the number and location of low-frequency variants detected using the three methods, highlighting the importance of selecting appropriate methods to obtain high quality sequence data from low viral load samples for public health and genomic surveillance purposes.

INTRODUCTION (4741 words, including methods)

The rapid implementation of genomic epidemiology has enabled unparalleled understanding and monitoring of viral evolution during the SARS-CoV-2 pandemic. The first reported SARS-CoV-2 case in Australia was notified on January 25, 2020, and by the end of 2020, 28,381 SARS-CoV-2 cases had been identified nationwide (https://www.health.gov.au/resources/publications/coronavirus-covid-19-at-a-glance-30-december-2020-0). Australia's low prevalence of COVID-19 is due to the implementation of

strong public health measures, which in New South Wales (NSW) has included integrated genomic surveillance to inform public health responses and contact tracing efforts [1].

Whole genome sequencing (WGS) of SARS-CoV-2 was implemented in NSW, Australia within two weeks of the first reported case in anticipation of increasing SARS-CoV-2 infections [2]. A pooled long-amplicon (long-amp) based sequencing approach was initially selected based on reagent and resource availability and was quickly adapted to fit existing WGS workflows and infrastructure [3]. By March 28, 2020, 209 samples from NSW had been sequenced and released on the Global Initiative on Sharing All Influenza Data database (GISAID; www.gisaid.org) [3], representing 13% of all SARS-CoV-2 cases diagnosed in NSW at the time. The initiative to promptly release genomic data has mirrored other national and international efforts focused on near real-time monitoring of the evolution and intercontinental spread of the SARS-CoV-2 [4-6]. Prospective WGS of SARS-CoV-2 cases in NSW has continued and to date (April 30, 2021), 1710 genomes representing 32% of confirmed cases have been generated.

A large array of SARS-CoV-2 sequencing protocols have been developed since the start of the pandemic. They range from numerous target enrichment techniques which can be applied to multiple sequencing technologies, to suites of bioinformatics and data visualisation workflows [7]. The rapid development of all aspects of SARS-CoV-2 WGS was aided in part by efforts from the global genomics community in developing viral WGS methods (https://artic.network/ncov-2019). However, accurate SARS-CoV-2 genomic surveillance has been hampered by several common challenges: firstly, a high level of variability exists between sequencing protocols in obtaining complete SARS-CoV-2 genomes, particularly from clinical samples with low viral loads (as reflected by real-time PCR (RT-PCR) cycle threshold (Ct) values), such as those collected from patients without symptoms, mild disease, or late in the course of infection. Secondly, the accuracy required to detect and call variants using different protocols has not been adequately validated. All of these factors, sequencing method, reproducibility and thresholds for variant calling may affect the quality and impact of genomic surveillance and ultimately public health efforts to contain outbreaks.

Synthesis of SARS-CoV-2 genomic data with detailed epidemiological exposure and contact tracing information can provide definitive evidence of importation events and identification of local SARS-CoV-2 transmission chains [3, 8]. SARS-CoV-2 clusters, transmission chains, or networks linked to superspreading events are often differentiated

66 genomically by single nucleotide polymorphisms (SNPs) within the SARS-CoV-2 genome [9].
67 The ability to rapidly and accurately characterise SNPs and other variants has become even
68 more important after the identification of several 'variants of concern' (VOC). VOC contain
69 specific mutations identified as important and relevant for COVID-19 control due to mounting
70 evidence of positive selection of specific non-synonymous spike protein mutations that can
71 increase the duration, severity and transmission of COVID-19 by affecting host immune
72 responses [10-13]. Complete genomes generated using highly sensitive and specific
73 sequencing methods are therefore required to inform and enable genomics guided surveillance
74 to provide the information necessary for COVID-19 control and policy decisions, particularly
75 as widespread SARS-CoV-2 vaccination is underway [14-17].

76 This study systematically evaluated three different sequencing methods for their
77 sensitivity and ability to generate complete SARS-CoV-2 genome sequences suitable for public
78 health surveillance. We assessed and compared (i) the pooled long-amplicon method [2] with
79 (ii) ARTIC v3 network tiled amplicon protocol (https://artic.network/ncov-2019) that has been
80 adopted widely since the start of the pandemic, and (iii) a probe capture-based Respiratory
81 Viral Oligo Panel (RVOP) (Illumina). Additionally, we investigated the pattern of low
82 frequency variants generated by these methods, which can be important in defining and
83 highlighting transmission chains [18].
84
85 **RESULTS**
86 ***Viral isolates, viral loads and genome profiles.*** Seven SARS-CoV-2 positive clinical
87 specimens were cultured as representatives of different SARS-CoV-2 genomic clusters that
88 were co-circulating in NSW between February - April 2020 [3]. Details of the genome obtained
89 from each clinical specimen, including GISAID ID, lineage and SNP profile are listed in
90 supplementary Table 1. Two of seven isolates lost a SNP compared to the genome obtained
91 directly from the original clinical specimens. The genome of Isolate 2 reverted to wild-type at
92 position C:26213; however, the SNP C:26213:T detected in the original clinical specimen was
93 still present as a low frequency variant. In Isolate 7, all reads at position 13730 were the wild-
94 type allele (C). To investigate the effect of low viral load on detection of variants, serial
95 dilutions of cell culture supernatant were performed. RT-PCR results from each culture
96 dilution demonstrated that a 10-fold decrease in viral load corresponded to a Ct increase of ~3-
97 4 cycles (Fig.1). A total of seven dilutions were made, five of which remained consistently
98 SARS-CoV-2 RT-PCR positive with corresponding viral loads decreasing from a median load

of 71,062 copies/µL (median Ct 25.42, range Ct 24.29 – 26.65, viral load range 47,482 – 1,178,540 copies/µL) to a median of 112 copies/µL (median Ct 36.62, range 34.7 – 38.19, viral load range 18 – 1,584 copies/µL). Culture dilutions with Ct >39 were deemed too low to attempt sequencing and were excluded from further analysis.

Synthetic control. Using the long-amp method, only 57% (8/14 amplicons) of the synthetic control genome were able to be sequenced up to ~Ct 32, after which no amplicons were produced. Regions which did not amplify at higher viral loads were A2, A3, A4, B4, B5 and B6, signalling that that these primer pairs span two contiguous but separate segments of the synthetic genome. The smaller tiled amplicons from ARTIC v3 produced a higher proportion (93.9%, 92/98 amplicons) of the genome; however, amplicons 16, 17, 33, 50, 66 and 82 did not amplify. Missing regions from both ARTIC v3 and long-amplification methods overlapped, confirming six distinct segments of the synthetic control. Due to the non-amplification of larger products from the long-amplification method, less of the genome was able to be recovered, meaning that subsequent variant calling from these missing regions could not be performed. Complete genomes (>99% coverage) for the synthetic control was able to be obtained using RVOP up to Ct 28.

Comparison of genome coverage across three sequencing methods. At Cts 25-29 (up to 2000 copies/µL), all three WGS methods generated near complete SARS-CoV-2 genomes with >10x coverage (Figure 2, Supplementary Figure 1). The highest level of genome coverage across all five dilutions was achieved using ARTIC v3 with >90% genome coverage achieved at viral loads down to ~Ct 38 (2 copies/µL). For each of the complete genomes (expected genome size of 29,903bp), there were fewer than 1000 ambiguous bases (N's) from the MN90894.3 reference genome (Figure 2). On the other hand, genome coverage decreased substantially using long-amplicon and RVOP methods at a median Ct 32 (Ct range 30.7 – 33.4, median viral load 1340 copies/µL, range 725 – 14,613 copies/µL) (Supplementary Figure 1) however the differences observed were not significant (Figure 2). This decreasing trend continued at lower dilutions for both long-amp and RVOP, resulting in significant differences of the genome coverage obtained using the ARTIC, long-amplification and RVOP methods ($p<0.05$) (Figure 2).

