Application Note

Genomics

Agilent

Trusted Answers

Whole Genome Sequencing of SARS-CoV-2 by Hybridization Capture Method

Authors

Daisuke Motooka,
Genome Information Research Center, Next-Generation Sequencing Core Facility, Research Institute for Microbial Diseases,
Osaka University

Yoshiyuki Ishii and
Fumiko Yoshizaki,
Diagnostics and Genomics Group,
Agilent Technologies, Inc.

Abstract

In this application note, we present a method for whole-genome sequencing of SARS-CoV-2 using the Agilent SureSelect XT Low Input kit and a custom NGS panel. Target enrichment libraries were prepared on the automated Agilent Bravo NGS workstation and sequenced on an Illumina MiSeq system, allowing us to obtain reads that covered the entire SARS-CoV-2 genome and enabled the detection of alpha, delta and omicron variants. As this hybridization-based method is capable of detecting a wide array of variants, we expect this approach to be able to detect multiple SARS-CoV-2 strains without further optimization.

Materials and Methods

Custom panel design

120 bp probes were designed against the SARS-CoV-2 reference genome (NCBI Accession: NC_045512.2) with 6x tiling density using a 20 bp sliding window. Probe designs were uploaded into the Agilent SureDesign software and subjected to the Optimized Performance XT HS/XT HS2/XT LI/QXT boost (90 min hyb) option to optimize probe counts for various GC% regions. The custom panel design information can be obtained from Agilent Technologies Japan upon request.

cDNA synthesis

First-strand cDNA was synthesized from 15 µL of extracted RNA using the ProtoScript II First Strand cDNA Synthesis kit (New England Biolabs, p/n E6560) and Random Primer (hexadeoxyribonucleotide mixture; pd (N)6) (Takara Bio, p/n 3801), with a 2.5-fold greater volume than that of the original protocol (total reaction volume of 50 µL). As RNA extraction was performed in outside labs using a variety of methods, input was based on volume rather than mass.

Second-strand cDNA synthesis was performed using the NEBNext Ultra II Non Directional RNA Second Strand Synthesis Module (New England Biolabs, p/n E6111).

For the second-strand synthesis, the entire volume of first-strand cDNA (50 µL) was used rather than the 20 µL described in the original protocol, instead reducing the volume of nucleic-acid-free water to adjust the total reaction volume to 80 µL. The entire second-strand cDNA synthesis reaction (80 µL) was mixed with 96 µL of Agencourt AMPure XP beads (Beckman Coulter, p/n A63880), purified, and eluted with 8 µL of nuclease-free water. The final concentration was determined using a Qubit 4 Fluorometer with the Qubit dsDNA HS Assay with 1 uL of eluted cDNA.

Library preparation and target enrichment using the SureSelect XT Low Input kit with enzymatic fragmentation

Seven µL of purified double-stranded cDNA was used for each library preparation regardless of yield. cDNA was fragmented using the Agilent SureSelect XT HS and XT Low Input Enzymatic Fragmentation kit according to the manufacturer's protocol. Library preparation was performed using the SureSelect XT Low Input kit with modified pre-capture PCR cycles (described in Table 1) for low-input samples.

In accordance with the protocol outlined in the application note, "*Utilization of Agilent SureSelect Target Enrichment for Whole Genome Sequencing of Viruses and Bacteria,*" 1,000 ng of adaptor-ligated libraries were used as input with a 1:10 dilution of capture probes and 21 cycles of post-capture PCR. Library preparation and target enrichment were performed on the Bravo Automated Liquid Handling system. Final libraries were analyzed with the High Sensitivity D1000 ScreenTape assay and the 4200 TapeStation system for size distribution. Quantification was performed on a Qubit 4 Fluorometer with Qubit dsDNA HS Assay. Libraries were mixed at equal masses and final concentrations of the pooled libraries were determined using KAPA Library Quantification kits (Roche, p/n 07960140001) on a PCRmax Eco 48 real time PCR system. Final pooled libraries were sequenced on a MiSeq instrument with 150 bp paired-end reads.

Data Analysis

After removing adaptor sequences with the Agilent Genomics NextGen Toolkit (AGeNT), reads were mapped to the reference genome NC_045512.2 using the Burrows-Wheeler Aligner (BWA-mem) package and indexed by the Picard toolkit after sorting and duplicate detection.

