Clear Labs

WHITE PAPER

The Clear Dx™ Rapid, Fully Automated Whole Genome Sequencing Platform for SARS-CoV-2 Detection and Genomic Surveillance

Whole Genome Sequencing Plays a Critical Role in the Ongoing COVID-19 Pandemic

Coronavirus disease 2019 (COVID-19) is a severe respiratory disease that was initially identified in Wuhan, China, but has become a global pandemic. The virus that causes COVID-19 is designated as Severe Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). As of February 2022, there are over 437 million cases and close to 6 million deaths worldwide due to COVID-19. The pandemic created an immediate need for genomic characterization tools to determine not only that SARS-CoV-2 was present, but also to track how the virus was transmitted in the human population, how it impacted human health in different demographics and how the changes in the genetic sequence impacted viral spread and transmissibility. Thus, whole genome sequencing (WGS) quickly became the tool of choice that was essential for understanding the functional and epidemiological consequences of SARS-CoV-2.

The evolution of SARS-CoV-2 has been unpredictable. Factors such as inequitable global vaccine distribution, long-haul COVID among immunocompromised patients and possible transmission between humans and other mammals, have contributed to the rapid increase in the number of mutations. The continued emergence and spread of new variants reinforce the critical role that genomic sequencing has in the enhanced surveillance of COVID-19. In order to keep pace with an ever-evolving virus, it is paramount to employ rapid and reliable genomic surveillance tools that track viral evolution and its impact on strategies such as diagnostic tests, vaccines, etc, that are aimed at curbing the pandemic.

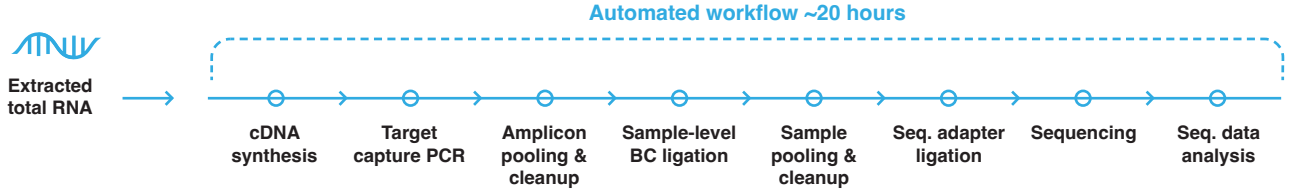
A Turnkey Automated WGS Workflow Solution to Enhance Genomic Surveillance

While WGS is a powerful tool, manual WGS workflows are time-consuming, labor-intensive and require specialized expertise for analysis which severely limits the practical application of WGS in many clinical and public health laboratories (PHLs), a major roadblock to the adoption of this technology for routine use in the laboratory. As a response to this, Clear Labs developed a fully automated, targeted next-generation sequencing based platform as a turnkey solution for WGS of the SARS-CoV-2 virus. The Clear Dx: WGS SARS-CoV-2 assay (RUO)[1] can process extracted RNA for up to 32 clinical samples in one run with a sample-to-report turnaround time of 20 hours. The platform does this by automating PCR, library preparation, sequencing and bioinformatic consensus genome creation (see Figure 1).

The Clear Dx platform is custom-built with thermocyclers and a MinION sequencer from Oxford Nanopore Technologies (ONT). The system is controlled using customized instrumentation