Read depth affects genome coverage and variant calling. Read depth across amplicons differed substantially between ARTIC v3 and long-amp methods creating highly uneven genome coverage. ARTIC v3 amplicons 9, 17, 23, 64, 67, 70, 74 and 91 were amplified inconsistently at higher Ct values (Ct >34). A2, B3 and B6 from the long-amplification protocol were the poorest performing, often not amplified in samples with Ct <30. These 400 bp-5 kb

missing amplicons created large genomic gaps, which made variant calling problematic. In contrast, the amplicons which amplified with high efficiency using ARTIC v3 (44, 57, 62), had consistently higher average read depths regardless of Ct value. The RVOP achieved the most consistent read depth across the genome, with relatively even distribution of missing bases compared with either amplification-based sequencing method. However, average read depth of samples at ~Ct 32 (Ct range 30.7 – 33.4) was low (Figure 3) with inconsistent genome coverage <10x, also resulting in problems with variant calling.

ARTIC rebalanced pools*:* Using the COVID-19 Genomics Consortium (COG-UK) guidelines, we rebalanced ARTIC v3 primers in an attempt to improve amplification of specific amplicons and obtain more even sequencing coverage across the genome. Figure 4 shows the performance of rebalanced primers compared with original primer concentrations prior to rebalancing. Unsurprisingly, as viral load decreases, coverage across poorer performing amplicons decreased in parallel (Figure 4 and Supplemental Figure 1). No significant changes in coverage were observed (across all dilutions) with amplicons 15, 27, 73, even though the primer concentrations were increased by 1.5x-2.1x. However, amplicons 64, 67, 70 and 74 (for which primer concentrations were increased by a factor of between 6-7.8x), performed significantly better than original unbalanced primer pools. Other amplicons (i.e. 36, 54 and 66) whose primers were increased by a factor of >3x performed worse than expected (likely due to potentially poorly designed primers or excessive interactions with other primers). Regardless of individual primer rebalancing factors, sufficient depth (>10x) to meet variant calling QC at Ct 35 was obtained for all amplicons.

Comparative sensitivity of three SARS-CoV2 sequencing methods. Sensitivity of each method was defined as the ability to accurately call SNPs, based on a clear consensus amongst all the dilutions. All three methods exceeded 90% sensitivity with median Ct 28.7 (Ct range 27.6 - 31.3, median viral load 12,025 copies/µL) (Figure 5). The sensitivity for ARTIC remained high for samples up to Cts >38, whereas sensitivities for both pooled long-amplicon and RVOP dropped below 80% at Ct >30. No differences in sensitivity or specificity were observed between ARTIC original primer pooling compared with rebalanced primer pools. Interestingly, at higher Ct values, the RVOP was more likely to represent true SNPs as low frequency variants (discussed in further detail below).

RVOP and the detection of other respiratory pathogens. The RVOP can detect 43 common human respiratory viruses and 60 human control genes (which serve as internal positive controls for library construction and sequencing steps) in individual clinical samples [19]. Trimmed reads from all seven diluted cultures prepared using the RVOP were mapped against

203 reference sequences of 43 respiratory pathogens. Human rhinovirus 89 (NC_001617.1) and adenovirus C (NC_001490.1) were detected, although coverage across both viral genomes was less than 2%. An in-house respiratory panel RT-PCR reconfirmed the presence of both rhinovirus (Ct 27) and adenovirus (Ct 26) in the negative respiratory matrix (Supplementary Table 3). Twenty-seven reads mapped to human coronavirus 229E, but when BLAST was used to check the identity of these reads, the majority of mis-mapped reads also had high homology with SARS-CoV-2 and were subsequently found to have short read lengths (<40 bp).

Low frequency variant detection. On average 16.7 low frequency variants were detected using all three techniques per sample (range 12-25). However, almost half of these low frequency variants were removed, due to their detection in a single dilution per isolate. Generally, these non-replicated low frequency variants were only detected in low viral load dilutions (1x 10-6 and 1x10-7). Low frequency variants repeatedly detected in at least two dilutions were most commonly detected using RVOP (median number of sites 10, range 6-16) followed by ARTIC (median number of sites 1, range 0-5) and long-amplification (median number of sites 1, range 0-4) and (Supplementary Figure 1, Supplementary Table 2). The presence of low frequency variants was confirmed, at the same genome position by all three methods, in two culture isolates (median number of sites 2, range 0-4): Isolate 1 at positions 657, 27972, 29585, Isolate 2 at positions 12299, 16466 (Supplementary Table 2). No additional low frequency variants were detected using the ARTIC v3 rebalanced pools. Despite using a simulated respiratory matrix to control for background artefacts, there was little consistency in the number and location of low-frequency variants detected across the diluted genomes using each of the three methods.

DISCUSSION

This study highlighted important quality requirements for high-throughput sequencing of SARS-CoV-2 for the purpose of public health surveillance. These parameters are critical for the application of SARS-CoV-2 genomics in tracking transmission pathways and monitoring ongoing viral evolution in circulating virus populations. Sequencing of samples with low viral loads and high Ct values (eg. Ct >33) has been challenging regardless of methodology used [20-23]. Sequencing of such samples can still be attempted, but the resulting genomes often have a substantial portion of missing bases, making it difficult to infer genomic clusters or identify VOC.

Our findings demonstrated the rapid loss of genome coverage using pooled long-amplicon sequencing and the RVOP at Ct >32 (median viral load 1340 copies/µL), indicating

that low viral load or suboptimal RNA quality can be a limiting factor that must be considered when using these methods to generate reproducible genomic data. In contrast, near complete genomes can be recovered using ARTIC v3 at Ct >38, suggesting that the ARTIC protocol is either more sensitive at low viral loads or less impacted by reduced RNA quality. Indeed, the ARTIC protocol has performed well for samples with higher viral loads (Ct <25) [24-26] and has been implemented in numerous laboratories worldwide. However, at lower viral loads, we found both amplification-based methods inconsistently produced data in genomic regions of known significance. Analogous to the findings presented here, uneven amplification efficiencies and coverage bias have been widely reported for low viral load specimens [25, 26]. Increasing coverage over underperforming regions of the genome may be achieved by sequencing at higher depths, but this approach is costly and impractical in outbreak situations where high and rapid throughput is necessary. Rebalancing primer concentrations for ARTIC v3 improved coverage over previously poorly sequenced regions, and it is likely that additional manipulation of primer pooling or primer design would further enhance coverage.

In contrast, the RVOP generates consistent and even SARS-CoV-2 genome coverage over a range of Ct values, despite the sensitivity being only marginally higher than for long-amplicon sequencing. While not examined fully in this study, the RVOP can simultaneously detect other pathogens in a single sample, reducing delays in diagnosis and treatment options for patients who test negative for SARS-CoV-2. Similar to the genome coverage achieved for SARS-CoV-2 in this study, the RVOP should also be able to generate whole genomes of other respiratory viral pathogens targeted by the panel. We were unable to confirm complete coverage of adenovirus and rhinovirus (despite their presence confirmed using RT-PCR) as the pooled respiratory matrix used for this study consisted of a convenient sample of SARS-CoV-2 negative universal transport media (UTM). Poor sample quality as a result of suboptimal transport and storage conditions may have been another contributing factor to the limited and inconsistent coverage of other respiratory pathogens.

The loss of informative sequencing data, especially in genomic regions of interest, can hamper public health efforts to monitor changes in circulating viral populations. Given that numerous VOC have been identified worldwide [14, 27-29], amplicon drop-outs, particularly within the spike region, are problematic. For instance, B6 from the long-amplification protocol, and amplicons 70 and 74 from the original ARTIC v3 protocol encompass part of the spike protein, but all performed poorly and often did not amplify at Ct >32. Rebalancing the ARTIC v3 primer pools increased sequencing coverage and depth over amplicons 70 and 74. However, it is important to note that both long-amplicon and ARTIC v3 methods involve primer binding

prior to amplification and are therefore prone to amplicon drop-outs if variants arise within primer sites. The risk of amplicon dropouts can be overcome by redesigning primers away from variant sites; such protocol changes can be time consuming and difficult to implement but will be necessary given the rapid rise and spread of VOC. The constantly changing population dynamics of the circulating SARS-CoV-2 viruses will require ongoing, high quality genomic surveillance to track the evolution of circulating isolates and help inform necessary changes to sequencing methodologies.