Results

Yields of double stranded cDNA and adaptor-ligated libraries

The yield of double-stranded cDNA varied from less than 0.5 to more than 200 ng. The synthesized cDNA was subjected to library preparation with a modified number of PCR cycles, resulting in sufficient yield for hybridization (1,000 ng) with all the samples (Table 1). A sample trace of these final libraries can be seen in Figure 1. All samples had sufficient yield for sequencing.

Table 1. Yield and pre-capture PCR cycles.

Sample	Ct Value	cDNA Yield (ng)	Pre-capture PCR Cycle	Adaptor-Ligated Library Yield (ng)	Post-Hybridization Final Library Yield (ng/µL)
1	27.6	5.306	16	1932	59.3
2	33.3	1.015	18	2028	61.3
3	26.7	0.742	18	1788	50.1
4	30.7	0.497	18	2052	49.1
5	29.6	3.304	16	1968	49.5
6	23.3	140.7	12	1212	61.5
7	22.6	46.69	12	1416	63
8	26.7	241.5	12	1236	20.2
9	22.6	15.47	16	1368	54.4
10	24.4	12.25	16	1488	51.6

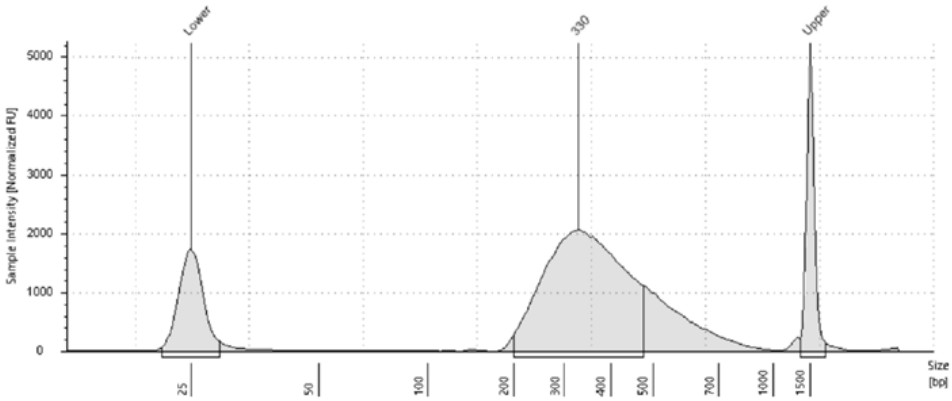

Figure 1. Example trace of a final library analyzed on the 4200 TapeStation system with the Agilent High-Sensitivity D1000 ScreenTape assay.

Sequencing results

In spite of the wide range of cDNA inputs and Ct values, the per-sample sequencing reads fell within a 2.3-fold range (Figure 2), indicating efficient use of sequencing capacity. Sample 2 had the highest Ct value (33.3) among these samples and was expected to have the lowest number of viral copies. However, while higher duplication percentage (72.7%, Figure 3) and lower average coverage (Figure 2) may have resulted from the lower viral copy number, this sample still achieved sufficient coverage for variant detection (>100X coverage at 99.8% of the total region). We achieved >95% on-target percentage with all the samples (Figure 3), indicating this target enrichment protocol enables efficient sequencing of a viral genome from samples containing human RNA.