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

Figure 1: Clear Dx: WGS SARS-CoV-2 assay workflow

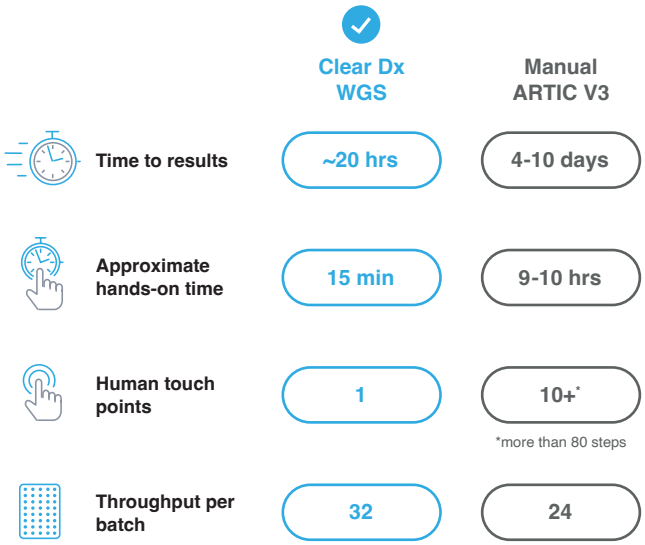


software and housed inside a Hamilton STAR liquid handling robot. The Clear Dx: WGS SARS-CoV-2 sequencing workflow leverages the Clear Labs modified Midnight primer panel with 29 primer pairs distributed into 2 pools, tiled across the entire genome to capture and amplify nearly 99.4% of the SARS-CoV-2 genome with amplicons ranging about 1200bp in length. The long amplicons from each sample are ligated with unique barcodes after a post-amplification clean up, pooled together and sequenced in a single MinION flow cell. The fewer target capture primers and consequently fewer primer binding regions reduce the possibility of future erosion in target capture performance due to mutations in the primer binding regions in emerging SARS-CoV-2 variants. Barcode classification of sequencing reads and assembly of the consensus genome sequences are done using a modified version of the ARTIC bioinformatics pipeline, optimized for Midnight primers.

Target capture through PCR amplification used in this platform enables high specificity; efficient, fast, and near-complete SARS-CoV-2 genome capture; and abundant and efficient use of sequencing read depth. Since the amplicon library is sequenced directly without further processing, the size distribution of the sequenced library is very tight, leading to a high percentage of sequencing reads passing read length filters. These features produce high quality genome sequences in the form of FASTA files for accurate lineage classification, submission to public databases such as GISAID (the Global Initiative for Sharing Avian Influenza Data) and NCBI (the National Center for Biotechnology Information), and for identification of mutations necessary for viral research and vaccine development.

The fully automated Clear Dx turnkey system reduces hands-on time from 9 hours to less than 0.5 hours and narrows the number of human touch points from more than 10 steps down to just one, decreasing turnaround time from sample receipt to reporting from greater than 4 days to less than 24 hours (see Figure 2).

Figure 2: Clear Dx: WGS system versus the manual sequencing process



Clear Dx: WGS SARS-CoV-2 Performance

Methods

Synthetic SARS-CoV-2 RNA controls procured from Twist Biosciences along with 977 real-world SARS-CoV-2 samples of various lineages obtained from several public health and reference lab partners were used for the purpose of evaluating the performance of the modified Midnight primer set on the Clear Dx system.

Results

Stock solution of synthetic SARS-CoV-2 RNA controls was serially diluted from 10,000 copies/µL to 1 copy/µL[2] and the resultant RNA samples were processed on the Clear Dx automated assay. As shown in Table 1, the Clear Dx: WGS assay has high sensitivity producing greater than 95% genome coverage[3] at >800x mean sequencing depth at 10

2 17ul was utilized per sequencing reaction.

3 Genome coverage is normalized to the maximum theoretical coverage possible with the synthetic Twist RNA controls that have known gaps in the sequence.