Detecting and locating genomic positions of low frequency variants from culture derived specimens can provide insight into the reliability of intra-host single nucleotide variants (iSNVs) called from clinical specimens. The role of intra-host genomic variability in SARS-CoV-2 may be important in inferring transmission events [30] and may be responsible for significant complications in patients with malignancies [31]. Thus, such low frequency variants require ongoing detection and surveillance. There have been suggestions that iSNVs can be detected at a frequency as low as 2% [18], however only those iSNVs occurring at a frequency of >10% and a minimum coverage of 100x were investigated in the present study. At this threshold, substantial variability of low frequency variants was observed using the methods tested in this study even after controlling for background artefacts generated during the WGS process (via the use of viral cultures in a defined respiratory matrix). The inconsistency in low frequency variants calls can be attributed to the unique sequencing chemistries of each method, and the impact of upstream amplification and hybridisation procedures, highlighting the importance of recognising and accounting for biases that arise during both laboratory preparation and downstream bioinformatic processes.

While we have systematically tested and determined the threshold at which complete genomes can be generated for each method, we have not yet addressed issues with poor quality specimens. Quality and quantities of RNA in clinical specimens for WGS is highly dependent on sample types, collection methods, transport and processing. Suboptimal processes are not uncommon and inherent with high throughput and often centralised testing. Sample degradation as a result of these factors has been highlighted as a significant problem to generating high quality genome sequences [23].

In conclusion, our systematic evaluation of sensitivity and ability to detect low frequency variants demonstrated that overall, the ARTIC v3 protocol was the most sensitive method for generating complete SARS-CoV-2 genomes. The additional advantages of the

ARTIC protocol are better capacity to recover genomes from clinical samples with low viral loads and the ability to detect low frequency variants. Ongoing updates to the ARTIC v3 protocol, such as the rebalancing of primer pools (through the COG-UK and efforts from research institutions), will ensure continual improvements to the WGS process. The optimisation of SARS-CoV-2 genome sequencing can increase the utility of SARS-CoV-2 genomics for COVID-19 cluster detection, transmission tracking and public health responses.

METHODS

Clinical specimens. The study period and region included the 4 months between March and July 2020, in NSW, Australia. SARS-CoV-2 RT-PCR positive specimens which were subsequently cultured at NSW Health Pathology-Institute of Clinical Pathology and Medical Research (ICPMR) in the study period were included for selection. Respiratory samples in UTM which were RT-PCR negative for SARS-CoV-2 were collected and stored at 4˚C. These negative specimens were de-identified and pooled together totalling 40 mL before RNA was extracted. This RNA was used to dilute SARS-CoV-2 isolates, referenced in the manuscript as negative respiratory matrix. Ethical and governance approval for the study was granted by the Western Sydney Local Health District Human Research Ethics Committee (2020/ETH02426).

Viral isolation. SARS-CoV-2 positive respiratory specimens were cultured in Vero C1008 cells (Vero 76, clone E6, Vero E6 [ECACC 85020206]) as previously outlined [32]. Briefly, Vero cell cultures were seeded at 1-3 x 10^4 cells/cm^2 in Dulbecco's minimal essential medium (DMEM, LONZA, Alpharetta, GA, USA) supplemented with 9% foetal bovine serum (FBS, HyClone, Cytiva, Sydney, Australia). Media was replaced within 12 hours with inoculation media containing 1% FBS with the addition of penicillin, streptomycin and amphotericin B deoxycholate to prevent microbial overgrowth and then inoculated with 500 µL of SARS-CoV-2 positive respiratory sample. The inoculated cultures were incubated at 37˚C in 5% CO_2 for 5 days (days 0 to 4). Cell cultures were observed daily for cytopathic effect (CPE). Routine mycoplasma testing was performed to exclude mycoplasma contamination of the cell line and all culture work was undertaken in physical containment laboratory level 3 (PC3) biosafety conditions. The presence of CPE and increasing viral load was indicative of positive SARS-CoV-2 isolation. RT-PCR testing was performed on day 1, 2, 3 and 4 by conducting RNA extraction and SARS-CoV-2 RT-PCR on 200 µL of culture supernatant. Culture supernatant was harvested four days after inoculation and stored at -80˚C.

RNA extraction from viral culture. A total of 600 µL (three x 200 µL) of day 4 SARS-CoV-2 culture supernatant was used as input into the RNeasy Mini Kit (Qiagen) for RNA extraction with minor modifications. 600 µL of RLT buffer was added to 200 µL of sample and mixed well. An equal volume (800 µL) of 70% ethanol was then added and mixed well by pipetting, before loading onto RNeasy column in successive aliquots until the entire volume was extracted. RNA was eluted in 30 µL, pooled for a total of 90 µL and stored at -80˚C prior to dilution. Total RNA was extracted from pooled SARS-CoV-2 negative clinical specimens as above.

Respiratory virus detection by RT-PCR. A previously described RT-PCR [33] targeting the N gene was employed to estimate the viral load of cultured RNA and ensure the absence of SARS-CoV-2 in the negative respiratory matrix. Additional RT-PCRs were used to investigate the presence of common viral respiratory viruses: human influenza viruses A & B, parainfluenzaviruses 1, 2 & 3, respiratory syncytial virus, adenovirus, and rhinovirus in negative UTM extract [34].

Synthetic control. A commercially available synthetic RNA control reference strain (Wuhan - 1 strain, TWIST Biosciences) containing six non-overlapping fragments replicating the most commonly used reference sequence (NCBI GenBank accession MN908947.3) was used as a control. Serial 10-fold dilutions starting at 20,000 copies/µL to 2 copies/µL were made and used to generate a standard curve and quantify the viral load of each culture spiked dilution per reaction. The N gene SARS-CoV-2 RT-PCR was used to determine the viral load of the neat culture RNA after extraction. The synthetic control was also serially diluted 10-fold in respiratory matrix (as outlined below), enriched using each of the methods below and sequenced in parallel with diluted cultures.

Normalisation and serial dilution of viral culture RNA into negative respiratory matrix Based on the viral load of the neat culture RNA (Ct 12.57– 14.48, viral load $2.0x10^8$ - $6.0x10^7$ copies/µL), each culture RNA extract was diluted 1:10 with negative RNA extract. Then 10-fold serial dilutions were made in negative RNA extract until an estimated concentration of >10 copies/µL (Ct 37-40) was reached for each isolate. cDNA was generated for all serially diluted RNA samples using LunaScript RT SuperMix Kit (New England BioLabs). Sufficient volume was prepared to perform duplicates for each method at each dilution. RNA and corresponding cDNA dilutions were aliquoted and stored at -80°C and -20°C, respectively. RT-PCR was then performed for each sample dilution to determine Ct value and corresponding viral load.

Viral enrichment and genome sequencing. For each of the serially diluted samples, viral enrichment was performed using three methods: ARTIC v3, a 14-pool long-amplicon approach, and probe capture using Illumina RNA Prep with Enrichment with the Respiratory Viral Oligo Panel (RVOP). Details of each enrichment method are outlined below.

ARTIC v3 nCoV-2019 sequencing protocol - ARTICv3 (https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye): The ARTIC v3 protocol was performed with the following modifications. Tiling PCR was used to amplify the whole genome according to ARTIC nCoV2-2019 sequencing protocol. Each PCR included 12.5 µL Q5 High Fidelity 2x Master Mix (New England Biolabs), 3.6 µL of either pool 1 or pool 2 10µM primer master mix (final concentration of each primer was ~10-11pM), 5 µL of template, molecular grade water was added to generate a total volume of 25 µL. Cycling conditions were as follows: initial denaturation at 95°C- 2 min, followed by 35 cycles of: 95°C for 30 s, 63°C for 2 min 45 s, and a final extension step of 75°C for 10 min. Pool 1 and pool 2 amplicons were combined and purified with a 1:1 ratio of AMPureXP beads (Beckman Coulter) and eluted in 30 µL of sterile water. Purified products were quantified using Qubit™ 1x dsDNA HS Assay Kit (ThermoFisher Scientific) and diluted to the desired input concentration for library preparation. Sequencing libraries were prepared using Nextera XT (Illumina) according to manufacturer's respective instructions. Sequencing libraries were then sequenced as 2x150bp runs on either the Illumina iSeq or MiniSeq platforms.