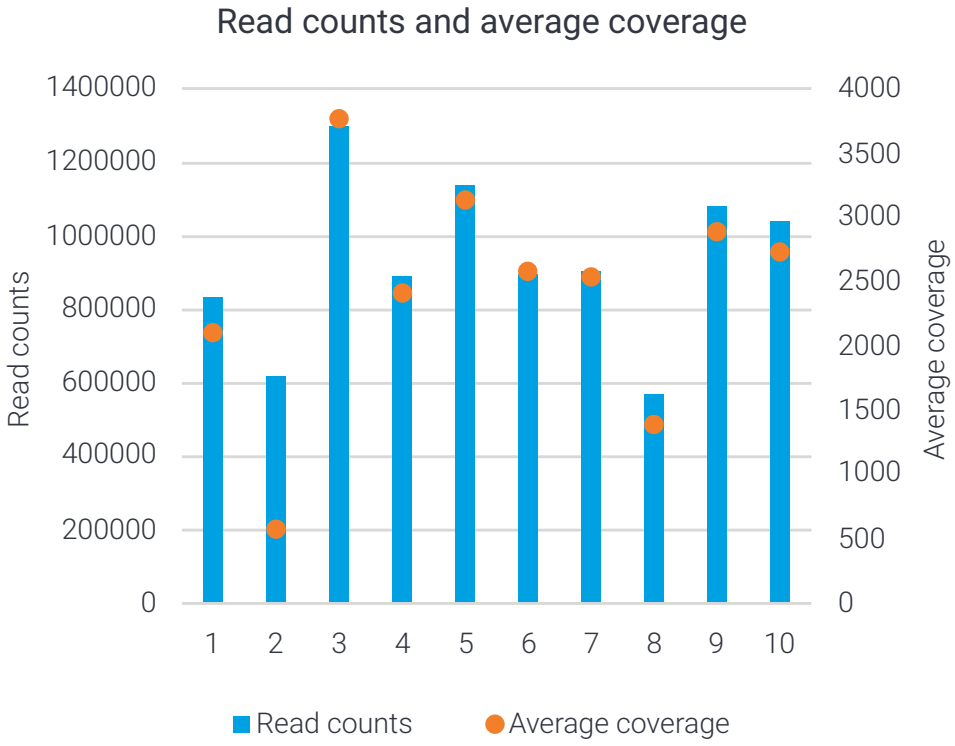

Figure 2. Read counts and average coverage of sequenced libraries. Read counts of each sample (left axis) and average coverage (right axis).

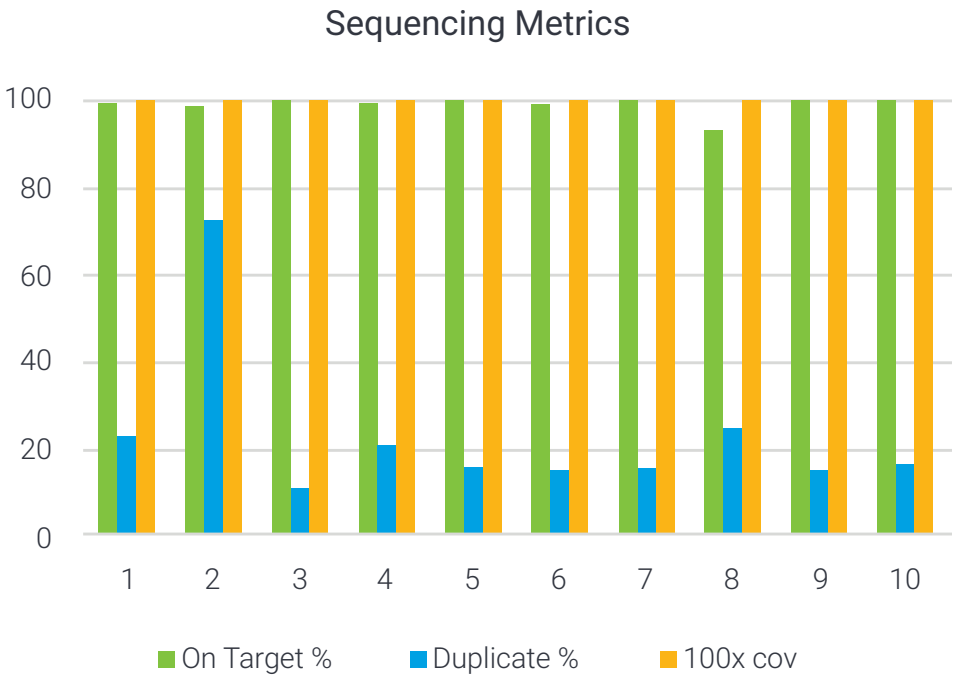

Figure 3. Sequencing metrics. The percentage of the reads mapped to the SARS-CoV-2 reference genome (On Target %), percentage of duplicated reads (Duplicate %), and percentage of regions with more than 100X coverage depth (100x cov) were calculated by the Picard software.

Sequencing reads covered the entirety of the SARS-CoV-2 genome (Figure 4) and characteristic variants, such as L452R for the delta variant and N501Y for the alpha variant, were detected (Figure 5).

High-throughput option

While we tried to standardize our protocol within a 96-well plate for more efficient handling, two parameters were fixed regardless of the input cDNA amount and pre-capture library yield: pre-capture PCR cycles were fixed at 18 cycles and hybridization input volume was fixed at 12 µL. Even with these fixed parameters among samples, high-quality and consistent results were obtained. We have already prepared more than 6,000 samples with successful detection including Omicron strain, and observed failure in less than 10 samples, all of which showed a Ct value around 35.