Table 1: Analytical performance of the Clear Dx: WGS SARS-CoV-2 assay

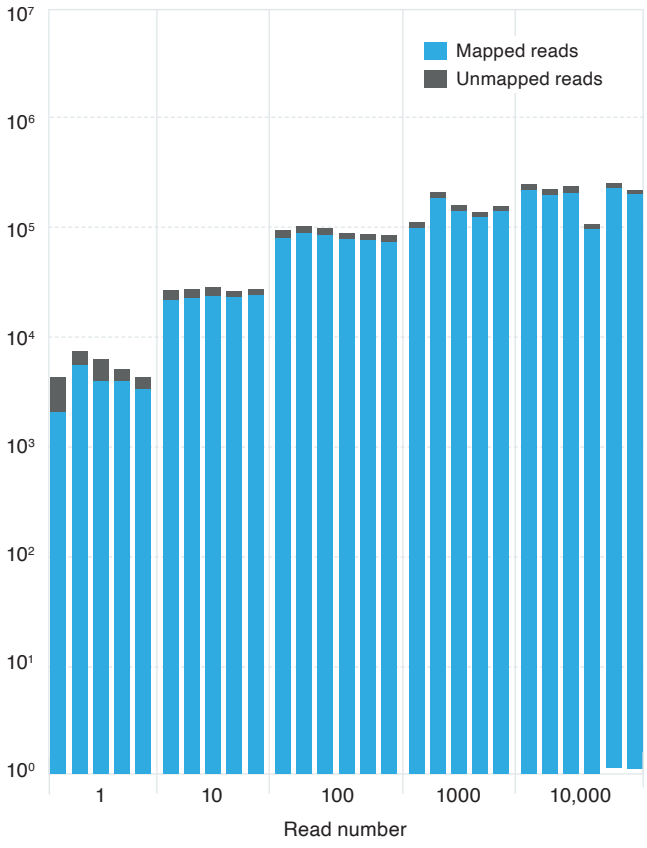
Concentration (copy/ul)	Mean adj. genome coverage*	Mean sequencing depth
10,000	98.74%	6511
1000	98.21%	4399
100	97.98%	2601
10	96.27%	862
1	68.78%	143

* Genome coverage is normalized to the maximum theoretical coverage possible with the synthetic Twist RNA controls that have known gaps in the sequence.

genome copies/µL. In addition, as indicated in Figure 3, there are greater than 1000 number of reads generated down to 1 copy/µL and on average, greater than 80% of the reads mapping successfully to the SARS-CoV-2 genome. This means that some of the SARS-CoV-2 virus genome regions and signature mutations may be covered even at 1 copy/µL and accurate lineage identification may be still possible at this very low concentration.

A similar study with 977 real-world COVID-19 samples including all SARS-CoV-2 variants ranging from Alpha to Omicron were assessed utilizing the Clear Dx platform. The results, as shown in Figure 4, illustrate that the Clear Dx: WGS SARS-CoV-2 assay produces greater than 95% mean genome coverage for the SARS-CoV-2 present in clinical samples with a wide range of viral load, with a cycle-threshold (CT) ranging from <15 up to 30. In addition, the Clear Dx: WGS SARS-CoV-2 assay, utilizing the Clear Labs modified[4] Midnight primer panel, produces highly uniform sequencing read coverage at high depth across all regions of SARS-CoV-2 genome for samples with the aforesaid range of CT values as displayed in Figure 5.

Figure 3: Read number and mapped reads for the Clear Dx WGS SARS-CoV-2 assay



4 Clear Labs modified Midnight primer panel uses additional equimolar amounts of modified sequence of primer 28 (as shown below) going right-to-left in pool 2 to account for the mutation in the 3' binding region of p28 leading to dropout of this amplicon in some cases of Omicron noticed earlier by a few in the community. SARS CoV_1200_28_LEFT_27837T: TTTGTGCTTTTTAGCCTTTCTGtT.

