TECH NOTE

Hybrid Capture Sequencing is More Robust than Amplicon Sequencing for Viral Surveillance

INTRODUCTION

Following its outbreak in late 2019, the SARS-CoV-2 virus contributed to hospitalizations and deaths globally, as well as disrupted the way of life for many. As the virus circulates within populations, novel mutations can lead to the emergence of more virulent strains. Close monitoring of the viral population allows researchers and health agencies to identify novel strains rapidly. Therefore, our ability to intercept and combat these strains is contingent on robust and accurate technologies for viral genome sequencing.

A common method for viral genome sequencing employs the polymerase chain reaction (PCR) to increase the concentration of the viral genome in a given sample before sequencing, known as amplicon sequencing. Its effective implementation relies on the precise design of primer sequences to target the amplification site of interest. Although cost-effective, amplicon sequencing can fail when sequencing a genome that contains mutations that are not present in the primer design. Mismatches between the primer and the genome lead to poor primer binding efficiency or total amplification failure, resulting in wide coverage variations and potential sequencing dropouts (Kuchinski et al., 2021).

In mid-2020, the ARTIC network designed the most widely implemented primer sets for the amplicon sequencing of SARS-CoV-2 (Tyson et al., 2020). Researchers have spiked in alternative versions of existing primers to circumvent the mismatch problem as the virus has evolved. ARTIC has also regularly updated their primer list (ARTIC network, 2021) . However, this iterative approach is costly concerning time and resources, given the pressing need to monitor variants as SARS-CoV-2 continues to mutate (Itokawa et al., 2020; Tyson et al., 2020). An alternative method that is robust to mutations in the SARS-CoV-2 genome is therefore desirable.

In answer to this need, Twist Bioscience has partnered with Biotia to release the SARS-CoV-2 NGS Assay - RUO, a highly sensitive nucleic acid hybridization capture-based assay for detecting SARS-CoV-2 RNA that is robust to mutations in the regions of interest. Additionally, It has previously been shown that target capture probes from Twist Bioscience are robust to mismatches (Twist Bioscience, 2019). Not only can this assay determine the presence or absence of the virus, but the software used can also detect genetic variants and lineages of the SARS-CoV-2 genome.

To demonstrate the advantages of hybrid capture, we implemented predictive modeling to assess areas in the SARS-CoV-2 genome that may be more prone to mutations that could impact primer efficiency in the multiplexing PCR step during amplicon sequencing. Our predictions identified four isolates as Variants of Concern (VoC), containing mutations within the binding sites for ARTIC primers. We then produced the synthetic genomes of these four VoC strains and used them to compare the performance of ARTIC amplicon sequencing with Twist Bioscience's SARS-CoV-2 NGS Assay - RUO. To understand how both methods perform on current VoCs, we also compared their performance on clinical SARS-CoV-2 samples.

Overall, our results show that hybrid capture led to fewer dropouts when sequencing a synthetic control of the Wuhan-1 variant (Twist Synthetic SARS-CoV-2 RNA Control 2; GenBank ID MN908947.3). Additionally, we observed dropouts in coverage of the predicted region for the four VoC synthetic genomes using amplicon sequencing but not for hybrid capture. Two of the four VoC synthetic genomes (EPI_ISL_1366445 and EPI_ISL_837547) had additional incidental dropout caused by a single mismatch in an ARTIC amplicon primer. Finally, we show that the Twist SARS-CoV-2 NGS Assay - RUO provides complete coverage of clinical VoC samples, whereas ARTIC amplicon sequencing presents dropouts in the S (encoding the spike protein) and ORF1ab genes. Our results demonstrate that hybrid capture is more robust to genomic variation than amplicon sequencing and leads to fewer dropout events. Therefore, we suggest that hybrid capture is a more robust and comprehensive solution for viral surveillance compared to alternatives.

