CPHM-SATURDAY-279

Xuan Xu[1], William Hull[1], David Plunkett[1], Zheng Jin Tu[1], Sima Mortazavi[2], Praneetha Avutala[2], Justin Ng[2], Ramin Khaksar[2], Daniel D. Rhoads[1], Hannah Wang[1]

[1] Department of Pathology and Laboratory Medicine, Cleveland Clinic, Cleveland, OH, USA [2] Clear Labs, Inc., San Carlos, CA, USA

Balancing Automation and Accuracy: A Comparative Study of SARS-CoV-2 Sequencing Workflows

Xuan Xu
Cleveland Clinic
xux2@ccf.org

Introduction

- Next-generation sequencing (NGS) is pivotal for SARS-CoV-2 genomic surveillance, especially with the emergence of new Omicron sublineages.
- We developed a SARS-CoV-2 whole-genome sequencing (WGS) protocol using ARTIC v5.3.2 primer set on Illumina MiSeq platform and an in-house bioinformatics pipeline.
- Separately, Clear Dx SARS-CoV-2 WGS v3.0 workflow was developed by Clear Labs, using Midnight Native primers on Oxford Nanopore Technologies (ONT) MinION platform, and it offers a fully automated solution.
- This study evaluates the performance of these two workflows, using 30 SARS-CoV-2 (Omicron) positive specimens with varying viral loads (Cycle threshold values 12-30) tested in parallel to assess their capabilities across a broad range of sample qualities.

Methods

- SARS-CoV-2 positive nasopharyngeal swabs were collected from clinical laboratories in Ohio between 2022 and 2024.
- A Panther Fusion qRT-PCR assay targeting the *ORF1ab* gene had been previously performed to identify SARS-CoV-2 in clinical routine in our laboratory.
- A total of 30 samples with Ct ranging from 12 to 30 were included in this study. Viral RNA was extracted using the KingFisher Flex Viral RNA Extraction system.
- Long-read sequencing libraries were generated and sequenced on ONT MinION platform using the Clear Dx SARS-CoV-2 WGS v3.0 workflow by Clear Labs.
- For the short-read sequencing, ARTIC v5.3.2 tiled amplicons were generated using the ARTIC v5.3.2 primer set, and libraries were prepared with the Illumina COVIDSeq Assay. Sequencing was performed on Illumina MiSeq platform with reagent kit v2 using 300 cycles with paired-end reads at Cleveland Clinic (CCF).
- Data analysis was carried out independently by each laboratory. Viral genomes were assembled, and consensus sequences were generated using each respective pipeline. Quality control (QC) metrics, lineage assignments, and mutation characterizations were performed with Nextclade. The acceptance criteria for Clear Labs workflow requires > 95% genome coverage with a minimum depth of 20x. In comparison, CCF workflow used a threshold of > 90% genome coverage with > 10x depth, and includes additional criteria: average coverage ≥ 100x, mean quality score ≥ 30, Q30 rate ≥ 75%, S gene coverage ≥ 85%, and manual frameshift mutation review.

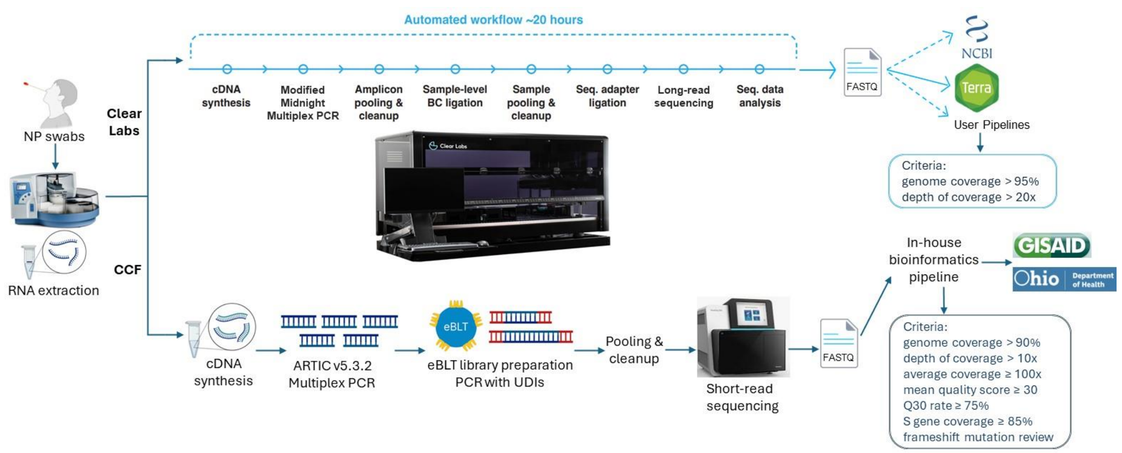

Figure 1 Schematic overview of the sequencing workflows compared. SARS-CoV-2 positive samples were extracted and sequenced using CCF workflow on the Illumina MiSeq and Clear Labs workflow on the ONT MinION platform.