Figure 4: CT value and genome coverage correlation

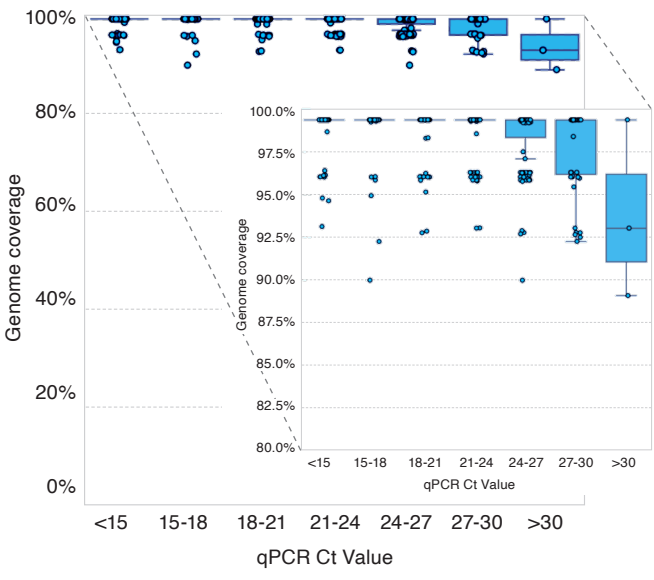
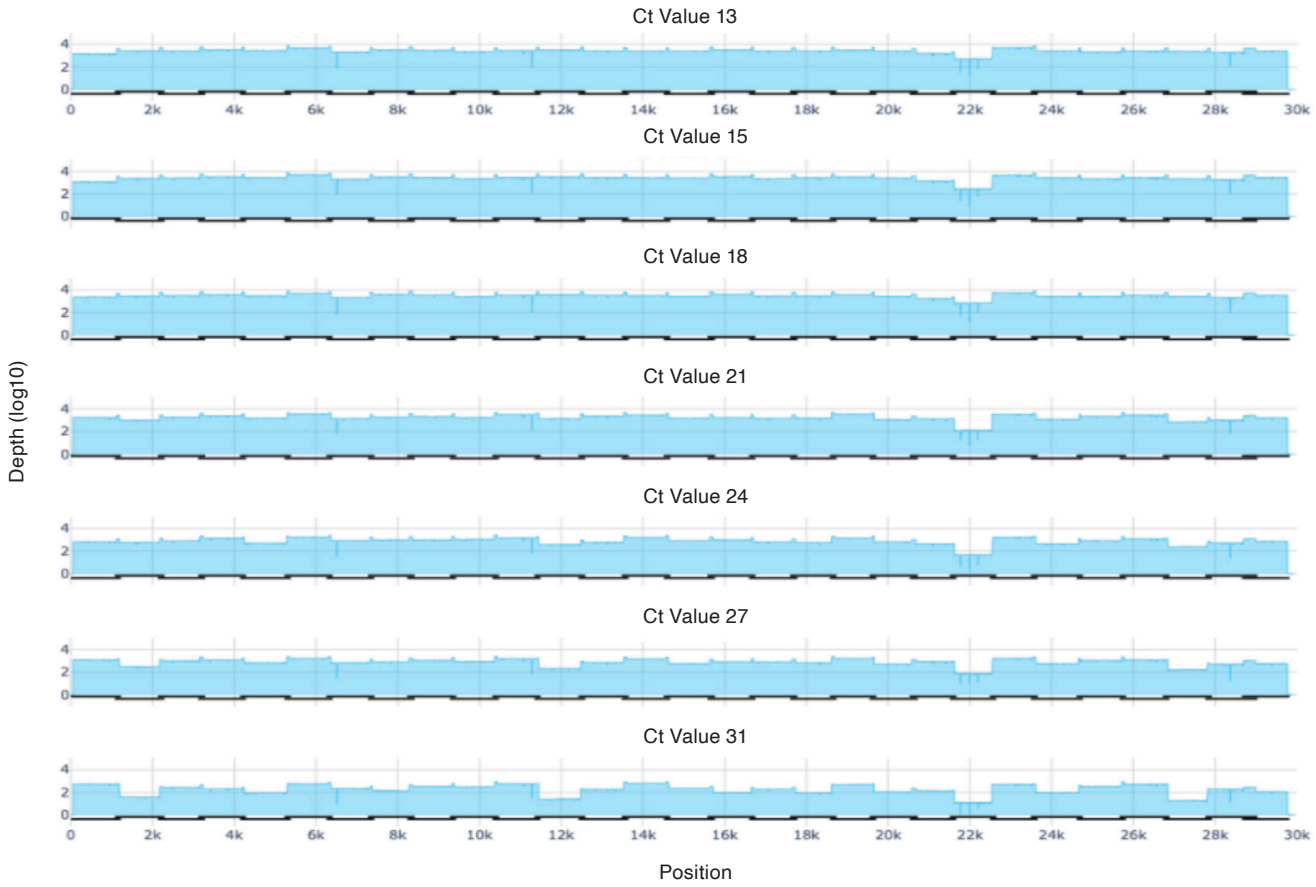