METHODS

Design and Production of Synthetic SARS-CoV-2 Genomes

Twist Bioscience acquired 1,067,579 SARS-CoV-2 genomic sequences from GISAID (Elbe and Buckland-Merrett, 2017; dated 2021-04-21) in a multiple sequence alignment format and called variants using the UCSC faToVcf tool while designating the Wuhan isolate (EPI_ISL_402125) as the reference strain (Kent et al., 2002). Restricting to existing VoC lineages, a total of 6,777 unique alleles were observed (588,922 mutations total) in regions where ARTIC (v3) primers primers target. Viruses with mutations in the last six bases from the 3' end of ARTIC primers were selected, resulting in 101,432 isolates. Starting with these mutations, synthetic controls of four viral genomes were constructed to test if the mutations would lead to dropouts after amplicon sequencing (Table 1).

GISAID ACCESSION ID	EPI_ISL_1366445	EPI_ISL_1108224	EPI_ISL_837547	EPI_ISL_1540525
WHO LABEL	Epsilon	Alpha	Epsilon	Alpha
PANGO LINEAGE	B.1.429	B.1.1.7	B.1.429	B.1.1.7
COLLECTION DATE	2021-02-16	2021-02-08	2020-12-23	2021-02-25
LOCATION	California, USA	England, UK	Washington, USA	Prague, CR
ARTIC AMPLICON TESTED	24	41	73	87
MUTATED PRIMER	Left	Right	Left	Left
MUTATIONS	NC_045512.2:g.7056_7059CTGG>AAAA	NC_045512.2:g.12466_12468AGC>GAA	NC_045512.2:g.21984_21985del, NC_045512.2:g.21986_21988GTG>CAT	NC_045512.2:g.26214_26215del, NC_045512.2:g.26216_26217TG>AC
AMPLICON LOCATION	7059-7389	12134-12465	21991-22324	26220-26566
GENE	ORF1a	ORF1a	S	ORF3a, E, M
ARTIC AMPLICON COVERAGE OF AMPLICON INSERT: 10^4, 10^6	0.0, 1	57.8, 52.5	0.0, 2.5	0.5, 3
TWIST COVERAGE OF AMPLICON INSERT: 10^4, 10^6	137.6, 189.5	141.8, 247.3	74.5, 131	104, 195

Table 1. Sequencing Dropout of Select SARS-CoV-2 Variants. Comparison of the performance of amplicon sequencing and hybrid capture on four select SARS-CoV-2 VoC strains with mutations clustered near the 3' end of amplicon primers. We report the median depth of coverage for both 10^4 and 10^6 viral copies for positions covering the tested amplicon insert. Note that for sample EPI_ISL_1108224, the median coverage is significantly diminished for amplicon sequencing compared to hybrid capture even though we used the same number of reads and viral copies in the experiment.

SARS-CoV-2 Genome Sequencing

For the four synthetic genomes and an additional synthetic control genome of the Wuhan isolate (Twist Synthetic SARS-CoV-2 RNA Control 2), we carried out amplicon sequencing and hybrid capture sequencing using the ARTIC SARS-CoV-2 FS Library Prep Kit (protocol v3) and the Twist SARS-CoV-2 NGS Assay - RUO, respectively. We carried out sequencing on an Illumina NextSeq 500. Reads were trimmed to 74 bp and downsampled to 100,000 reads. Sequences were aligned to the NCBI reference sequence NC_045512.2 using the BWA-MEM algorithm while marking duplicates (Li, 2013; Wu et al., 2020). Since we constructed the viral synthetic genomes in 5kb fragments, we only considered primer pairs that fully spanned an individual fragment, resulting in the omission of 5–6 ARTIC primer pairs depending on the sample.

SARS-CoV-2 Nasopharyngeal Clinical Specimen Genome Sequencing by Biotia

Biotia collected nasopharyngeal specimens from the New York area that tested positive with FDA EUA authorized RT-PCR tests in reference clinical laboratories (Spring 2020–Summer 2021). Their specimens were prepared for amplicon sequencing and hybrid capture sequencing using the ARTIC SARS-CoV-2 FS Library Prep Kit (protocol v3) and the Twist SARS-CoV-2 NGS Assay - RUO, respectively. Biotia sequenced their prepared samples on an Illumina NextSeq 550 (Read counts: B.1.1.7: target capture - 4.4M; amplicon - 3.2M. B.1.617.2: target capture - 1M; amplicon - 3.4M. B.1.526.1: target capture - 4M; amplicon - 2.3M). The viral genomes were then profiled with the COVID-DX software (Biotia) to observe the depth and evenness of coverage and identify mutations across the genome.