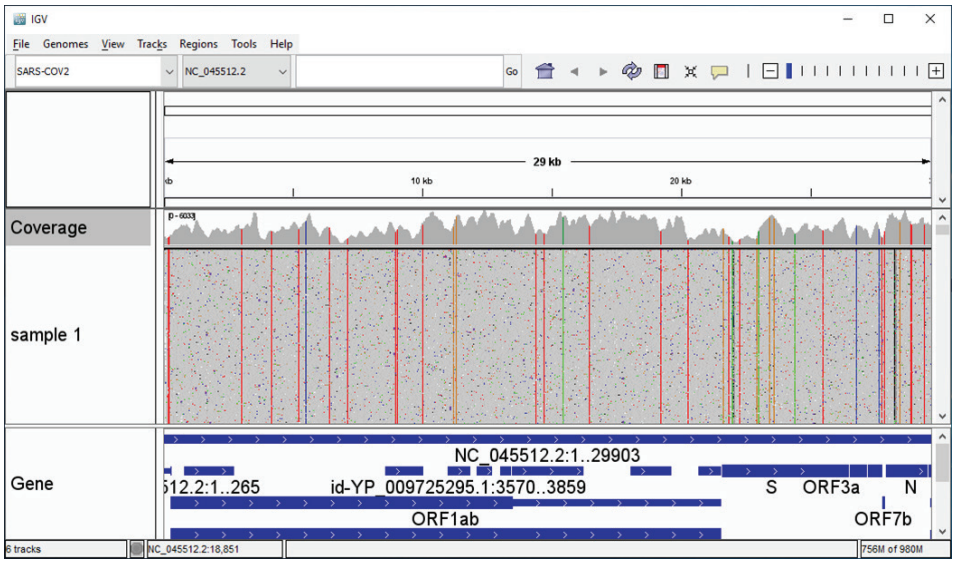

Figure 4. Representative coverage achieved by this protocol across the SARS-CoV-2 reference genome.

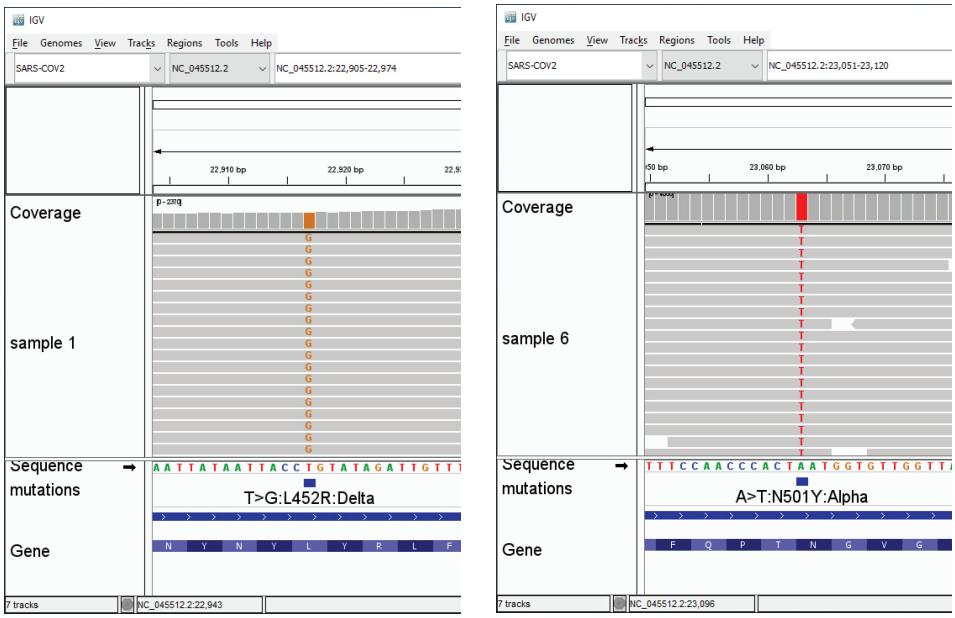

Figure 5. Examples of SARS-CoV-2 variants detected by this protocol. We were able to successfully identify the delta-strain characteristic variant L452R (Panel A) and the alpha strain characteristic variant N501Y (Panel B).

Conclusion

In this application note, we demonstrated that this targeted NGS protocol is robust and can provide accurate information on the SARS-CoV-2 genome across samples with a variety of Ct values and double-stranded cDNA yield. In addition, automation of this protocol on the Bravo system enables high-throughput analysis of the viral genome. Combined, this protocol helps address the growing demands for epidemiological surveillance and is expected to support basic research such as viral evolution and viral propagation paths.

www.agilent.com

For Research Use Only. Not for use in diagnostic procedures.
PR7000-3237

This information is subject to change without notice.

Published in the USA, February 25, 2022
5994-4538EN

Agilent
Trusted Answers