Surveillance of infectious disease through wastewater sequencing

Detect SARS-CoV-2 variants and other respiratory viruses in the community

- Wastewater sequencing can aid in tracking specific viral variants for COVID-19 epidemiology

- Illumina COVIDSeq™ Assay shows robust coverage of the spike protein gene using ARTIC v4 primers or ARTIC v3 primers

- Respiratory Virus Oligo Panel v2 detects SARS-CoV-2 plus additional respiratory viruses

illumına®

For Research Use Only. Not for use in diagnostic procedures.

M-GL-00429 v1.0 | 1

Introduction

Monitoring the spread and evolution of SARS-CoV-2 aids public health efforts to counter the ongoing COVID-19 pandemic. One surveillance method that offers complementary information to clinical case data is wastewater-based epidemiology. The SARS-CoV-2 virus is shed in the stool of infected individuals and appears in samples from wastewater treatment plants. Detecting this viral RNA in wastewater offers a way to track infections at a community scale. Wastewater data can also predict surges of infections up to two weeks before clinical cases present in the health care system.[1] Early warning from wastewater data can prompt local action to prevent the spread of disease and prepare hospitals for imminent surges.

Various methods exist for examining RNA in wastewater, including reverse transcription quantitative polymerase chain reaction (RT-qPCR) and next-generation sequencing (NGS) of RNA (RNA-Seq). While RT-qPCR can detect the virus, RNA-Seq provides the most detail to discern the presence of specific viral lineages. This application note compares the performance of targeted amplicon- or enrichment-based RNA-Seq library preparation methods for detecting SARS-CoV-2 variants in wastewater samples.

Considerations for wastewater sequencing

In contrast with clinical samples that usually contain a single pathogen, wastewater is a mixed sample that may contain multiple viral lineages and other microbes. To capture only sequences of interest, amplicon- or enrichment-based library preparation methods are recommended.

Amplicon sequencing involves analyzing genomic regions of interest with deep sequencing of PCR amplicons. The Illumina COVIDSeq Assay with ARTIC v3 primers or updated ARTIC v4 primer pools is an amplicon-based NGS assay that can identify and characterize SARS-CoV-2.[2,3]

Target enrichment sequencing captures genomic regions of interest via hybridization to target-specific probes. Illumina RNA Prep with Enrichment and the Respiratory Virus Oligo Panel v2 (RVOP) allow targeted sequencing of the SARS-CoV-2 genome plus other common respiratory viruses in the same assay.[4,5]

Here we demonstrate how the COVIDSeq and RVOP enrichment assays can be adopted for wastewater surveillance of SARS-CoV-2 variants in a population.

Wastewater comes with additional challenges, compared to standard clinical samples. Nucleic acids in wastewater are degraded, with an average fragment length < 200 bp. Wastewater also contains chemical inhibitors. Samples often show a very low viral titer, as measured by a cycle threshold (Ct) value > 30.

To compensate for these factors, we increased the input RNA and the number of reads per sample by 10-fold, where possible. We also reduced the consensus sequence generation threshold for data analysis. For clinical samples, submission of sequences to the GISAID* EpiCoV repository requires over 90% non-N bases with at least 10× coverage for 90 out of 99 SARS-CoV-2 genome amplicons.[6] For the purposes of this study, we lowered the threshold to at least 10× coverage for 50 out of 99 amplicons. This allowed data generation for low-titer samples that showed coverage just below the GISAID standard, yet high enough to detect SARS-CoV-2 and call specific variants present in the community.

Methods

Sample Preparation

Biobot Analytics, a wastewater epidemiology company, collected samples from various wastewater treatment plants in the US in spring 2021 ("BB"-labeled samples, received by Illumina on June 22, 2021) and in fall 2021 ("E"- or "F"-labeled samples, received by Illumina on October 12, 2021) (Table 1). To concentrate the samples, Biobot loaded 45 ml of wastewater into an Amicon Ultra-15 centrifugal ultrafiltration unit (Millipore, Catalog no. UFC903096) three times and then eluted samples in 100 µl nuclease-free water.[7,8] RNA was extracted using the QIAGEN RNeasy method. Biobot reported Ct values for each sample based on RT-qPCR analysis (Table 1). DNA complementary to the RNA (cDNA) was generated by reverse transcriptase with random hexamers.