RESULTS

Assessment of SARS-CoV-2 Genome Coverage

It was observed that hybrid capture produced fewer dropouts for the control SARS-CoV-2 genome. At a viral titer of 10^6, hybrid capture covered 100% of bases at 1X compared to 98.9% of bases at 1X for amplicon sequencing. At a lower titer of 10^4 viral copies, hybrid capture could still cover 100% of sequenceable bases in contrast to 98% of bases at 1X for amplicon sequencing (Figure 1). Although both methods can effectively cover most of the SARS-CoV-2 genome, surveillance of extant and emerging SARS-CoV-2 strains is contingent on variant calling, requiring high coverage for accurate interpretation. At a minimum coverage of 50X, hybrid capture covered 99.5% and 90.4% of the genome at 10^6 and 10^4 viral copies compared to 92.1% genome covered at 10^6 copies and 81.6% genome covered at 10^4 copies for amplicon sequencing.

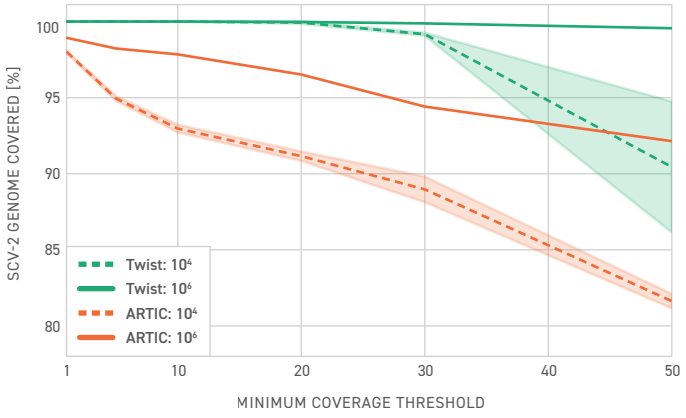

Figure 1. Twist's SARS-CoV-2 NGS Assay - RUO covers more of the SARS-CoV-2 genome than amplicon sequencing. The total percent of the SARS-CoV-2 genome covered by hybrid capture and amplicon sequencing was compared at 10^4 and 10^6 viral copies. Shaded regions represent one standard deviation from the mean for two replicates.

Assessment of the Effect of Mutations in Amplicon Primer Binding Sites on Sequencing Coverage

Sequencing dropouts were observed for every strain tested with ARTIC amplicon sequencing, whereas Twist's SARS-CoV-2 NGS Assay - RUO presented no dropouts. When comparing sequencing coverage to a non-variant positive control (Figure 2: ARTIC SARS-CoV-2 Control and Twist SARS-CoV-2 Control), no discernible difference was present in sequencing depth for either method. Therefore, the presence of these mutations is causing amplification failure in amplicon libraries. Such is the case for a strain from the Alpha/B1.1.7 lineage (Figure 2a), which has a near-complete dropout of sequencing due to the presence of a 3 bp substitution near the 3' end of the reverse primer for ARTIC amplicon 41. Although methods were limited to mismatches, deletions were present like the one found in EPI_ISL_1540525 (Figure 2b), where a 2 bp deletion coupled with a 2 bp substitution in the sense primer of amplicon 87 led to a dropout of 200 bp.

Surprisingly, strain EPI_ISL_1366445 (Epsilon/B.1.429) had both the expected dropouts caused by a 4 bp substitution (Figure 2c) and an additional incidental dropout caused by a 1 bp mismatch in the reverse primer for ARTIC amplicon 72 (Figure 2d). Similarly, another isolate from the Epsilon/B1.429 lineage had an equivalent incidental dropout caused by the same mutation (c.22018G>T; Figure 2e) in addition to the predicted dropout caused by a complex mutation of a 2 bp deletion adjacent to a 3 bp mismatch. These results demonstrate that hybrid capture is more resilient than amplicon sequencing to sequencing dropouts caused by mutations in probe targeting regions. Furthermore, observing that a 1 bp mismatch in an amplicon primer led to an incidental sequencing dropout highlights the severe limitation of amplicon sequencing for a virus with a high mutation rate.

