

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

Bede Constantinides^{*1,4}, Hermione Webster^{*1,3}, Jessica Gentry², Jasmine Bastable², Laura Dunn², Sarah Oakley², Jeremy Swann¹, Nicholas Sanderson^{1,3}, Philip W Fowler^{1,3}, Geoffrey Ma^{1,3}, Gillian Rodger^{1,4}, Lucinda Barrett², Katie Jeffery^{2,5}, Timothy EA Peto^{1,2,3}, Nicole Stoesser^{1,4}, Teresa Street^{1,3}, Derrick W Crook^{1,2,3}

¹ NDM Experimental Medicine, University of Oxford, Oxford, United Kingdom

² Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

³ NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom

⁴ NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance in partnership with Public Health England, Oxford, United Kingdom

⁵ Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

Abstract

Genome sequencing is pivotal to SARS-CoV-2 surveillance, elucidating the emergence and global dissemination of acquired genetic mutations. Amplicon sequencing has proven very effective for sequencing SARS-CoV-2, but prevalent mutations disrupting primer binding sites have necessitated the revision of sequencing protocols in order to maintain performance for emerging virus lineages. We compared the performance of Oxford Nanopore Technologies (ONT) Midnight and ARTIC tiling amplicon protocols using 196 Delta lineage SARS-CoV-2 clinical specimens, and 71 mostly Omicron lineage samples with S gene target failure (SGTF), reflecting circulating lineages in the United Kingdom during December 2021. 96-plexed nanopore sequencing was used. For Delta lineage samples, ARTIC v4 recovered the greatest proportion of $\geq 90\%$ complete genomes (81.1%; 159/193), followed by Midnight (71.5%; 138/193) and ARTIC v3 (34.1%; 14/41). Midnight protocol however yielded higher average genome recovery (mean 98.8%) than ARTIC v4 (98.1%) and ARTIC v3 (75.4%), resulting in less ambiguous final consensus assemblies overall. Explaining these observations were ARTIC v4's superior genome recovery in low viral titre/high cycle threshold (Ct) samples and inferior performance in high titre/low Ct samples, where Midnight excelled. We evaluated Omicron sequencing performance using a revised Midnight primer mix alongside prototype ARTIC v4.1 primers, head-to-head with the existing commercially available Midnight and ARTIC v4 protocols. The revised protocols both improved considerably the recovery of Omicron genomes and exhibited similar overall performance to one another. Revised Midnight protocol recovered $\geq 90\%$ complete genomes for 85.9% (61/71) of Omicron samples vs. 88.7% (63/71) for ARTIC v4.1. Approximate cost per sample for Midnight (£12) is lower than ARTIC (£16) while hands-on time is considerably lower for Midnight (~7 hours) than ARTIC protocols (~9.5 hours).

Introduction

SARS-CoV-2 sequencing is pivotal to global surveillance and the international and national public health response. The ARTIC Network's open source tiled amplicon protocol is widely used for sequencing SARS-CoV-2, has been updated to reflect changes in circulating lineages internationally, and is suitable for low cost, highly-multiplexed sequencing using popular Illumina and Oxford Nanopore Technologies (ONT) platforms (Tyson et al., 2020; DNA Pipelines R&D et al., 2020). ONT's Midnight protocol is one of a number of commercially available alternative protocols, and offers reduced turnaround time and lower cost per sample than ARTIC v4 for nanopore sequencing (Freed et al., 2020). We sought to compare the performance of ARTIC and Midnight protocols, while also considering the practice.

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

resource requirements of the both methods, to support establishing a routine SARS-CoV-2 sequencing service in a clinical laboratory. For Omicron lineages we also evaluated the latest revision of both tiling amplicon protocol, to which we refer as Midnight v2 and ARTIC v4.1. Reflecting the prevailing lineages in the United Kingdom in late 2021, we sequenced clinical samples comprising the WHO-designated Delta (Pango AY.4, AY.4.2 and B.1.617.2-like) and Omicron (BA.1-like) lineages as assigned by Pangolin (Rambaut et al., 2020).

Results

As part of ongoing of evaluation work, we characterised 196 PCR-positive clinical samples of Delta lineage collected during November 2021, and 71 mostly Omicron samples collected in December 2021 exhibiting S gene target failure (SGTF), of which 48 were assigned Omicron by Pangolin's rule-based Scorpio classifier.