An updated ARTIC v3 protocol with rebalanced primer pools was also evaluated in this study. Primers for each ARTIC v3 pool were combined according to updated COG-UK consortium guidelines (https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-bky5kxy6). Subsequent PCR and sequencing using the rebalanced ARTIC primer pools were performed as above.

Pooled long-amplicon PCR (dx.doi.org/10.17504/protocols.io.befyjbpw): Pooled long-amplicon sequencing was performed as described previously [2]. Briefly, 14 overlapping PCR amplicons were independently generated and pooled together in equal volumes. Pooled products were purified with 0.8x AMpure XP beads (Beckman Coulter) and eluted in 30 µL of sterile water. Qubit™ 1x dsDNA HS Assay Kit (ThermoFisher Scientific) was used to quantify pooled amplicons before diluting to the desired input concentration for library preparation. Sequencing libraries were prepared using the Nextera XT kit (Illumina) and sequenced on either iSeq or MiniSeq (Illumina) using 2x76bp paired end reads. No other changes were made to the protocol.

Respiratory Viral Oligo Panel (RVOP): Diluted culture RNA extracts were used as input into the RNA Prep with Enrichment kit (Illumina). RNA denaturation, first and second strand cDNA synthesis, cDNA tagmentation, library construction, clean up and normalisation were performed according to manufacturer's instructions. Individual libraries were then combined in 3-plex reactions for probe hybridisation. The Respiratory Viral Oligo Panel v2 (Illumina) was used for probe hybridisation with the final hybridisation step held at 58°C overnight. Hybridised probes were then captured and washed according to manufacturer's instructions and amplified as follows: initial denaturation 98°C for 30 s, 14 cycles of: 98°C for 10 s, 60°C for 30 s, 72°C for 30 s, and a final 72°C for 5 min. Library quantities and fragment size were determined using Qubit™ 1x dsDNA HS Assay and Agilent HS Tapestation and sequenced using 2x76bp runs on the Illumina MiniSeq.

Bioinformatic analysis. Raw sequence data was processed using an in-house quality control procedure prior to further analysis. Demultiplexed reads were quality trimmed using Trimmomatic v0.36 (sliding window of 4, minimum read quality score of 20, leading/trailing quality of 5 and minimum length of 36 after trimming) [35]. Reference mapping and variant calling was performed using iVar version 1.2 [36]. Briefly, reads were mapped to the reference SARS-CoV-2 genome (NCBI GenBank accession MN908947.3) using BWA-mem version 0.7.17, with unmapped reads discarded. Primer positions were supplied to iVar trim to soft-clip any reads in the bam file which matched primer sequences. Average genome coverage was estimated by determining the number missing bases (N's) in each sequenced genome. Variants were called using iVar variants (min. read depth >10x, quality >20, min. frequency threshold of 0.1). SNPs were defined based on an alternative frequency ≥0.9 whereas low frequency variants were defined by an alternative frequency between 0.1 and 0.9. Low frequency variants with <100x depth were excluded over concerns over reliability of calls where the frequency of either allele dropped below 10. Low frequency variants were only included if they were detected in 2 or more dilutions of each spike culture sequenced. Variants falling in the 5' and 3'UTR regions were excluded due to poor sequencing quality of these regions. Polymorphic sites that have previously been highlighted as problematic were monitored [37]. SARS-CoV-2 lineages were inferred using Phylogenetic Assignment of Named Global Outbreak LINeages v2 (PANGOLIN) (https://github.com/hCoV-2019/pangolin) [38]. The frequency and positions of polymorphisms were compared between dilutions of the same culture and also against the original genome generated from the respiratory specimen and between cultures. Median genome coverage was calculated using the median depth in 50 bp bins across the reference

genome for each method and dilution. Median read depth per amplicon was assessed in non-overlapping segments of each ARTIC v3 amplicon which was then converted to a factor of the expected read coverage (total mapped reads/genome size*150 bp). These factors were compared between original and rebalanced ARTIC v3 sequencing runs. Graphs were generated using R (version 3.6.1).

Analytical performance- sensitivity and specificity. Sensitivity and specificity were calculated for each sequencing method using a consensus SNP approach. For each isolate, a SNP called in any method was considered a 'true positive' SNP if it occurred in two or more sequencing methods at the highest dilution. SNPs identified by a single sequencing method only (and not detected in the original clinical specimen) were considered false positives. Sensitivity was calculated using the following formula A/(A+C) x 100 where A was the number of true positive SNPs and C is the number of false negative SNPs. Specificity was calculated using the formula D/(D+B) x100 where D was the number of true negative bases (within the CDS region) and B was the number of false positive SNPs. Pairwise statistical comparisons were conducted between genome coverage and sensitivities at each dilution across each method using Friedman Test or Mann-Whitney tests with a significance level of $p<0.05$.

Acknowledgements: The authors thank the Sydney Informatics Hub at The University of Sydney, NSW Ministry of Health, and staff from the Microbial Genomics Reference Laboratory and the Virology Laboratory, NSW Health Pathology-ICPMR for their assistance. We are grateful to partner laboratories within the NSW Health Pathology network and private laboratories for their participation in the COVID-19 genomics surveillance and to GISAID for collating international data on SARS-CoV-2.

Funding: This work was conducted with funding from NSW Health as part of the COVID-19 priority grants scheme.

Data availability: Fastq files have been deposited in BioProject PRJNA723901 for all 118 genomes produced in this study. Individual SRA and GISAID accessions can be found in Supplementary Table 4 and 1 respectively.

REFERENCES

1. McAnulty, J.M. and K. Ward, *Suppressing the Epidemic in New South Wales.* N Engl J Med, 2020. **382**(21): p. e74.
2. Eden, J.S., et al., *An emergent clade of SARS-CoV-2 linked to returned travellers from Iran.* Virus Evol, 2020. **6**(1): p. veaa027.
3. Rockett, R.J., et al., *Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling.* Nat Med, 2020. **26**(9): p. 1398-1404.
4. Oude Munnink, B.B., et al., *Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands.* Nat Med, 2020. **26**(9): p. 1405-1410.
5. Meredith, L.W., et al., *Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study.* Lancet Infect Dis, 2020. **20**(11): p. 1263-1271.
6. Gudbjartsson, D.F., et al., *Spread of SARS-CoV-2 in the Icelandic Population.* N Engl J Med, 2020. **382**(24): p. 2302-2315.
7. Hadfield, J., et al., *Nextstrain: real-time tracking of pathogen evolution.* Bioinformatics, 2018. **34**(23): p. 4121-4123.
8. Deng, X., et al., *Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California.* Science, 2020. **369**(6503): p. 582-587.
9. Worobey, M., et al., *The emergence of SARS-CoV-2 in Europe and North America.* Science, 2020. **370**(6516): p. 564-570.
10. Martin, D., et al., *The emergence and ongoing convergent evolution of the N501Y lineages coincided with a major global shift in the SARS-Cov-2 selective landscape.* Virological.org, 2021.
11. Cele, S., et al., *Escape of SARS-CoV-2 501Y. V2 variants from neutralization by convalescent plasma.* medRxiv, 2021.
12. Williams, T.C. and W.A. Burgers, *SARS-CoV-2 evolution and vaccines: cause for concern?* Lancet Respir Med, 2021.
13. Ho, D., et al., *Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization.* Res Sq, 2021.
14. Tegally, H., et al., *Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa.* medRxiv, 2020.