To understand the impact sequencing dropouts may present in a clinical setting, Twist Bioscience collaborated with pathogen detection company Biotia, who employ both the Twist SARS-CoV-2 NGS Assay - RUO and an ARTIC Amplicon assay for the detection and characterization of SARS-CoV-2. Biotia obtained three nasopharyngeal samples presenting Alpha (B.1.1.7), Delta (B.1.617.2), and a sublineage of Iota (B.1.526.1, previously NYC) variants from patients in the New York City area. Biotia assessed percent genome coverage at 5X depth for each sample with both the Twist and ARTIC workflows. Due to mutations in the SARS-CoV-2 Genome falling within some target regions for primer binding, all three strains presented dropouts with amplicon sequencing. The Alpha variant presented three Spike protein dropouts (NC_045512.2:g.22895G>A, NC_045512.2:g.23012G>A, and NC_045512.2:g.23063A>T) caused by the NC_045512.2:g.22895G>A substitution falling within the reverse 75 primer binding sequence. Two dropouts were observed in the delta variant caused by mutations NC_045512.2:g.21618C>G and NC_045512.2:g.22029_22034del present within the forward 73 and reverse 72 primer binding sequences respectively. Dropouts in both the S (21682C>T) and in ORF1ab (11020C>T) genes were observed for the Iota variant as variants landed in the forward 72 and forward 37 primer binding sites respectively. As before, Twist's SARS-CoV-2 NGS Assay - RUO showed complete uniformity of coverage (>99%) across the viral genome.