For Delta samples, using a criterion of at least 90% coverage at depth ≥ 20 , ARTIC v4 recovered the greatest proportion of nearly-complete genomes (81.1%; 159/193), followed by Midnight (71.5%; 138/193), and ARTIC v3 (34.1%; 14/41). However, in terms of the proportion of the genome recovered with a depth ≥ 20 (cov20), Midnight protocol yielded greater average genome recovery (mean 98.8%; median 99.4%) than both ARTIC v4 (mean 98.1%; median 98.8%) and ARTIC v3 (mean 75.4%; median 82.5%), resulting in less ambiguous consensus assemblies than ARTIC in the majority of cases (Figures 1 and 2). Midnight typically recovered more genome in samples with Ct value < 25 (see Figure S1 for Ct distributions), but above this its performance deteriorated more rapidly than that of ARTIC v4 (Figure 2A). Genome recovery (cov20) and unambiguous consensus sequence coverage were strongly correlated (Pearson's $r > 0.99999$, $p < 0.00001$; Figure S2). ARTIC v4 notably suffered from a ubiquitous dropout at 6850-7080bp (ORF1A) in our testing (Figure 3). In Delta samples, Midnight yielded greater genome recovery in 62.2% (122/196) of samples.

For Omicron samples, we compared performance of the existing commercially available Midnight and ARTIC v4 protocols alongside the latest Midnight v2 and ARTIC v4.1 protocols, both featuring revised primer mixtures designed to address defects associated with the constellation of mutations defining Omicron lineage SARS-CoV-2. The revised protocols both offered improved performance compared to respective previous versions, and exhibited similar overall performance to one another in the test panel of SGTF-selected clinical Omicron samples. Using the same criterion of $\geq 90\%$ cov20, ARTIC v4.1 and v4 recovered the same proportion of nearly-complete Omicron genomes (88.7%; 63/71), followed by Midnight v2 (85.9%; 61/71), and Midnight (83.1%; 59/71). Using a more stringent measure of genome recovery ($\geq 99\%$ cov20) highlighted much greater differences between protocols: Midnight v2 recovered 74.6% (53/71) of samples, followed by ARTIC v4.1 (57.7%; 41/71), followed by Midnight (19.7%; 14/71), followed by ARTIC v4 (0%; 0/71). See Figure 4 for the empirical distribution of cov20 by protocol. Giving rise to the revised protocols' improved performance were reductions in amplicon failure with respect to existing versions. A dropout in Midnight protocol affecting genes ORF8, ORF9b and N appeared fully resolved in Midnight v2, as were ARTIC v4 dropouts in S gene (Figure 5). Dropouts in ARTIC v4's E and M genes appeared largely if not entirely addressed by the v4.1 protocol (Figure 5), although it should be noted that the V4.1 primers used here were unbalanced, and that balancing would be expected to further improve performance.

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

Discussion

Community-developed open source amplicon sequencing protocols have seen widespread adoption in the SARS-CoV-2 pandemic, enabling reliable and cost-effective whole genome sequencing, in doing so maximising the global reach of genomic surveillance. Both ARTIC v4 and Midnight protocols performed well in our testing for Delta lineage sequences, with ARTIC attaining considerably greater coverage in high Ct clinical samples and Midnight providing marginally better performance in low Ct samples. Omicron lineage reduces the performance of both existing protocols, although is arguably more problematic for ARTIC v4, causing dropouts in the gene encoding spike protein. The latest versions of the Midnight and ARTIC protocols seem to effectively mitigate these issues, attaining similarly high genome recovery in samples below Ct ~28, albeit with tendencies for dropouts in different regions. It should be noted that while neither Midnight v2 nor ARTIC v4.1 protocols as evaluated here used finalised commercially available kits, the Midnight v2 primers used were balanced, while the ARTIC v4.1 primers were unbalanced, likely to the detriment of ARTIC v4.1 as tested. By using transpose-based 'rapid barcoding' in place of the more time-consuming 'native barcoding' process, Midnight reduces hands-on time from approximately 9.5 to ~7 hours at an estimated cost per sample of £12 vs. ARTIC's £16. Given its broadly comparable performance to ARTIC with currently circulating Delta and Omicron lineages, Midnight protocol therefore presents an appealing alternative for rapid turnaround SARS-CoV-2 sequencing using nanopores.