15. Galloway, S.E., et al., *Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021.* MMWR Morb Mortal Wkly Rep, 2021. **70**(3): p. 95-99.
16. Leung, K., et al., *Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020.* Euro Surveill, 2021. **26**(1).
17. Naveca, F., et al., *Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein.* Virological. org, 2021.
18. Lythgoe, K.A., et al., *Shared SARS-CoV-2 diversity suggests localised transmission of minority variants.* bioRxiv, 2020.
19. Illumina, *Detection and characterization of respiratory viruses, including SARS-CoV-2, using Illumina RNA Prep with Enrichment*. 2020, Illumina.
20. Paden, C.R., et al., *Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2.* Emerg Infect Dis, 2020. **26**(10): p. 2401-2405.
21. Doddapaneni, H., et al., *Oligonucleotide capture sequencing of the SARS-CoV-2 genome and subgenomic fragments from COVID-19 individuals.* bioRxiv, 2020.
22. Gohl, D.M., et al., *A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2.* BMC Genomics, 2020. **21**(1): p. 863.
23. Pillay, S., et al., *Whole Genome Sequencing of SARS-CoV-2: Adapting Illumina Protocols for Quick and Accurate Outbreak Investigation during a Pandemic.* Genes (Basel), 2020. **11**(8).
24. Charre, C., et al., *Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation.* Virus Evol, 2020. **6**(2): p. veaa075.
25. Itokawa, K., et al., *Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR.* PLoS One, 2020. **15**(9): p. e0239403.
26. Tyson, J.R., et al., *Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore.* bioRxiv, 2020.
27. Fiorentini, S., et al., *First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020.* The Lancet. Infectious Diseases, 2021.
28. Thomson, E.C., et al., *The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity.* bioRxiv, 2020.
29. Zhang, W., et al., *Emergence of a novel SARS-CoV-2 strain in Southern California, USA.* medRxiv, 2021.

30. Worby, C.J., M. Lipsitch, and W.P. Hanage, *Shared Genomic Variants: Identification of Transmission Routes Using Pathogen Deep-Sequence Data.* Am J Epidemiol, 2017. **186**(10): p. 1209-1216.

31. Siqueira, J., et al., *SARS-CoV-2 genomic and quasispecies analyses in cancer patients reveal relaxed intrahost virus evolution.* bioRxiv, 2020.

32. Basile, K., et al., *Cell-based culture of SARS-CoV-2 informs infectivity and safe de-isolation assessments during COVID-19.* Clin Infect Dis, 2020.

33. Rahman, H., et al., *Interpret with caution: An evaluation of the commercial AusDiagnostics versus in-house developed assays for the detection of SARS-CoV-2 virus.* J Clin Virol, 2020. **127**: p. 104374.

34. Ratnamohan, V.M., et al., *Pandemic clinical case definitions are non-specific: multiple respiratory viruses circulating in the early phases of the 2009 influenza pandemic in New South Wales, Australia.* Virol J, 2014. **11**: p. 113.

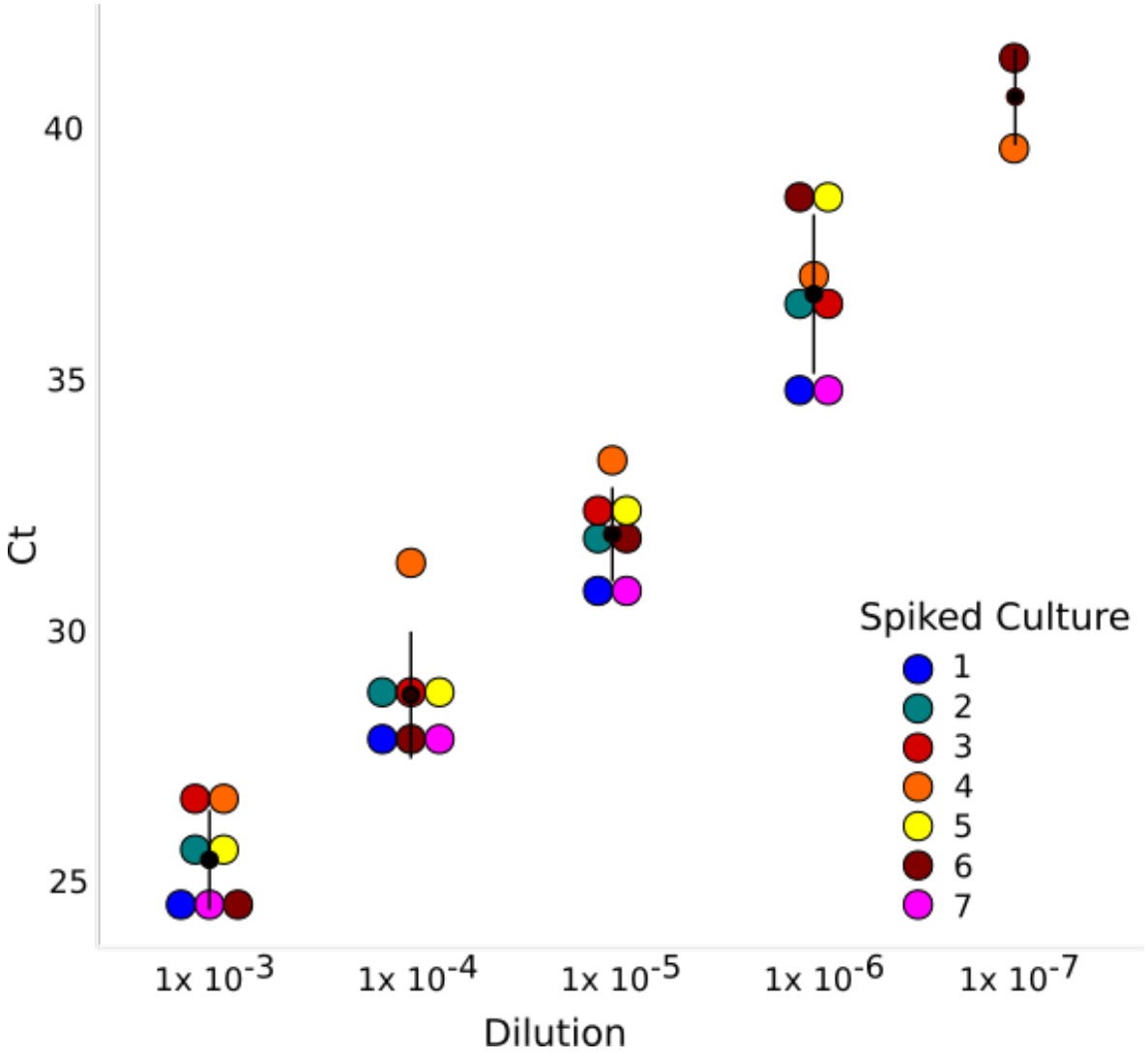
35. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.

36. Grubaugh, N.D., et al., *An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar.* Genome Biol, 2019. **20**(1): p. 8.

37. De Maio, N., et al., *Issues with SARS-CoV-2 sequencing data.* Virological, 2020(6): p. 80-92.

38. Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology.* Nat Microbiol, 2020. **5**(11): p. 1403-1407.

520



521

522 **Figure 1:** Viral load of SARS-CoV-2 cultures spiked in respiratory matrix. RT-PCR
523 quantification of seven serially diluted SARS-CoV-2 cultures demonstrates an increase of 3-4
524 cycles for each ten-fold dilution of viral culture.

