Rapid, high-frequency SARS-CoV-2 wastewater sequencing with the Genexus™ Integrated Sequencer

Jingwei Ni[1], Rob Knight[2,3], Smruthi Karthikeyan[2], David Merrill[1]

1. Thermo Fisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA 94080. 2. Department of Pediatrics, University of California San Diego, La Jolla, CA 92093. 3. Department of Bioengineering, Department of Computer Science & Engineering, and Center for Microbiome Innovation, University of California San Diego, La Jolla, CA 92093

Abstract

Purpose: Quick wastewater SARS-CoV-2 mixed lineage deconvolution

Methods: Full-genome sequencing with Genexus™ Integrated Sequencer and variant analysis using Freyja tool

Results: Sample to results (strain info and abundance) with a one-day turn around time

Introduction

Environmental wastewater surveillance is becoming a powerful and minimally invasive genomic sequencing tool to monitor the emergence and spread of SARS-CoV-2 variants such as Delta and Omicron. A key advantage of wastewater sequencing is its ability to potentially provide early, comprehensive surveillance data within a community, especially with frequent sampling and sequencing. Rapid sequence data availability offers the potential for earlier public health messaging, guidance, or interventions to help decrease the severity of outbreaks. However, current approaches are hindered by pooled sampling and long sequencing turnaround times, which can delay results. Wastewater sequencing can also be impacted by poor sample quality due to degradation, low viral titers, and mixed samples for analysis.

To address these challenges, we report a protocol for SARS-CoV-2 wastewater surveillance using the Ion AmpliSeq™ SARS-CoV-2 Insight Research Assay on the Genexus™ Integrated Sequencer. Data were obtained by sequencing archival wastewater samples from University of California, San Diego (UCSD). We achieved >95% base coverage at 20X for samples with Ct values up to 35 with a one-day turnaround time. Using a sample de-mixing tool, Freyja, (https://www.medrxiv.org/content/10.1101/2021.12.21.21268143v1), we identified mixed strains matching expected percentages, with sensitivity as low as 1%. The automated workflow of the Genexus™ Integrated Sequencer enables quick turnaround times of one day with minimal hands-on time, which is ideal for high-frequency SARS-CoV-2 surveillance with wastewater samples.

Materials and methods

Sample preparation and sequencing

Archival wastewater samples from UCSD, previously quantified using qPCR, up to 25ul each were loaded onto the Genexus™ sequencer, in conjunction with the Ion AmpliSeq™ SARS-CoV-2 Insight Research Assay, following the protocol described in the Ion AmpliSeq™ SARS-CoV-2 Insight Research Assay – GX User Guide. Automated library preparation and full-genome sequencing were performed by the Genexus™ sequencer. Up to 16 samples were loaded on the same run, and the results were reported in about one day.

Data analysis

For QC purposes, coverage statistics for each sample were downloaded from the SARS_CoV_2_coverageAnalysis plugin output. A processed, primer-trimmed BAM file for each sample was downloaded from the Genexus™ web interface and manually run through Freyja software (v1.3.4) to deconvolute mixed strain lineages and relative abundances. Summarized variants of concern/variants of interest (VOC/VOI) based on World Health Organization (WHO) designation from outbreak.info were reported in a curated metadata file.

Freyja is an open-source tool developed by Dr. Joshua Levy in Dr. Kristian Andersen's lab at Scripps Research, it is freely available from GitHub: https://github.com/andersen-lab/Freyja.

Results

More than half wastewater samples produced sufficient coverage for lineage analysis

11 out of 20 archival wastewater samples with a Ct range of 28-37 produced sufficient coverage for strain lineage identification based on > 20X target base coverage. Fresh samples are expected to obtain better sequencing coverage than archived samples.

Table 1. Sample QC stats from SARS_CoV_2_coverageAnalysis plugin.

	Ct Value	Mapped Reads	Average reads per amplicon	Percent base reads on target	Uniformity of base coverage	Target base coverage at 20X
S23_StrainMix	n.a.	2,004,870	8,447	100%	95%	**100%**
S24_StrainMix	n.a.	5,773,132	24,307	100%	95%	**100%**
S59	34.49	527,436	2,033	99%	81%	**99%**
S61	34.22	725,098	2,913	100%	79%	**99%**
S62	33.95	1,345,700	5,605	100%	86%	**100%**
S63	34.42	1,658,537	6900	100%	85%	**100%**
S64	33.28	1,474,932	6,127	100%	92%	**100%**
S66	33.47	1,644,400	6,834	100%	90%	**100%**
S67	34.31	1,148,331	4,717	100%	87%	**100%**
S68	32.91	384,006	1,310	97%	75%	**95%**
S71	37.76	452,087	1,648	98%	80%	**94%**
S74	28.06	243,268	857.6	97%	75%	**96%**
S78	33.65	786,910	3237	100%	84%	**99%**

Sample control mixtures with known strains produced expected lineages and abundance info

A mixture of RNA samples containing 5 different SARS-CoV-2 strains (Alpha, Beta, Gamma, Delta and A) with different known ratios were sequenced on Genexus™ in replicate and produced expected, largely consistent lineage and abundance using Freyja.