Methods

Samples

Clinical samples used in this evaluation were laboratory discards, used only after routine diagnostic workflows were complete. Sample sets 1-3 comprised a total of 196 distinct PCR-positive samples. Clinical negative controls were randomly included throughout each plate. Sample processing and analysis was covered by ethical approval for the use of anonymised oro- or nasopharyngeal specimens from patients for the diagnosis of influenza and other respiratory pathogens, including SARS-CoV-2 (North West-Greater Manchester South Research Ethics Committee [REC], REC Ref:19/NW/0730).

RNA extraction

RNA was extracted from nasopharyngeal swabs using the KingFisher Flex platform MagMax™ 96 Viral/Pathogen kit (Thermo Fisher Scientific, Waltham, MA, USA), following the 400µl extraction protocol. To achieve a higher yield of RNA for sequencing set 4 (predominantly Omicron samples), RNA was extracted from two 400µl sample inputs eluted into 100µl each which were combined prior to RT-PCR. TaqPath RT-PCR identified positive COVID-19 samples based on N-gene, S gene and ORF1ab amplification, as well as providing cycle threshold (Ct) values.

Reverse Transcription

Reverse Transcription (RT) conditions were identical for both ARTIC and MIDNIGHT sequencing protocols. A 4:1 concentration of RNA to Luna Script RT SuperMix was

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

incubated at 25°C for 2 minutes, 55°C 10 mins and 95°C for 1 minute. Each sample set included a negative control at this step.

ARTIC protocols

For ARTIC v3, after reverse transcription, the NEBNext Var Skip Short Express Protocol (New England Biolabs, 2021) was followed using ARTIC v3 primers. For ARTIC v4, 10µM of v4 primers were substituted for v3 primers at the PCR step in the same protocol. The ARTIC v3 protocol was evaluated only on sample set one (41 positive samples), and performing poorly, was not evaluated further.

Midnight protocol

The Midnight PCR tiling of SARS-CoV-2 virus was used with rapid barcoding and the Midnight RT PCR Expansion (SQK-RBK110.96 and EXP-MRT001). PCR conditions differed between sample sets, reflecting changing manufacturer best practices. For sample set one, the Midnight revision H protocol was followed with PCR annealing and extension conditions set at 65°C for 5 minutes. For sample sets two, three and four, protocol revision M was followed, with PCR annealing and extension conditions respectively set at 61°C for 2 minutes and 65°C for 3 minutes.

Sequencing

Pooled libraries generated by both protocols were sequenced with an ONT GridION using a total of eleven ONT R9.4.1 flow cells.

Sequence analysis

Sequences were basecalled using Guppy 5.0.16 in high accuracy mode. Sequence analysis was performed using a forked version of the EPI2ME Labs wf-artic Nextflow pipeline (<https://github.com/bede/wf-artic/tree/0.3.6-patched>), namely incorporating ARTIC Fieldbioinformatics (Loman, 2020), Nextclade (Aksamentov et al., 2021) and Pangolin (Rambaut et al., 2020). Depth of coverage and cov20 was calculated from primer-trimmed BAM files using Samtools (Li et al., 2009) depth, including aligned deletions in the depth calculation. Pangolin version 3.1.17 with PangoLEARN version 2021-12-06 was used. JupyterLab (Kluyver et al., 2016) was used for further analysis and visualisation.

Data availability

Per-sample depth of coverage, consensus sequences and summary data tables are available at <https://github.com/bede/sars2-seq-eval>

Competing interests

Some Midnight sequencing kits used in this study were provided free of charge by Oxford Nanopore Technologies.

Acknowledgements

We thank Joshua Quick for developing and sharing the revised ARTIC v4.1 primers for testing.

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

Funding

This study is supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance (NIHR200915), a partnership between the UK Health Security Agency (UKHSA) and the University of Oxford. Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR or the UKHSA.