---

* GISAID, Global Initiative on Sharing Avian Influenza Data

Table 1: Ct values for wastewater samples provided by Biobot Analytics

	Samples from spring 2021							Samples from fall 2021					
Sample	BB13	BB3	BB2	BB7	BB9	BB1	BB16	F1	F3	E8	F9	F5	F7
Ct value	30.69	31.24	31.76	33.14	33.71	34.81	34.86	30.01	30.42	30.53	30.59	30.78	30.88

Library Preparation

Sequencing-ready libraries were prepared starting with cDNA from concentrated wastewater samples using one of three different methods: COVIDSeq Assay (Illumina, Catalog no. 20049393 and 20051772) with ARTIC v3 primer pools, COVIDSeq Assay with ARTIC v4 primer pools (Illumina, Catalog no. 20065135), or Illumina RNA Prep with Enrichment (Illumina, Catalog no. 20040536) and the Respiratory Virus Oligo Panel v2 (Illumina, Catalog no. 20044311).

Sequencing

Prepared COVIDSeq libraries were sequenced on the MiSeq™ System at 2 × 150 bp read length using the MiSeq Reagent Kit v3 (Illumina, Catalog no. MS-102-3003) at a depth of 1 million reads per sample. Prepared RVOP libraries were sequenced on the NextSeq™ 550 System at 2 × 150 bp read length using the NextSeq 500/550 High Output Kit v2.5 (Illumina, Catalog no. 20024908) at a depth of 10 million reads per sample. Viral titer, RNA quality, and the number of reads per sample impact the number of virus-specific reads obtained and coverage of the virus genome.

Data Analysis

FASTQ sequencing files were input to Illumina DRAGEN™ pipelines (accessible in BaseSpace™ Sequence Hub) or kallisto for analysis.

Analysis with DRAGEN COVID Lineage App

For COVIDSeq and RVOP enrichment libraries, FASTQ sequencing files from the MiSeq or NextSeq 550 system were input to the DRAGEN COVID Lineage App for alignment to a SARS-CoV-2 reference genome. The app performs mapping/alignment, variant calling, and consensus sequence generation.

For DRAGEN COVID Lineage analysis, default parameters were used except for the minimum alignment score (set at 22 vs 12) and the consensus sequence generation threshold. The alignment score was increased to reduce spurious alignment given the sequencing read configuration used. The consensus threshold was reduced to increase sensitivity of low-titer samples as mentioned previously.

Analysis with DRAGEN RNA Pathogen Detection App

For RVOP enrichment libraries, FASTQ sequencing files from the NextSeq 550 System were input to the DRAGEN RNA Pathogen Detection pipeline for analysis and viral detection with default parameters. The app performs analysis including alignment, variant calling, and protein-encoding transcript level detection of viruses, which increases the ability to identify novel and highly divergent viruses.

Analysis with kallisto

Reads aligned to the SARS-CoV-2 sequence by the DRAGEN COVID Lineage App were used as an input into a kallisto pipeline, developed by Baaijens et al, to estimate SARS-CoV-2 variant abundance in wastewater, following steps described at github.com/baymlab/wastewater_analysis.[8,9] The SARS-CoV-2 reference set (required by kallisto pipeline) was built from US GISAID sequences up to September 29, 2021. WHO[†] variants were defined according to the CDC[†] site cdc.gov/coronavirus/2019-ncov/variants/variant-info.html as of December 1, 2021.

---

† WHO, World Health Organization;
   CDC, Centers for Disease Control and Prevention

Results

Sequencing libraries from wastewater samples were prepared using COVIDSeq with ARTIC v3 primer pools, COVIDSeq with ARTIC v4 primer pools, or RVOP enrichment. SARS-CoV-2 sequence data was analyzed with the DRAGEN COVID Lineage App. Wastewater samples with Ct values close to 30 yielded better sequencing results than samples with lower viral titers (Table 1, Figure 1).

The COVIDSeq Assay with ARTIC v3 or ARTIC v4 primer pools generated better genome coverage of SARS-CoV-2 than did the RVOP enrichment method (Figure 1, Figure 2). The updated ARTIC v4 primer pools show more consistent coverage across the spike protein gene compared to the ARTIC v3 primer pools (Figure 2).