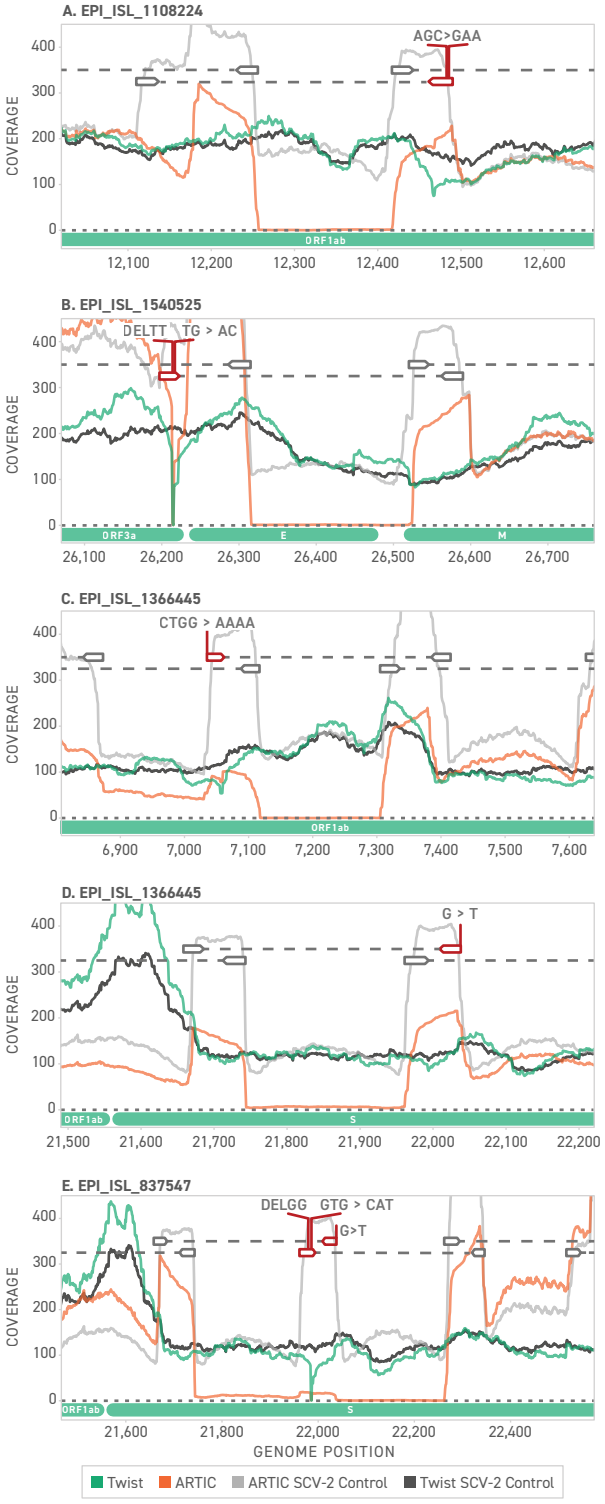

Figure 2. Mutations within Amplicon Primers can lead to Sequencing Dropout. Amplicon sequencing and hybrid capture were compared on four synthetic SARS-CoV-2 genomes listed in Table 1. Each strain had a mutation near the 3' end of an ARTIC PCR primer (shown in red). Sequencing was performed at 10^4 and 10^6 viral copies and included a positive control with no mutations (Twist Synthetic SARS-CoV-2 Control 2). All samples were processed in duplicate. Lines for Twist hybrid capture (green and dark gray) and ARTIC amplicon sequencing (orange and light gray) show the mean depth of coverage for all replicates and viral titers. ARTIC PCR primers are shown near the top of the figure in respective amplicon pools, while genes are depicted below.

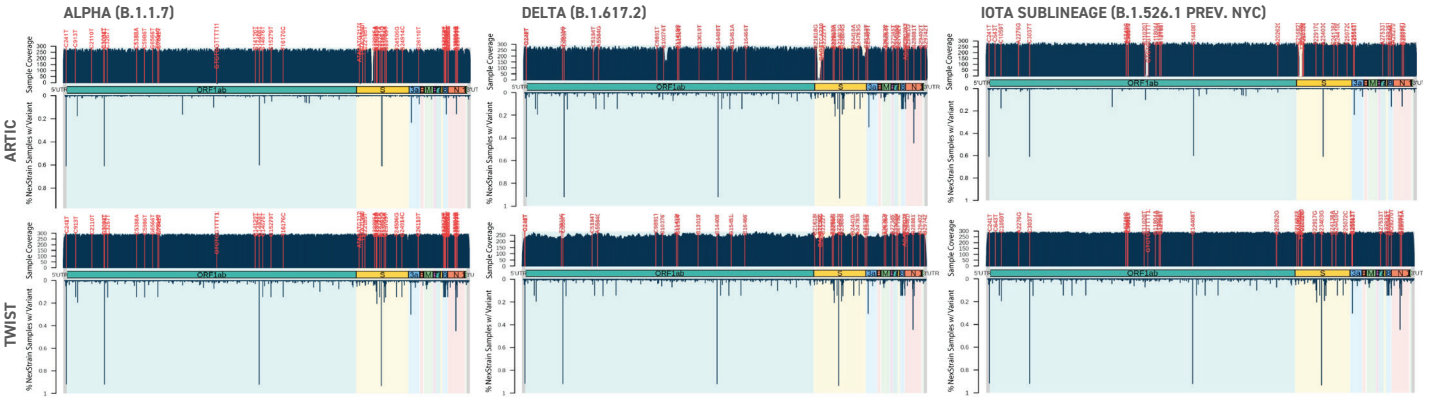

Figure 3. Amplicon Sequencing Leads to Dropouts in Clinical SARS-CoV-2 Samples, whereas Target Capture provides complete coverage. Detection of SARS-CoV-2 lineages Alpha (B.1.1.7; A), Delta (B.1.627.2; B), and Iota (B.1.526.1 variant; C) in three nasopharyngeal samples collected from the New York City area. The depth of sequencing recovered across the genome is plotted with detected mutations labeled in the top panels. The prevalence of each substitution across variants worldwide, as reported on NextStrain as of March 7, 2021, is shown in the bottom panels. Variants are only shown in genomic regions that have sufficient coverage.