References

- Aksamentov, I., Roemer, C., Hodcroft, E., & Neher, R. (2021). Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, 6(67), 3773. <https://doi.org/10.21105/joss.03773>
- DNA Pipelines R&D, Farr, B., Rajan, D., Betteridge, E., Shirley, L., Quail, M., Park, N., Redshaw, N., F Bronner, I., Aigrain, L., Goodwin, S., Thurston, S., Lensing, S., Bonfield, J., James, K., Salmon, N., Beaver, C., Nelson, R., K. Jackson, D., ... Johnston, I. (2020). *COVID-19 ARTIC v3 Illumina library construction and sequencing protocol v5* [Data set]. <https://doi.org/10.17504/protocols.io.bibtkann>
- Freed, N. E., Vlková, M., Faisal, M. B., & Silander, O. K. (2020). Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols*, 5(1), bpaa014. <https://doi.org/10.1093/biomethods/bpaa014>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90). IOS Press.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.

<https://doi.org/10.1093/bioinformatics/btp352>

Loman, N. J. (2020). *The ARTIC field bioinformatics pipeline*.

<https://github.com/artic-network/fieldbioinformatics>

New England Biolabs, N. (2021). *NEBNext® Varskip Short ARTIC SARS-CoV-2 FS Library Prep Kit (Illumina®) E7658 Express Protocol with One Clean-up Step v1* [Preprint].

<https://doi.org/10.17504/protocols.io.bx9apr2e>

Oxford Nanopore Technologies. (n.d.). *ARTIC SARS-CoV-2 workflow and reporting*.

<https://github.com/epi2me-labs/wf-artic>

Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5(11), 1403–1407.

<https://doi.org/10.1038/s41564-020-0770-5>

Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. H., Lapointe, H., Kamelian, K., Smith, A. D., Prystajeky, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020). *Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore* [Preprint]. Genomics. <https://doi.org/10.1101/2020.09.04.283077>

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

Figures

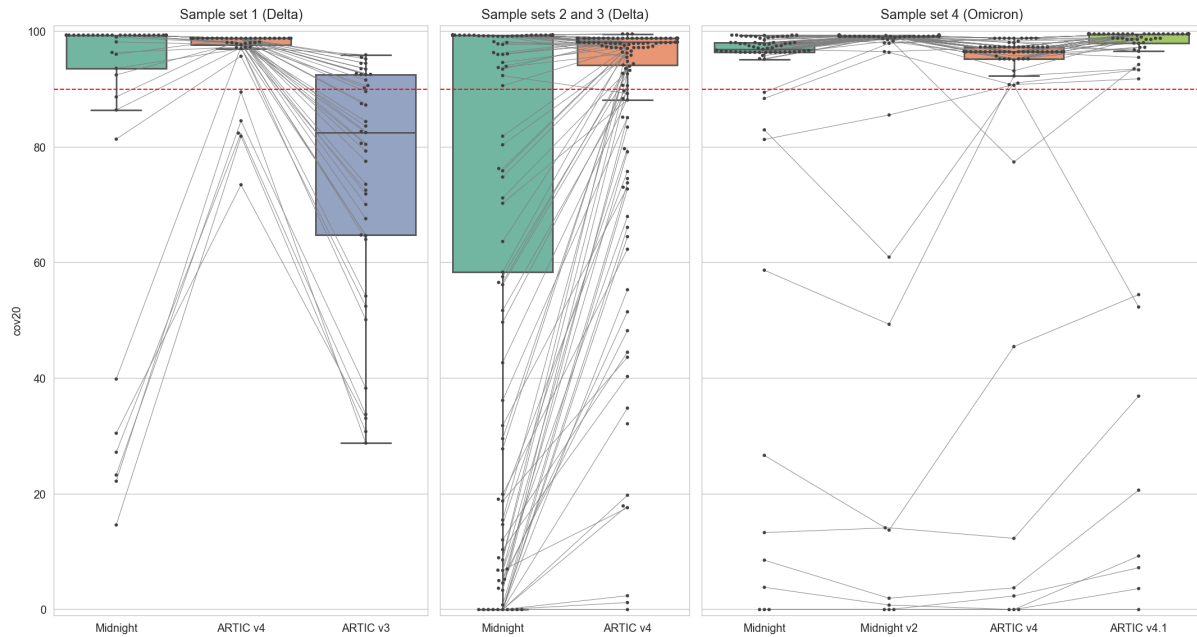


Fig. 1. Genome recovery at or above depth 20 in head-to-head sequencing of 196 Delta and Omicron lineage SARS-CoV-2 clinical samples over a total of 9 runs. Red lines correspond to a widely used lower bound on genome recovery for lineage assignment (90%).