The relative abundance of SARS-CoV-2 variants in wastewater samples was estimated using a specialized kallisto data analysis pipeline (Figure 3).[8,9] Samples collected in
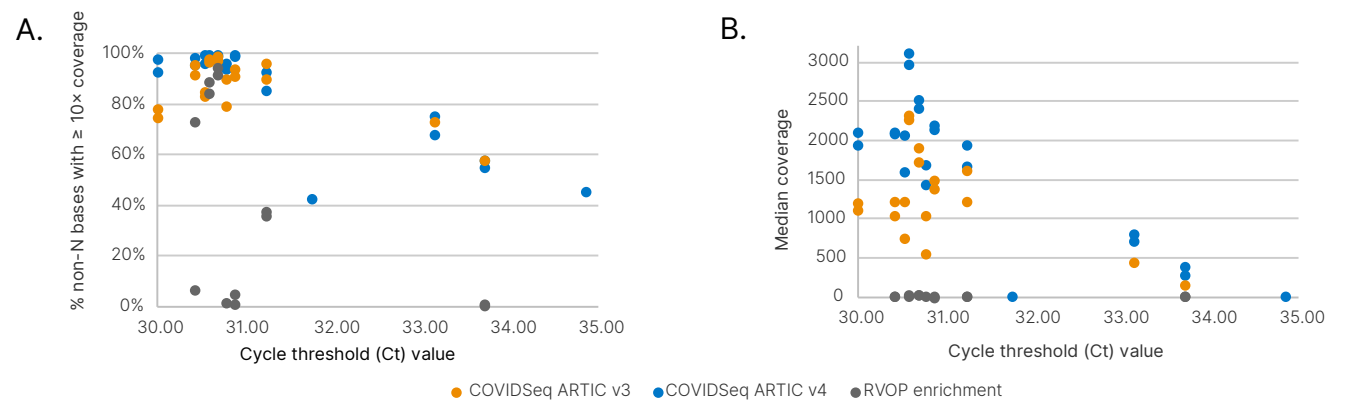
Figure 1: Viral titer affects genome coverage for SARS-CoV-2 in wastewater samples—Sequencing libraries from wastewater samples were prepared using COVIDSeq with ARTIC v3 primer pools (orange), COVIDSeq with ARTIC v4 primer pools (blue), or RVOP enrichment (gray) and data analyzed with the DRAGEN COVID Lineage App. (A) Percent non-N bases with at least 10× coverage and (B) median coverage of SARS-CoV-2 viral genome, compared to viral titer (Ct values).
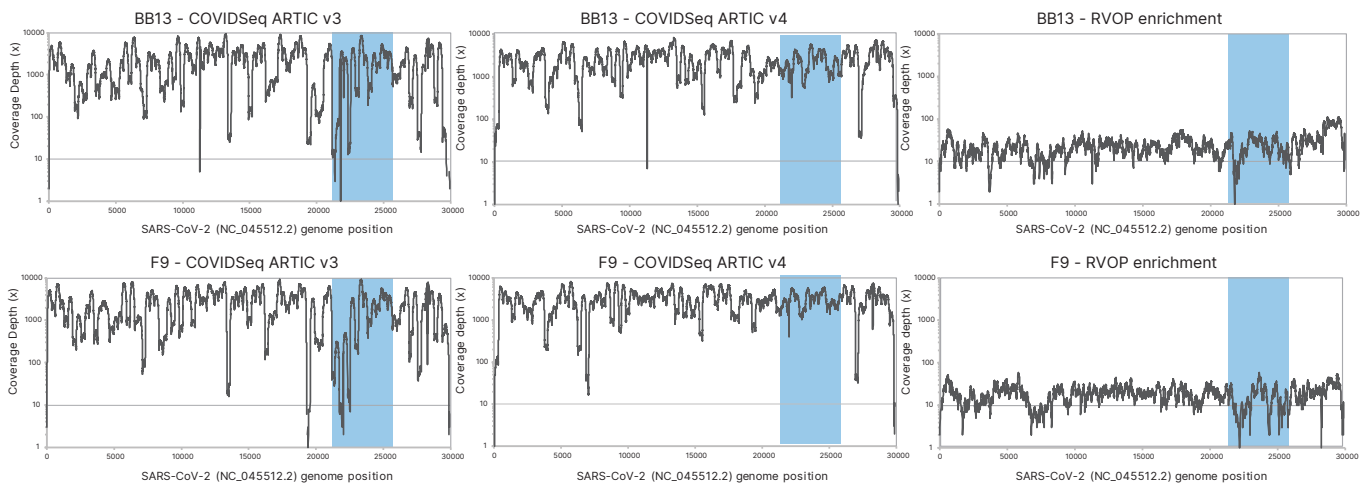

Figure 2: COVIDSeq Assay generates high coverage across SARS-CoV-2 genome in wastewater samples—Coverage plot comparison between COVIDSeq Assay with ARTIC v3 (left) and ARTIC v4 (middle) primer pools and RVOP enrichment (right) for wastewater sample received in June (BB13, Ct 30.69; top) and sample received in October (F9, Ct 30.59; bottom). Data analyzed with the DRAGEN COVID Lineage App. ARTIC v4 primer pools produce the most consistent coverage across the spike protein gene (blue).

spring 2021 show a predominance of the alpha variant across the population, while samples collected in fall 2021 demonstrate prevalence of the delta variant. This data correlates with the clinical strains observed during the time of sampling.