Figure 1. Lineage breakdowns for mixed strain samples

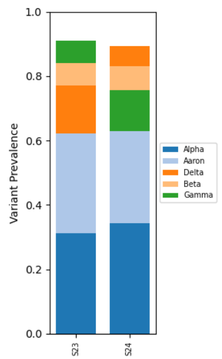
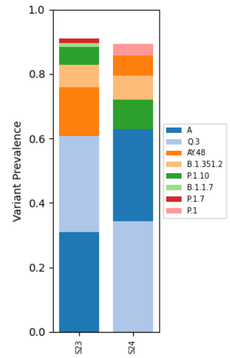


Figure 2. Summarized lineage abundance for mixed strain samples



Mixed lineage abundances are reported from wastewater samples

Since the samples were taken during the Delta outbreak period, most of the mixtures are different strains of Delta with a couple of exceptions.

Table 2. Strain and lineage for wastewater samples for Freyja output

	Summarized	Lineages	Abundances
S59	[('Delta', 0.934)]	['AY.25' 'AY.39' 'AY.100' 'AY.103' 'AY.44' 'AY.128']	[0.390 0.195 0.152 0.110 0.075 0.011]
S61	[('Delta', 0.907)]	['AY.100' 'AY.3' 'AY.44' 'AY.39' 'AY.122' 'AY.25.1' 'AY.20' 'AY.119.2' 'AY.65' 'AY.3.1' 'AY.3.3' 'AY.24' 'AY.4' 'AY.26' 'B.1.617.2']	[0.129 0.098 0.094 0.089 0.087 0.065 0.062 0.050 0.046 0.046 0.035 0.029 0.026 0.024 0.019]
S62	[('Delta', 0.887)]	['AY.103' 'AY.44' 'AY.100' 'AY.3' 'AY.25' 'AY.39' 'AY.20' 'AY.25.1' 'AY.4.9' 'AY.113' 'AY.3.1' 'AY.43' 'AY.119' 'AY.126' 'AY.32' 'AY.4.2' 'AY.119.2' 'AY.127']	[0.247 0.201 0.069 0.044 0.041 0.034 0.031 0.029 0.027 0.026 0.023 0.019 0.019 0.017 0.017 0.012 0.011 0.010]
S63	[('Delta', 0.916)]	['AY.25' 'AY.44' 'AY.100' 'AY.3' 'AY.119.2' 'AY.113' 'AY.119' 'AY.79' 'AY.39' 'AY.25.1' 'AY.26' 'AY.39.1' 'AY.124' 'AY.20' 'AY.103']	[0.322 0.154 0.093 0.071 0.045 0.041 0.028 0.028 0.027 0.019 0.018 0.017 0.017 0.017 0.013]
S64	[('Delta', 0.885)]	['AY.44' 'AY.103' 'AY.25' 'AY.100' 'AY.118' 'AY.125' 'AY.98' 'AY.3.1' 'AY.119' 'AY.25.1' 'AY.122.1' 'AY.126' 'AY.4' 'AY.75' 'AY.39']	[0.154 0.150 0.114 0.112 0.073 0.072 0.043 0.035 0.030 0.020 0.016 0.016 0.015 0.014 0.014]
S66	[('Delta', 0.831)]	['AY.25' 'AY.103' 'AY.44' 'AY.119' 'AY.100' 'AY.20' 'AY.125' 'AY.3' 'AY.25.1' 'AY.39' 'AY.117' 'AY.122' 'AY.119.2' 'AY.113' 'AY.124' 'AY.92' 'AY.3.1']	[0.153 0.150 0.143 0.062 0.060 0.039 0.034 0.030 0.026 0.025 0.018 0.017 0.016 0.015 0.014 0.011 0.010]
S67	[('Delta', 0.945)]	['AY.44' 'AY.103']	[0.561 0.383]
S68	[('Delta', 0.930)]	['AY.25' 'AY.122' 'AY.119' 'AY.113' 'AY.44' 'AY.95']	[0.307 0.217 0.181 0.165 0.045 0.013]
S71	[('Delta', 0.971)]	['AY.103' 'AY.126']	[0.960 0.010]
S74	[('Delta', 0.981)]	['AY.103']	[0.981]
S78	[('Gamma', 0.975)]	['P.1.12']	[0.975]

Conclusions

- Wastewater extracted RNA sample-to-results full-genome sequencing was achieved in one day using the Genexus ™ Integrated Sequencer. The system was easy to set up with only 10 min of hands-on time.
- Freyja is easy to use, and it can produce accurate lineages and abundances for mixed SARS-CoV-2 samples.
- The combination of Genexus™ and Freyja enables regular, high-frequency SARS-CoV-2 surveillance of wastewater samples.

References

1. Wastewater sequencing uncovers early, cryptic SARS-CoV-2 variant transmission: https://www.medrxiv.org/content/10.1101/2021.12.21.21268143v1

Trademarks/licensing