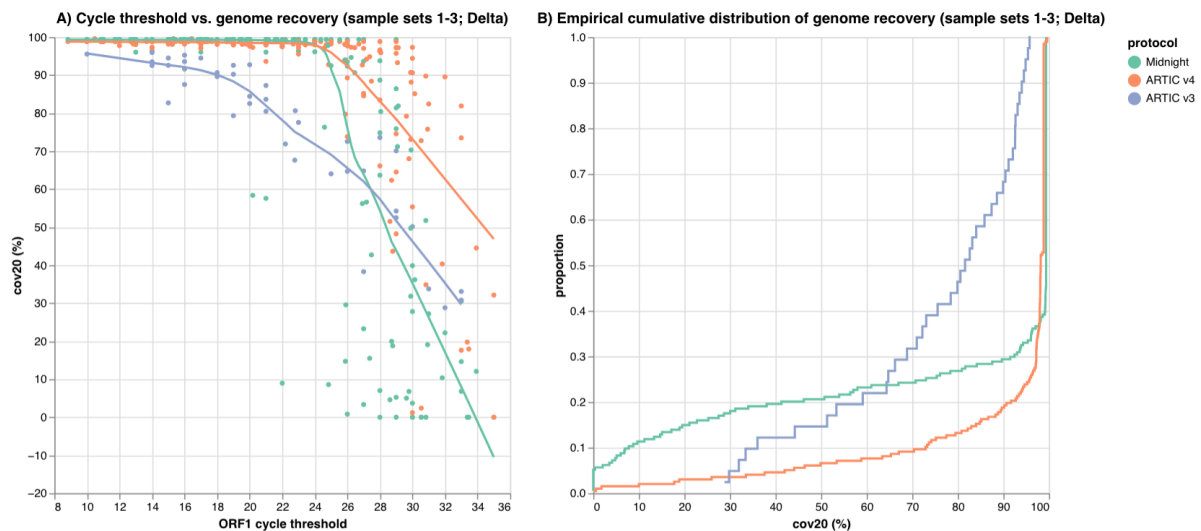


Fig. 2. A) TaqPath ORF1 gene cycle threshold (Ct) vs. genome recovery at \geq depth 20 (cov20) in 196 Delta lineage samples with LOESS regression lines. B) Empirical cumulative distribution of cov20. ARTIC v4 yielded much higher typical genome recovery than Midnight in low viral titre, high Ct samples while Midnight recovered more genome in high viral titre, low Ct samples. Shown are sample sets 1-3 for Midnight and ARTIC v4, and sample set 1 only for ARTIC v3, which performed poorly.