Even with lower coverage of SARS-CoV-2, the RVOP enrichment method and the DRAGEN RNA Pathogen Detection App identified additional respiratory viruses present in the wastewater samples (Figure 4). This information can also be critical for management of public health in local communities.
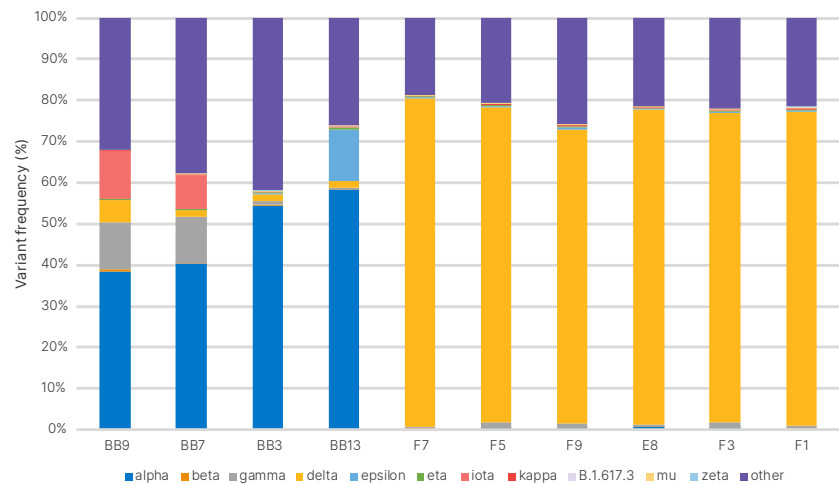
Figure 3: kallisto detects distinct SARS-CoV-2 lineages between wastewater samples collected at different times—To estimate the relative abundance of SARS-CoV-2 variants (as defined by WHO and CDC) in wastewater samples, SARS-CoV-2 reads from the COVIDSeq Assay with ARTIC v4 primer pools were fed into the kallisto pipeline developed by Baaijens et al (github.com/baymlab/wastewater_analysis).[8,9] Samples from spring 2021 (BB9, BB7, BB3, BB13) show large representation of alpha variant (bright blue); samples from fall 2021 (F7, F5, F9, E8, F3, F1) show dominance of delta variant (yellow).
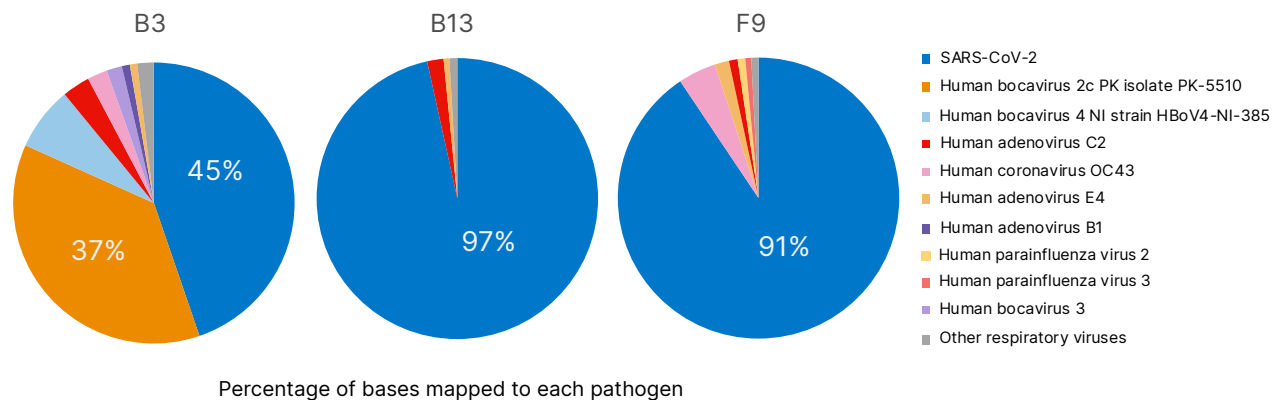

Percentage of bases mapped to each pathogen

Figure 4: RVOP enrichment detects various viruses in wastewater samples—Sequencing libraries from wastewater samples were prepared using RVOP enrichment, sequenced at 10 million reads per sample, and data analyzed with the DRAGEN RNA Pathogen Detection App. Different wastewater samples show different abundance of various respiratory pathogens. Samples with high SARS-CoV-2 titers (BB13, Ct 30.69 and F9, Ct 30.59) show predominance of that virus. Samples with lower SARS-CoV-2 titers (like BB3, Ct 31.24) may be dominated by other viruses, which affects genome coverage for COVID-19 sequencing.