525

526

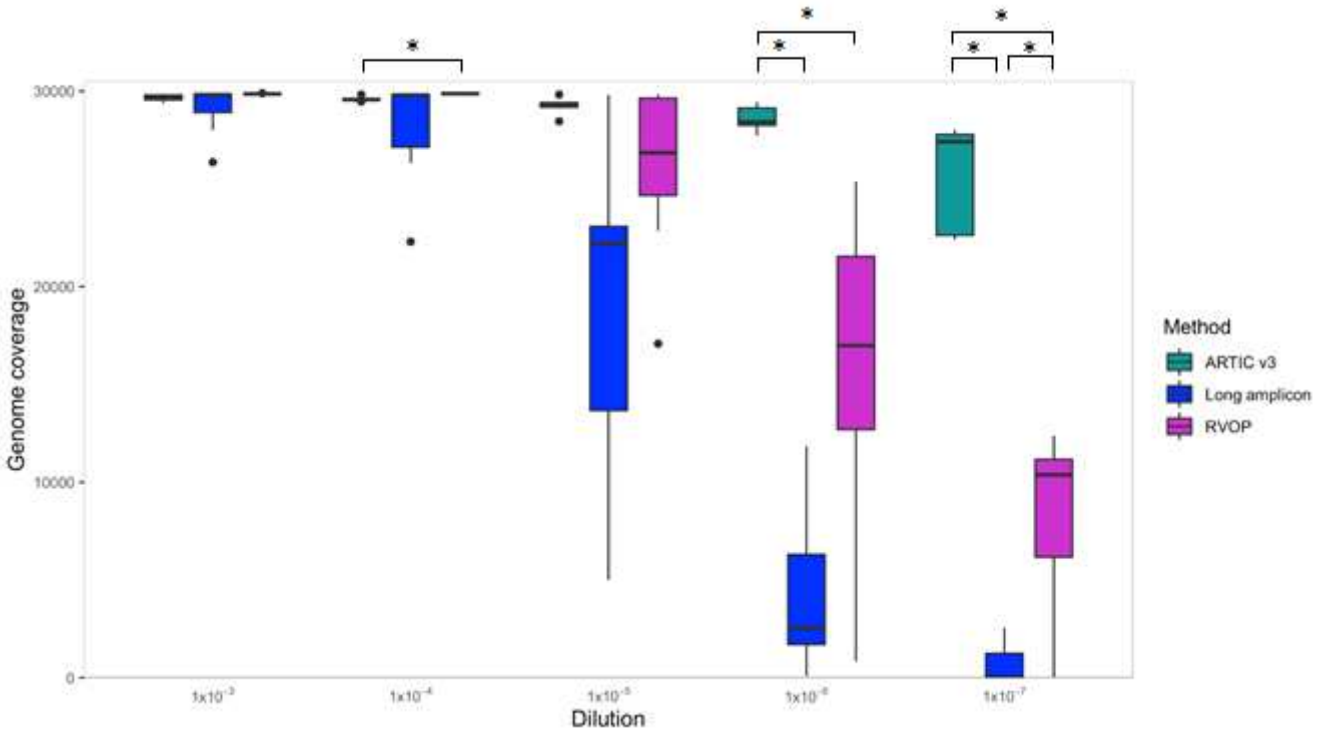


Figure 2. Boxplot showing the genome coverage of ARTIC v3, long-amplification and probe capture (RVOP) based whole genome sequencing methods of SARS-CoV-2 cultures. Significant differences (*) observed in genome coverage between different methods ($p<0.05$). NB pairwise comparisons between each methodology were only performed within each dilution.

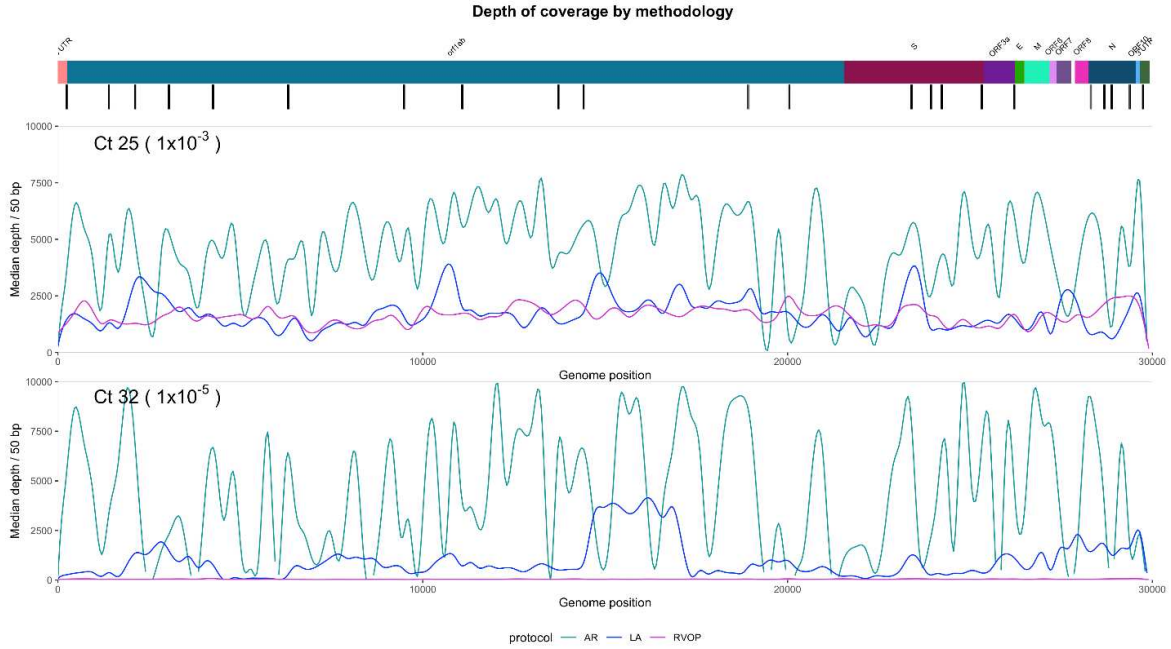


Figure 3: Overall read depth across the SARS-CoV-2 genome using ARTIC v3 (green line), long-amplification (blue line), and probe capture RVOP (pink line) whole genome sequencing methods. Depth is averaged across all samples for each method separately. Lines are smoothed by using the 'geom_spline' function in R. The coloured bar at the top represents the regions of the SARS-CoV-2 genome, and black bars represent informative single nucleotide polymorphisms.

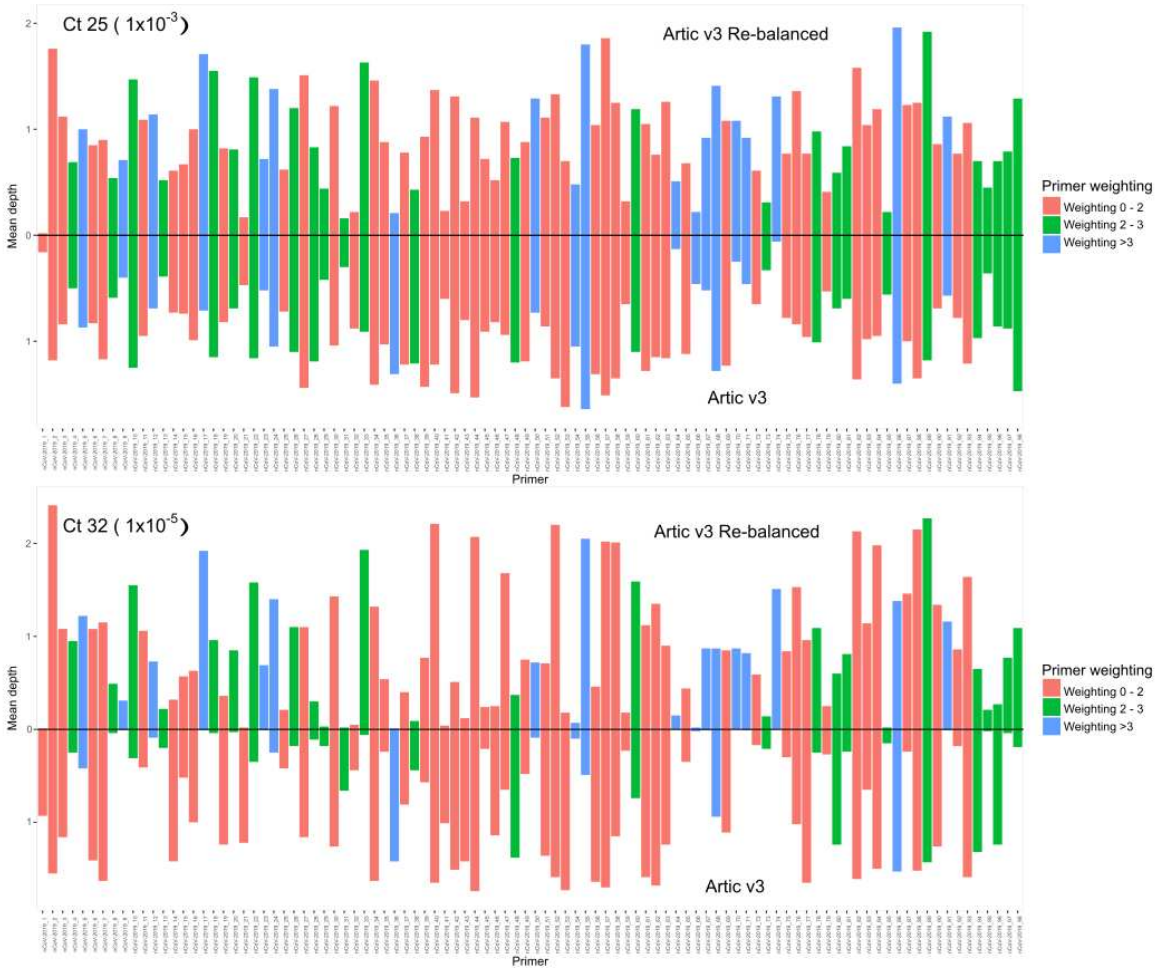


Figure 4: Comparison between ARTIC v3 original primer pooling and rebalanced primer pools at Ct 25 and Ct 32. Median depth is expressed as a factor of average read depth across each amplicon. Bars above zero represent sequencing depth achieved by ARTIC v3 rebalanced primer pools; height of bars below zero represent sequencing depth of ARTIC v3 original primer pools. ARTIC v3 primers are listed across the x-axis in sequential order across the genome. For exact primer weightings, refer to: https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-bky5kxy6).