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

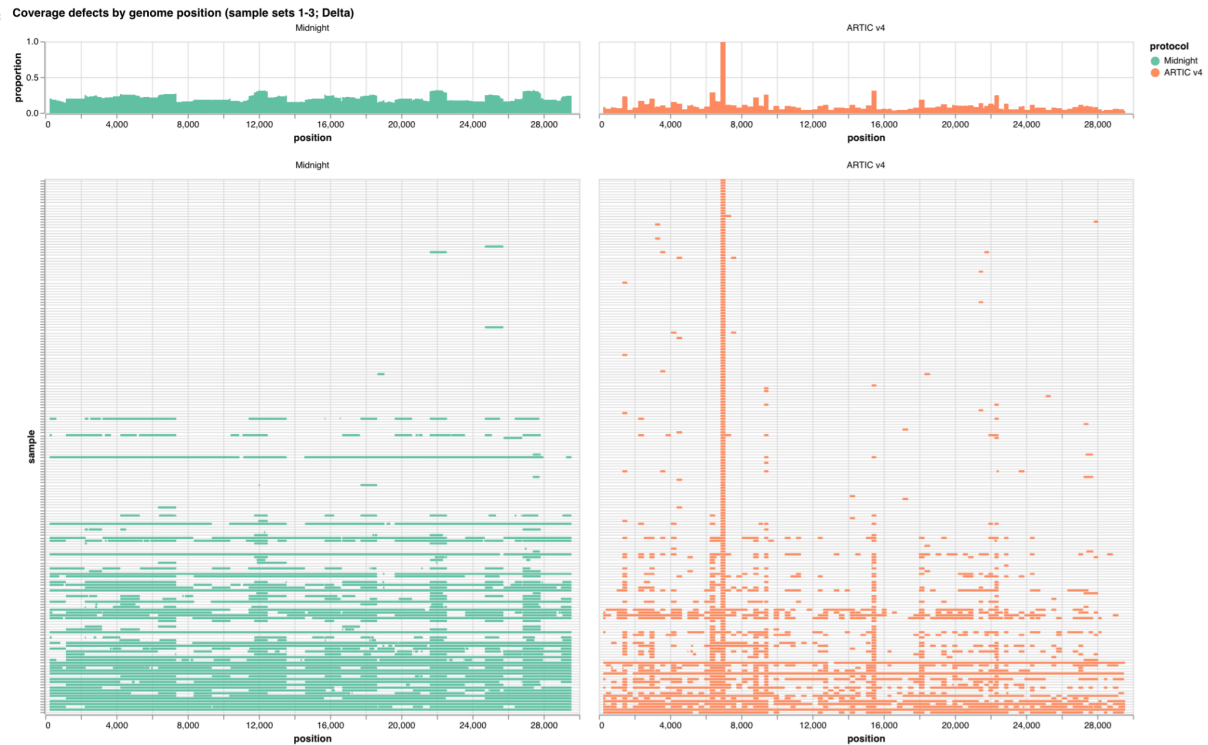


Fig. 3. Coverage defect distribution (depth < 20) in 196 Delta lineage SARS-CoV-2 clinical samples, sorted by ascending ORF1 gene Ct value. ARTIC v4 exhibited a systematic defect at approximately 6850-7080bp (S gene) (right).

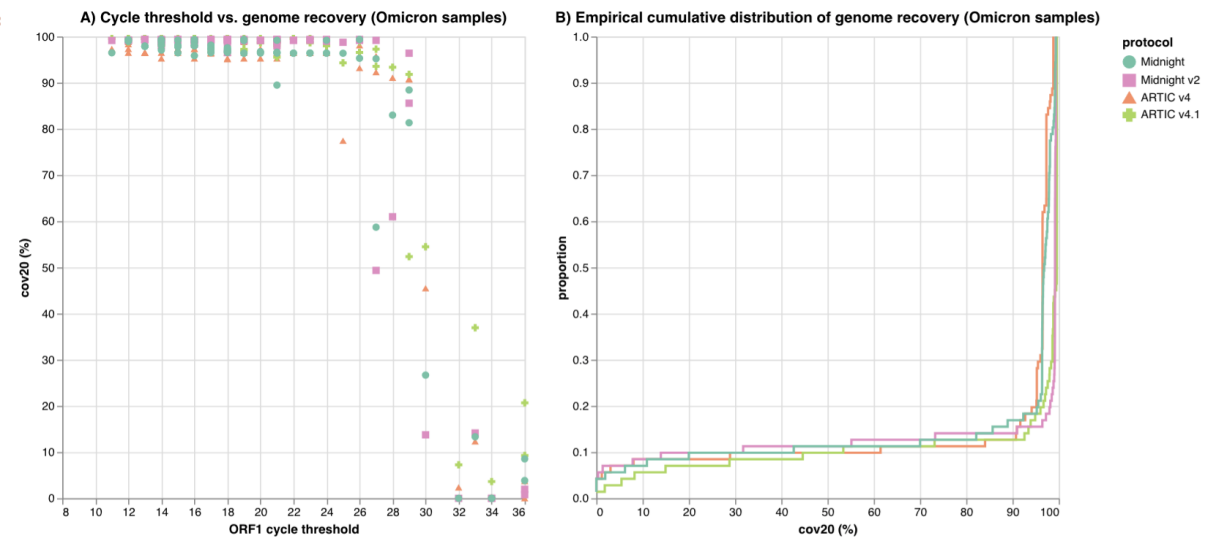


Fig. 4. A) Comparison of existing and revised (v2) Midnight protocols in terms of ORF1 cycle threshold vs. genome recovery (cov20) in sample set 4, which comprises Omicron lineage samples selected by S gene target failure. LOESS lines omitted here for lack of data. B) Cumulative empirical cov20 distributions for existing and revised Midnight protocols.