Summary

More public health labs are adopting wastewater sequencing to detect SARS-CoV-2 variants circulating in the population, independently of patient-based surveillance. This application note demonstrates an RNA-to-report solution for library preparation, sequencing, and analysis of wastewater samples that can be used to monitor the COVID-19 pandemic and other emerging infectious diseases at a community level. The COVIDSeq Assay with ARTIC v4 or ARTIC v3 primer pools offers sensitive detection of SARS-CoV-2 variants in wastewater samples, including their relative prevalence in the population. Wastewater sequencing using Illumina RNA Prep with Enrichment and Respiratory Virus Oligo Panel v2 allows simultaneous, broad-range detection of SARS-CoV-2 plus additional important pathogens. Illumina is committed to providing solutions to support the needs of public health efforts.

Learn more

NGS methods for infectious disease detection, illumina.com/areas-of-interest/microbiology/infectious-disease-surveillance.html

COVIDSeq Test (RUO Version), illumina.com/products/by-type/clinical-research-products/covidseq.html

COVIDSeq Assay, illumina.com/products/by-type/clinical-research-products/covidseq-assay.html

Respiratory Virus Oligos Panel, illumina.com/products/by-type/sequencing-kits/library-prep-kits/respiratory-virus-oligo-panel.html

Illumina RNA Prep with Enrichment, illumina.com/products/by-type/sequencing-kits/library-prep-kits/rna-prep-enrichment.html

Ordering information

Product	Catalog no.
COVIDSeq Assay (96 samples) index 1	20049393
COVIDSeq Assay (96 samples) index 2	20051772
COVIDSeq Positive Control	20051775
Illumina COVIDSeq v4 Primer Pools RUO (384 samples)	20065135
Respiratory Virus Oligos Panel V2	20044311
Illumina RNA Prep with Enrichment, (L) Tagmentation (16 samples)	20040536
Illumina RNA Prep with Enrichment, (L) Tagmentation (96 samples)	20040537
IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 indexes, 96 samples)	20027213
IDT for Illumina DNA/RNA UD Indexes Set B, Tagmentation (96 indexes, 96 samples)	20027214

References

1. Crits-Christoph A, Kantor RS, Olm MR, et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio*. 2021;12(1):e02703-20. doi:10.1128/mBio.02703-20
2. Illumina. Illumina COVIDSeq Test (Research Use Only) Data Sheet. illumina.com/content/dam/illumina-marketing/documents/products/datasheets/illumina-covidseq-test-data-sheet-1270-2020-008.pdf. Published 2020. Accessed December 6, 2021.
3. Illumina. Sequencing guidelines for COVID-19 surveillance using the Illumina COVIDSeq Test (RUO Version). illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-covidseq-ruo-read-length-depth-tech-note-m-gl-00088/illumina-covidseq-ruo-read-length-depth-tech-note-m-gl-00088.pdf. Published 2021. Accessed December 6, 2021.

4. Illumina. Detection and characterization of respiratory viruses, including SARS-CoV-2, using Illumina RNA Prep with Enrichment. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-rna-enrichment-coronavirus-app-note-470-2020-005/illumina-rna-enrichment-coronavirus-app-note-470-2020-005.pdf. Published 2020. Accessed December 6, 2021.

5. Illumina. Faster detection of respiratory viruses using the MiniSeq™ Rapid Reagent Kit and Illumina RNA Prep with Enrichment. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/miniseq-rapid-covid-app-note-470-2020-008/miniseq-rapid-covid-app-note-470-2020-008.pdf. Published 2020. Accessed December 6, 2021.

6. Association of Public Health Laboratories (APHL). Recommendations for SARS-CoV-2 Sequence Data Quality & Reporting. aphl.org/programs/preparedness/Crisis-Management/Documents/APHL-SARS-CoV-2-Sequencing.pdf. Accessed February 2, 2022.

7. Duvallet C, Wu F, McElroy KA, et al. Nationwide trends in COVID-19 cases and SARS-CoV-2 wastewater concentrations in the United States. *medRxiv*. 2021.09.08.21263283. doi:10.1101/2021.09.08.21263283

8. Baaijens JA, Zulli A, Ott IM, et al. Variant abundance estimation for SARS-CoV-2 in wastewater using RNA-Seq quantification. *medRxiv*. 2021.08.31.21262938. doi:10.1101/2021.08.31.21262938

9. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525-527. doi:10.1038/nbt.3519

illumına®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com