CONCLUSION

Since the onset of the SARS-CoV-2 outbreak in 2019, researchers have been studying the evolution of the virus's genome. By monitoring how variations in the viral genome emerge and propagate throughout the population through genome sequencing, health providers around the globe can make more informed decisions about public health. Likewise, pharmaceutical research in treatments, cures, and vaccines for SARS-CoV-2 is contingent on ongoing genome surveillance. When this work was carried out the Alpha lineage was dominant in the United States, which has since been subsumed by the Delta lineage, demonstrating the pressing need to implement a sequencing method that is robust to genomic variation. In this study, we have demonstrated the benefits of hybrid capture for genome sequencing of SARS-CoV-2 over amplicon sequencing.

The results presented in this technical note are corroborated by many recent studies. One found a recurrent 168 bp deletion that was repeatedly missed using ARTIC primers and methods (Brandt et al., 2021). Another study published by Klempt et al. (2020) compared three methods for sequencing SARS-CoV-2 samples: target capture from Twist and Illumina, and amplicon sequencing from Paragon. Target capture methods presented greater uniformity in coverage, and lower false-positive rates compared to amplicon sequencing. Additionally, a recent study by Doddapaneni et al. (2021) showed that Twist Bioscience's target capture-based assay has further benefits in SARS-CoV-2 surveillance as it allows for the simultaneous target capture and quantitation of subgenomic coronavirus RNAs, which is not possible with amplicon sequencing.

Although amplicon sequencing provides a low-cost platform for viral surveillance, it comes at the added cost of sequencing dropout which can cause incorrect classification of viral lineages. This is further compounded by the certainty that mutations will accumulate in SARS-CoV-2, making it a matter of time until more primers fail. In conclusion, effective viral surveillance requires comprehensive genome sequencing as a reliable platform for the continuous monitoring of variants as they occur.

REFERENCES

Elbe, S. and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges, 1(1), pp. 33–46. https://doi.org/10.1002/gch2.1018.

Itokawa, K. et al. (2020). Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. PLOS ONE, 15(9), p. E0239403. https://doi.org/10.1371/journal.pone.0239403.

Kent, W. J. et al. (2002). The human genome browser at UCSC. Genome research, 12(6), pp. 996–1006. doi: https://doi.org/10.1101/gr.229102.

Kuchinski, K. S. et al. (2021). Mutations in emerging variant of concern lineages disrupt genomic sequencing of SARS-CoV-2 clinical specimens. medRxiv. https://doi.org/10.1101/2021.06.01.21258181.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. arXiv:1303.3997

Tyson, J. R. et al. (2020). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. bioRxiv, p. 2020.09.04.283077. https://doi.org/10.1101/2020.09.04.283077.

Wu, F. et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), pp. 265–269. https://doi.org/10.1038/s41586-020-2008-3.

Brandt, D. et al. (2021) 'Multiple Occurrences of a 168-Nucleotide Deletion in SARS-CoV-2 ORF8, Unnoticed by Standard Amplicon Sequencing and Variant Calling Pipelines', Viruses, 13(9). https://doi.org/10.3390/v13091870

Klempt, P. et al. (2020). Performance of Targeted Library Preparation Solutions for SARS-CoV-2 Whole Genome Analysis. Diagnostics. 2020; 10(10):769. https://doi.org/10.3390/diagnostics10100769

Doddapaneni, H. et al. (2021). Oligonucleotide capture sequencing of the SARS-CoV-2 genome and subgenomic fragments from COVID-19 individuals. PLoS ONE 16(8): e0244468. https://doi.org/10.1371/journal.pone.0244468

ARTIC Network (2021). SARS-CoV-2 version 4 scheme release https://community.artic.network/t/sars-cov-2-version-4-scheme-release/312. (Accessed October 19th 2021)

Twist Bioscience (2019). The Effects of mismatches on DNA Capture by Hybridization. https://www.twistbioscience.com/resources/white-paper/effects-mismatches-dna-capture-hybridization. (Accessed October 19th 2021).