Figure 5: SARS-CoV-2 sequencing depth versus CT



The Clear Dx: SARS-CoV-2 WGS assay has been able to accurately characterize Variants of Interest (VOI) and Variants of Concern (VOC) since the detection of the Alpha variant. Of the 977 real-world samples assessed, none had any fatal Viral Annotation DefineR (VADR) alerts, thus allowing for successful submission to GISAID. Table 2 reports the performance of the assay in terms of accuracy of lineage calls and VADR alerts encountered for the different variants.

Lastly, multiple different extraction platforms were analyzed as part of this assessment as well. Seven different extraction platforms from vendors such as the KingFisher Flex, Qiagen, Roche, Perkin Elmer and Apostle provided comparable sequence data in terms of genome coverage and sequencing depth

Table 2: Real-world samples with non-fatal VADR alerts

Variant	Sample size	% of samples with correct lineage call	% of samples with VADR=0
Omicron	494	100%	100%
Delta	384	100%	99.48%
Epsilon	8	100%	100%
Alpha	13	100%	100%
Zeta	1	100%	100%
Gamma	11	100%	100%
Mu	3	100%	100%
Lamda	6	100%	100%
Other*	57	100%	100%

*Includes: B.1.2, B.1.234, B.1.243, B.1.436, B.1.628, A.2.5, A.21, B.1, B.1.1, B.1.1.135, B.1.1.192, B.1.1.222, B.1.1.318, B.1.1.322, B.1.1.519, B.1.2, B.1.234, B.1.361, B.1.369, B.1.401, B.1.415.1, B.1.541, B.1.561, B.1.565, B.1.568, B.1.575, R.1, XB, B.1.637.

(data not shown here) ensuring that the Clear Dx system is compatible with various different extraction methods.

Benefits

There are numerous benefits of utilizing the Clear Dx platform for WGS of SARS-CoV-2 samples:

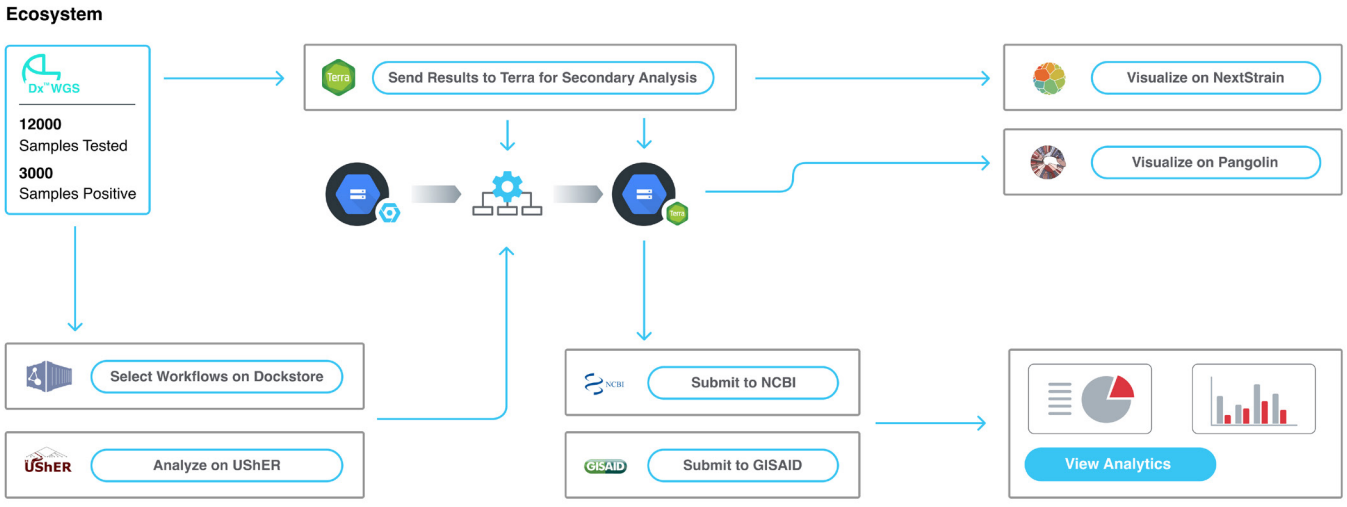
1. Compatibility with a broad range of RNA extraction platforms
2. No need for quantification of the amount of library going into the sequencing
3. Minimal personnel time requirement for sequencing set-up
4. Less than 24-hr turnaround time for genome assembly results, and
5. Ease of integration of genome assemblies for downstream phylogenetic analysis. These benefits will help to ensure that the laboratory can keep up with the demanding needs for SARS-CoV-2 during the pandemic.

Cloud Connectivity and Data Visualization

Clear Labs has also developed proprietary Cloud-based solutions for data storage. To address platform security, Clear Dx combines Google Cloud Platform (GCP) encryption methods such as Encryption-at-Rest using AES-256 bit, Encryption-in-Transit using HTTPS over TLS1.2, along with internal testing procedures. Together, these methods provide a robust and secure cloud genomics solution.

The Cloud-based connectivity provided by the Clear Dx platform is optimized to consolidate big genomic data by automating upload of sequencing data to analytical platforms such as the Terra platform for further analysis and phylogenetic mapping as seen in Figure 6. Thus, being able to leverage connected data across multiple systems offers timely, accurate and unique data that can inform vital outbreak decision-making for implementation of real-time public health countermeasures.

Figure 6: Clear Dx system's cloud integration scheme



The Future of Sequencing Impacts the Future of Pandemics

Clear Labs harnesses the power of NGS to simplify complex diagnostics for clinical and applied markets. By creating a fully automated platform that brings together DNA/RNA sequencing, robotics and cloud-based analytics, Clear Labs democratizes genomics applications to deliver increased clarity. Clear Labs' turnkey platform accelerates outcomes and improves accuracy – from foodborne pathogens to infectious diseases, including SARS-CoV-2. The automated sample-to-result solution eliminates the manual manipulation associated with sequencing, an important roadblock to the clinical adoption of WGS platforms. The automated platform liberates the use of WGS in laboratories that are not staffed with bioinformaticians and short-staffed on molecularly trained staff, which allows for public health departments to focus on the most critical epidemiological analysis in a timely manner.

As of February 2022, over 17.5 billion reads of data have been generated by Clear Labs clients. The Clear Dx platform has been placed in more than 50% of the U.S. State public health laboratories with anticipated adoption of more State and County public health laboratories and commercial reference laboratories in 2022. The SARS-CoV-2 virus will transition to become an endemic virus and thus continuous surveillance will be needed, particularly as some population groups remain unvaccinated and vulnerable to serious effects from infection. In addition to consistency with surveillance of endemic infectious organisms, it is a reality that pathogens emerge and re-emerge constantly. It will be integral to invest appropriately in sequencing efforts so that laboratories, clinicians and global health officials are equipped for routine response as well as for the next pandemic.

Published March 31, 2022

Clear Labs

To learn more, contact: (650) 257-3304 | inquiries@clearlabs.com

clearlabs.com