Results

- Both the Clear Labs and CCF workflows demonstrated comparable performance in generating high-quality SARS-CoV-2 consensus genomes.
- More than 90% genome coverage was achieved in 93% of samples sequenced by Clear Labs and 90% of samples sequenced by CCF. Median genome coverage was similar (Clear Labs: 99.2%, IQR 99%-99.2%; CCF: 99.7%, IQR 94.8%-99.7%).
- Clear Labs workflow generated higher sequencing depth, with a median depth of 2298x (IQR: 1752-2797x), compared to CCF workflow, which had a median depth of 1583x (IQR: 1021-2198).
- Three samples failed QC in CCF workflows, all with <90% genome coverage. Two of these also failed in Clear Labs workflow due to low genome coverage. The third sample passed QC in Clear Labs workflow (99% genome coverage) but was flagged by Nextclade for an incorrect frameshift call in the S gene. Another sample failed QC in Clear Labs workflow (90.4% genome coverage) but passed in CCF workflow (91.3% genome coverage), reflecting differences in QC thresholds.
- Among the 27 samples that passed QC in both workflows, two were assigned different SARS-CoV-2 lineages depending on the pipeline used. Manual review of BAM files revealed that a key mutation (ORF1a:T170I) was missed in the Clear Labs analysis, resulting in an incorrect lineage assignment.

Sample ID	Ct	% Genome coverage		Depth of coverage		QC.overall score[a]		Lineage	
		Clear Labs	CCF	Clear Labs	CCF	Clear Labs	CCF	Clear Labs	CCF
63	18.9	99.2	99.7	2027x	1617x	0	0	BA.5.1.22	BA.5.1.22
64	22.9	99.3	98.8	2111x	1015x	0	0	BF.4	BF.4
65	15.7	99.3	99.7	2417x	2236x	0	0	BE.1.2.1	BE.1.2.1
66	20.7	99.3	99.8	2355x	1041x	0	0	BF.10	BF.10
67	20.9	99.2	99.7	1896x	1841x	70.3	0.2	BA.5.2	BA.5.2
68	21	99.3	94.8	1902x	933x	0	18.1	BA.4.6	BA.4.6
69	15.8	99.3	99.7	2547x	2422x	0	0	BA.5.1.7	BA.5.1.7
70	24.1	98.9	98.9	887x	2494x	0	0	BA.4.6	BA.4.6
71	23.6	99.3	99.7	2551x	2086x	0	0	BA.5.2.6	BA.5.2.6
72	29	99.1	92.3	1750x	1125x	0	49.3	XBB.1.5.13	XBB.1.5.13
73	27.6	97.1	92.9	1339x	832x	4.3	40.7	BE.10	BE.10
74	22.9	99.2	99.8	2810x	2639x	0	0.2	BN.1.7	BN.1.7
75	19.7	99.3	99.7	2746x	551x	4.3	4.3	XBF	XBF
76	20.6	99.2	99.7	2370x	1949x	0	0	BQ.1	BQ.1
77[b]	27.7	75.6	89.9	416x	905x	674.5	153.1	JN.1.8.1	JN.1.8.1
78	14.5	99.2	99.7	2241x	2428x	0	0	JN.1.7	JN.1.7
79	13.7	99.3	99.8	1856x	1879x	0	0	JN.1.7	JN.1.7
80	18.6	99.3	99.7	2760x	925x	0.2	0	JN.1.7	JN.1.7
81	22.1	99.3	97.5	3003x	1549x	0	6.2	JN.1.8.1	JN.1.8.1
82[b]	26.3	43.5	73.2	114x	453x	3858.7	854.6	JN.1.18	JN.1
83	26.7	96.0	94.4	1737x	1383x	10.8	27.1	JN.1.4	JN.1.4
84[e]	17.8	99.2	99.9	1758x	1950x	0	0	JN.1.18	KQ.1
85[e]	24.4	99.4	94.8	3215x	1414x	1.6	29.6	JN.1.18	KQ.1
86	19.9	99.2	99.7	3007x	2415x	0	0	KQ.1	KQ.1
87	25.8	99.2	96.5	1746x	1395x	0	8.8	JN.1	JN.1
88	15.1	98.4	99.7	3423x	3280x	3.3	6.3	JN.1.18	JN.1.18
89	12.3	99.2	99.8	3672x	1966x	0	0	JN.1.7	JN.1.7
90[c]	30.2	90.4	91.3	607x	1001x	91.6	72.3	JN.1.13.1	JN.1.13.1
91[d]	22.5	99.0	88.0	3358x	1229x	156.3	165.3	KP.2	KP.2
92	16.5	99.2	99.8	3174x	2572x	0	0	LB.1.2	LB.1.2

Table 1. Comparison of genome coverage, sequencing depth, quality control scores (as indicated by "qc.overallScore" according to Nextclade), and lineage assignments of both ONT MinION (Clear Labs) and Illumina MiSeq (CCF) platforms.

Note:
[a] Quality scores are interpreted as follows: 0-29 = good, 30-99 = mediocre, and ≥ 100 = poor quality.
[b] Samples that failed QC in both workflows are highlighted in yellow.
[c] Sample that failed QC in Clear Labs workflow but passed in CCF workflow is highlighted in green.
[d] Sample that failed QC in CCF workflow but passed in Clear Labs workflow, with an incorrect frameshift call, is highlighted in orange.
[e] Samples with discrepant lineage assignments between the two workflows are highlighted in blue.

Conclusions

- Both ONT MinION- and Illumina MiSeq-based WGS workflows – implemented by Clear Labs and CCF – are effective for SARS-CoV-2 genomic surveillance, generating high-coverage sequences across a range of sample qualities.
- CCF's workflow demonstrated slightly superior consensus genome quality and lineage assignment accuracy, while Clear Labs offers full automation with minimal hands-on time, making it labor-efficient despite processing fewer samples per batch (32 vs. 96 for CCF). However, the per-sample consumable cost for Clear Labs platform is 3- to 5-fold higher than that of CCF workflow without accounting for labor overhead.
- These findings provide valuable insights into workflow selection based on laboratory resources, throughput needs, and surveillance goals.