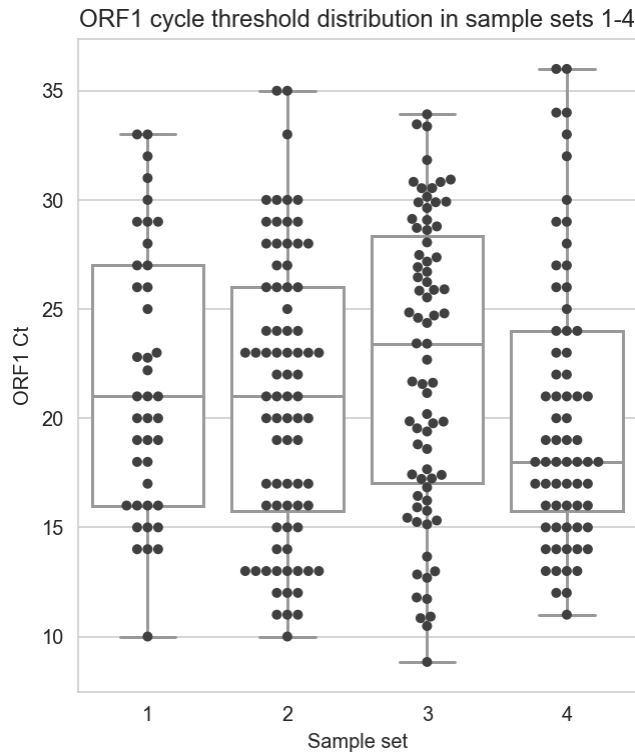
Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance



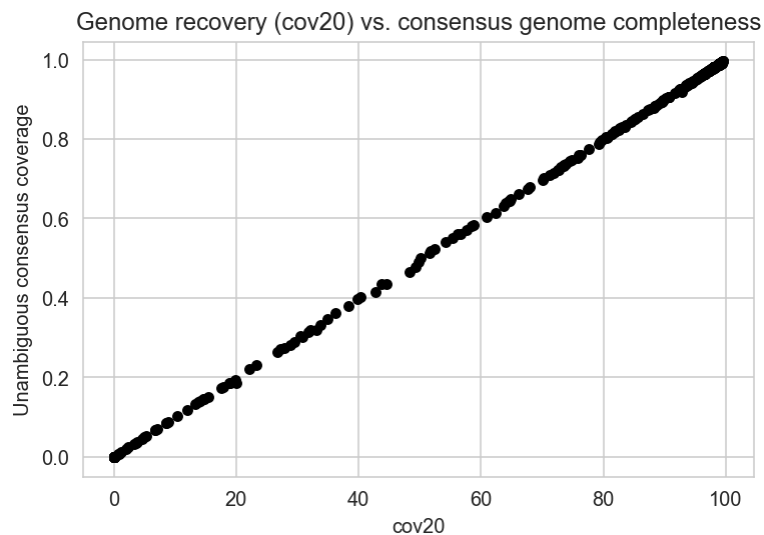
Fig. 5. Coverage defect maps of existing and revised Midnight and ARTIC protocols applied to Omicron lineage clinical samples. Top panels show the proportion of samples with defects by position while bottom panels show the per-sample distribution thereof, sorted by ascending ORF1 gene Ct. The additional primer sequence proposed by Freed and colleagues used in the Midnight v2 protocol effectively addresses the amplicon failure at ~28kb associated with the commercial kit still in widespread use at time of writing. For ARTIC v4.1, several new primers, some with revised coordinates, largely address the amplicon failures seen in Omicron lineage samples using the v4 primers.

Rapid turnaround multiplex sequencing of SARS-CoV-2: comparing tiling amplicon protocol performance

Supplementary figures



S. fig. 1. TaqPath RT-PCR ORF1ab cycle threshold distribution for sample sets 1-4.



S. fig. 2. Genome recovery vs. unambiguous consensus genome coverage. Pearson $R > 0.99999$, $p < 0.00001$.