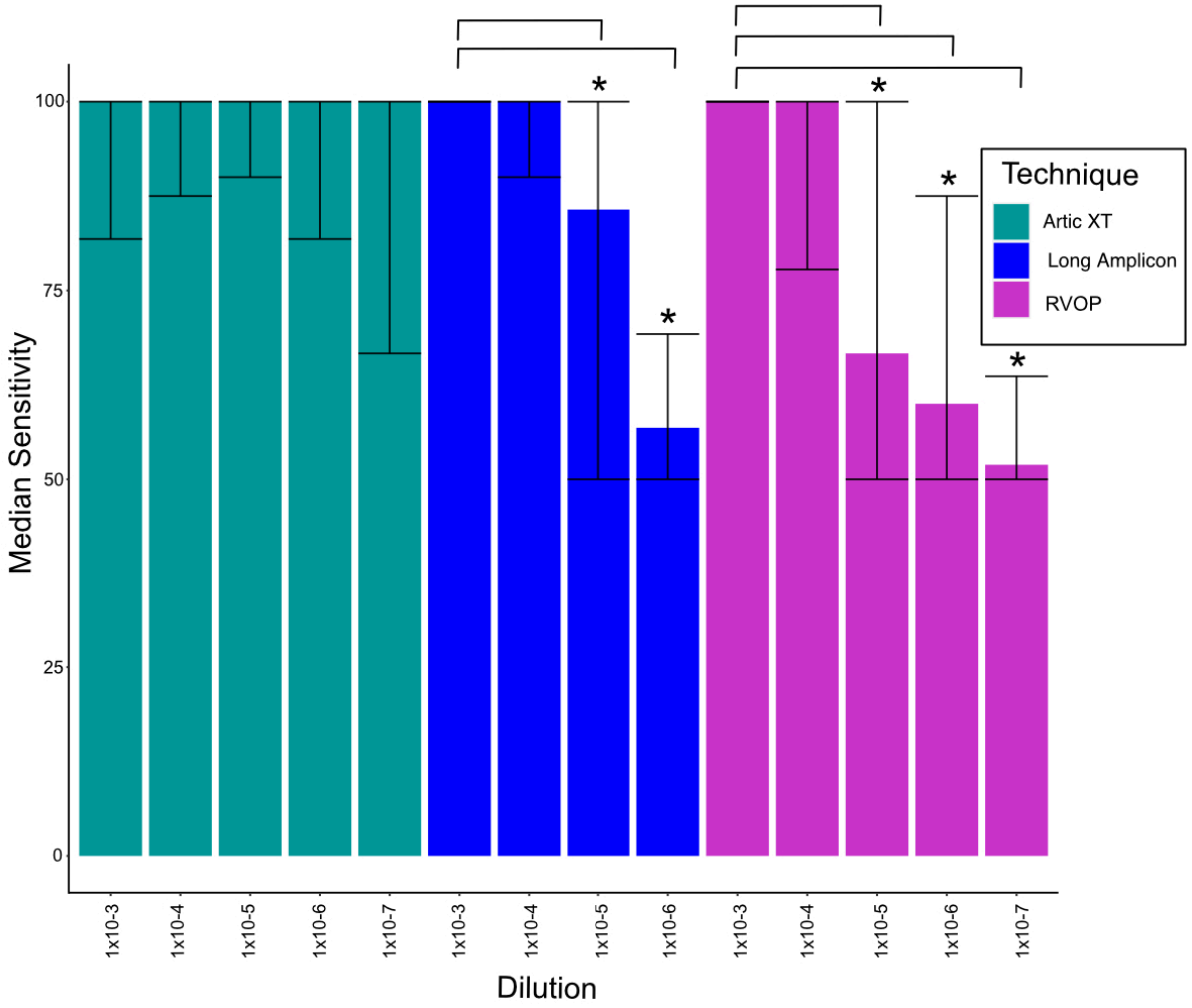


Figure 5: Sensitivity of ARTIC v3, long-amplification and RVOP (Respiratory Viral Oligo Panel, Illumina) whole genome sequencing methods. The sensitivity of ARTIC v3 was the highest across all viral load dilutions. No sensitivity calculations could be made for the pooled long-amplification method at 10^{-7} dilutions due to insufficient amplicons for variant calling. Significant differences (*) in sensitivity were observed between dilution 10^{-3} and dilutions 10^{-5} and 10^{-6} using the long-amplification method and dilution 10^{-3} and dilutions 10^{-5} , 10^{-6} and 10^{-7} using the RVOP method. No differences were observed between ARTIC v3 original primer pooling and rebalanced pools.

564 **Supplementary materials**

565 **Supplementary Table 1: Genomes of SARS-CoV-2 isolates used in the study**

Isolate ID	GISAID ID	Lineage	SNP profile
Isolate 1 (NSW05)	EPI_ISL_412975	B.4	G:1397:A, G:4255:A, *G:11083:T*, A:20047:G, T:28688:C, G:29742:T
Isolate 2 (NSW13)	EPI_ISL_413599	B.4	G:1397:A, C:2113:T, G:11083:T, C:18928:T, C:26213:T*, T:28688:C, G:29374:A, G:29742:T
Isolate 3 (NSW08)	EPI_ISL_413594	B	G:25323:T
Isolate 4 (NSW14)	EPI_ISL_413600	B.4	G:1397:A, G:4255:A, *G:11083:T*, A:20047:G, T:28688:C, G:29742:T
Isolate 5 (NSW155)	EPI_ISL_427647	B.1	C:241:T, C:3037:T, C:14408:T, A:23403:G, G:28881:A, G:28882:A, G:28883:C
Isolate 6 (NSW26)	EPI_ISL_417389	B.4	G:1397:A, C:2113:T, A:9483:G, *G:11083:T*, C:18928:T, G:24227:A, T:28688:C, G:29374:A, G:29742:T
Isolate 7 (NSW48)	EPI_ISL_417403	B.6	C:6312:A, *G:11083:T*, C:13730:T* , C:23929:T, C:28311:T

566 The * symbol denotes SNPs lost in culture. The mutation lost in Isolate 2 was detected as a low
567 frequency variant. The SNP lost during culture of Isolate 7 was not detected as a low frequency
568 variant. Italics indicate known hypervariable sites.

569

570 **Supplementary Table 2. Low frequency variants detected using each method**

Sample ID	Long-Amp	ARTIC v3	RVOP
TWIST Control - UTM	3350, 6669, *11074(+T)*, 11079	*11074(+T)*, 14707, 26791, 26793, 26794, 26796	**5765**, **5766**, 10001, *11074(+T)*, **12413**, **12926** (+C), **23652**, **26433**, 26791, 26793, 26794, 26794
TWIST Control - Water	3350, 6669	26791, 26793, 26794, 26796	N/A
Isolate 1.	657, *11082(-N)*, 27972, 29585	657, *11082(-N)*, 27972, 29585	657, **5765**, **5766**, *11082(-N)*, **12413**, **12926 (+C)**, 15071, 17561, 18408, 18848, 20079, 23403, **23652**, **26433**, 27870, 27972, 29585
Isolate 2.	*10323*, *11074(+T)*,*11082(-N)*, 12299, 16466	*10323*, *11082(-N)*, 12299, 16466	**5765**, **5766**, *11082(-N)*, 12299, 16466, 18402, 21055, 21949, **23652**
Isolate 3.	-	-	**5765**, **5766**, **12413**, **12926 (+C)**, **23652**, **26433**, 27870
Isolate 4.	*11082(-N)*	*11074(+T)*, *11082*	**5765**, **5766**, *11082(-N)*, **12413**, **12926 (+C)**, 17561, 20079, 23403, **23652**, **26433**, 27870
Isolate 5.	695, 29190	-	5765, 5766, 12413, 12926, 23652, 27870
Isolate 6.	11050, *11082(-N)*	*11082(-N)*	**5765, 5766**, *11082(-N)*, **12413**, **12926 (+C)**, 15071, 19812, 21949, **23652**, **26433**, 27870

Isolate 7.	6310, *11074(+T)*,	*11082(-N)*	**5765**, **5766**, 6310, *11082(-N)*, **12413**, **12926 (+C)**, 15071, **23652, 26433**, 27870

Low frequency variant positions are based on the reference SARS-CoV-2 genome (NCBI GenBank accession MN908947.3). Underline and italic positions are known regions of hypervariability commonly masked from phylogenetic analysis, indels are denoted with brackets indicating the inserted base or N for a deleted base. Bold positions indicate iSNV sites which occurred in both the TWIST control and spiked isolate RNA, indicative that the iSNV is likely an artefact and may occur from non-specific mapping of non-SARS-CoV-2 RNA.

Supplementary Table 3:

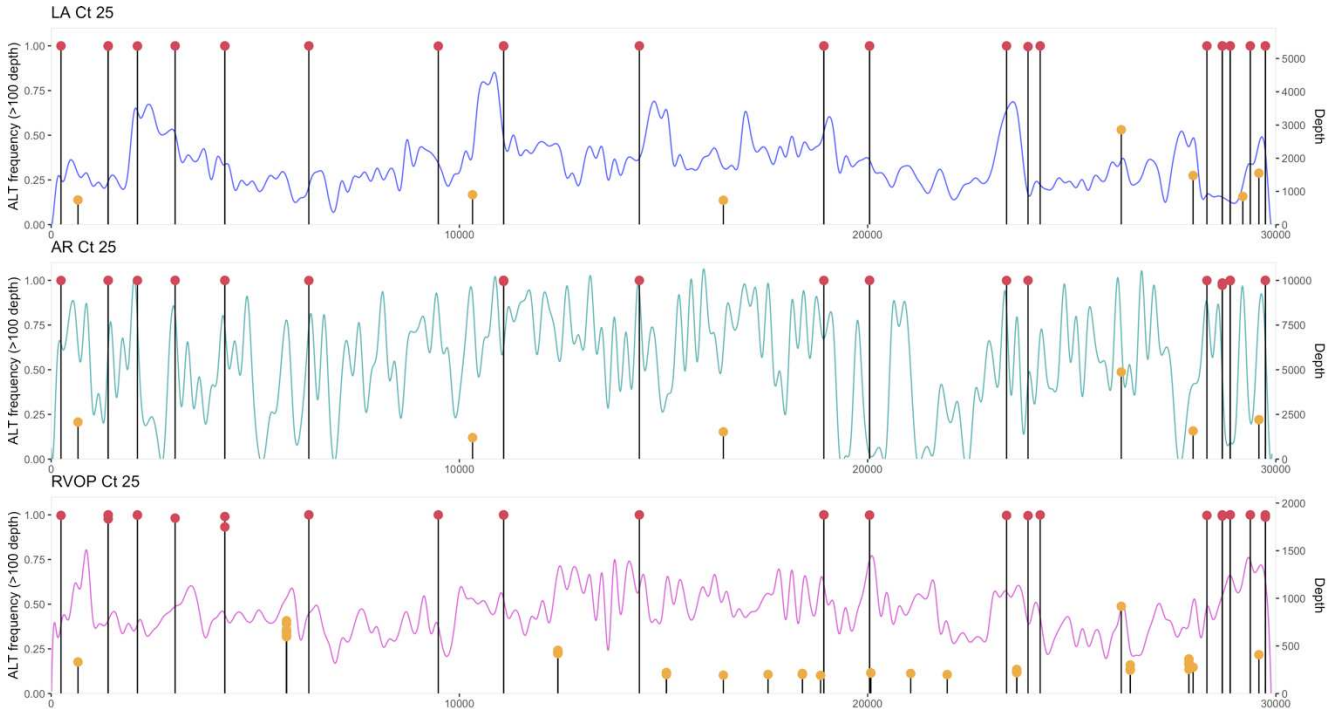
Negative SARS-CoV-2 RNA extract and detection of other respiratory pathogens

Respiratory pathogen	Ct Value
Rhinovirus	27
Adenovirus	27
Influenza	ND
RSV	ND
Parainfluenza	ND

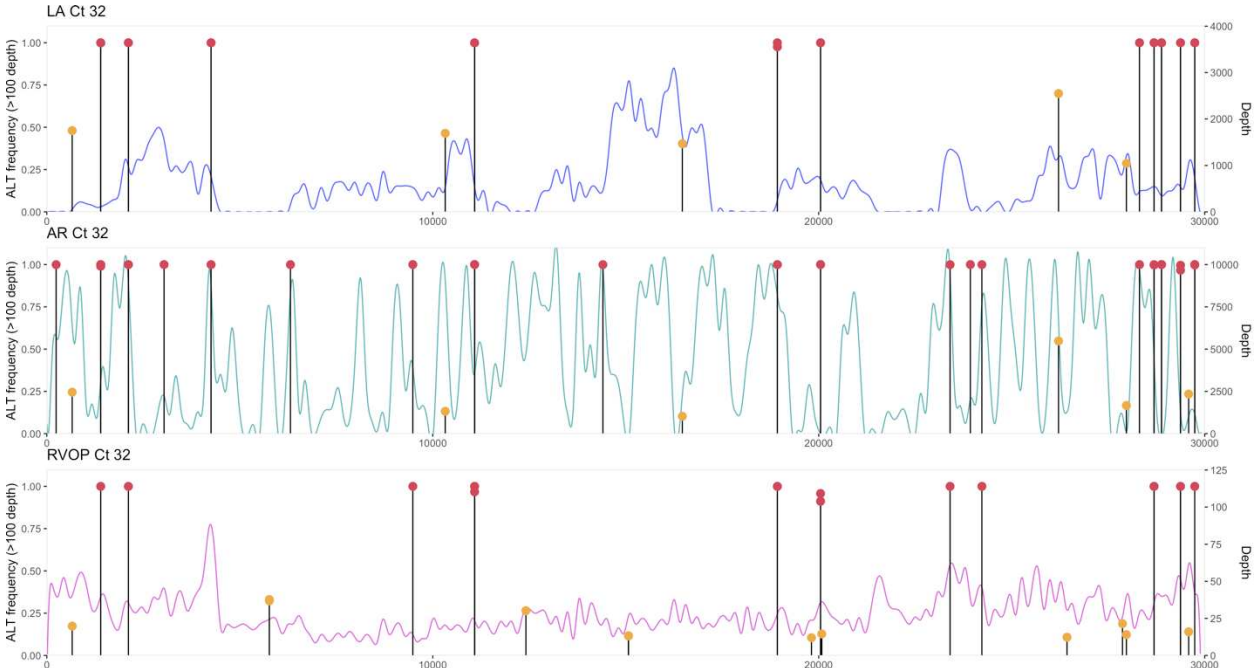
Supplementary Table 4: Details of SRA data availability for all 118 genomes produced in the study

Supplementary Figure 1: Comparison of read depth for Long-amplification method (LA-blue), ARTIC v3 (AR- green) and the respiratory viral oligo panel (RVOP- pink) at Ct 25 (panel A) and Ct 32 (panel B) and averaged across samples. The depth of coverage is shown on the right axis, while the proportion of reads for single nucleotide polymorphisms (SNPs) and low frequency variant calling is on the left axis. The position of SNPs (red circles) and low frequency variants (orange circles) across the genome are overlaid on top of the both coverage graphs.

A)



B)



603