Fully Automated Isolate Whole Genome Sequencing Platform to Enable Agnostic Characterization and Real-Time Outbreak Investigations of Foodborne Pathogens

Rick G. Kim, Sam Hoeffel, Mark Deiparine, Michael Balamotis, David Tran, Akshay Paropkari, Hua Gao, Anay Campos, James Maloney, Atul Singh, Han Jia Justin Ng, Derreck Carter-House, Norma A. Ordaz-Yuan, Andrew Lin, Alexander Jorjorian, Adam Allred, Charlie Vaske, Shadi Shokralla, Alfred Haddad, Ramin Khaksar, Prasanna Thwar

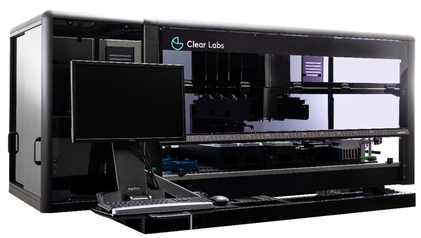
BACKGROUND

Approximately 48 million people get sick, 128,000 get hospitalized, and 3,000 die from foodborne diseases each year in the United States. Foodborne outbreaks are predominantly caused by bacteria such as *Salmonella*, *Listeria*, Shiga toxin-producing *Escherichia* and *Vibrio*. With the advancement of high-throughput sequencing technology, whole genome sequencing (WGS) of pathogens in public health laboratories has improved surveillance for foodborne disease outbreaks and antimicrobial resistance. Isolate WGS (iWGS) provides genome-wide high-resolution data for identifying and tracking outbreaks sooner and enables phylogenetic studies of bacterial genomes. Despite the growing demand and numerous benefits, the WGS library preparation workflow includes several critical steps that require careful human attention and labor, thereby limiting the widespread adoption.

METHODS

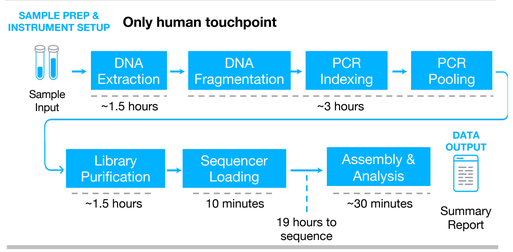
Clear Labs has successfully developed an end-to-end iWGS workflow on a completely integrated platform, the Clear Dx™ Microbial Surveillance, that includes a liquid handling robot, thermocyclers, plate racks, magnets and Illumina iSeq sequencer(s) (Figure 1). The workflow is fully automated and includes genomic DNA extraction, library preparation, amplification and purification of sequencing libraries, loading of libraires and sequencing cartridges onto the sequencer and the automatic upload of sequencing data onto a cloud server. Data is subsequently processed through our scalable, modular, dockerized, WDL-based bioinformatic pipeline housed in our cloud server, which includes genome assembly and a suite of secondary analyses.

Figure 1: The Clear Dx™ Automated Platform



The workflow is also simplified to minimize hands-on time required for sample preparation and sample loading at the beginning to eliminate any human intervention after the initiation of a run (Figure 2). The final reports, with FASTQs and other run metrics, are available for download from an intuitive web browser interface.

Figure 2: The Clear Dx™ Automation Workflow



RESULTS

Figure 3: Run-level Metrics of Automated Runs

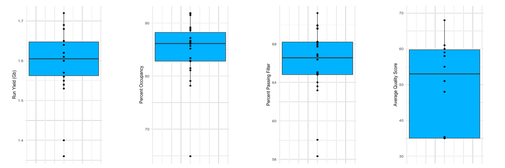


Figure 3 shows the run-level sequencing metrics such as Run Yield (Gb), Percent Occupancy, Percent Passing Filter and Average Quality Score from n = 26 fully automated Clear Dx™ Microbial Surveillance platform runs, performed on multiple instruments. The average run yield is ~1.6 Gb with the percent occupancy of flow cell at ~86%. Average Percent Passing Filter is around 67% and average read Q score is above 50.

Figure 4: Observed Mean Coverage Depth

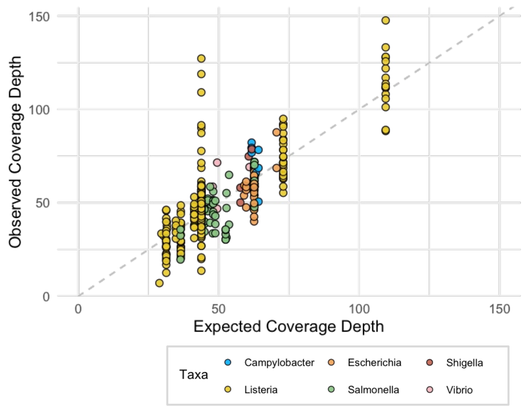


Figure 4 compares the observed coverage depths and the expected coverage depths for 376 samples from six PulseNet pathogens of interest. The results were generated from 54 sequencing runs using the Clear Dx™ Microbial Surveillance platform. The expected coverage is calculated using the pooling ratios of the sample libraries, the average run yield and the respective pathogen genome sizes. Since the data used in this figure includes sequencing runs with diverse sample throughput (ranging from 3 to 12), the expected coverage for some of the smaller pathogens spans a wider range.

Table 1: Summary of Sequencing and Assembly Metrics

Pathogen (genus)	Sample Size	Avg. Cov. Depth (x)	Avg. # of Contigs	Avg. N50 (bp)	Avg. MIS
Listeria	261	50.8	14	1,103,000	324 ± 64
Salmonella	65	50.0	49	276,000	363 ± 58
Escherichia	32	58.5	187	119,000	357 ± 78
Campylobacter	9	71.1	27	168,000	N/A
Shigella	5	65.4	400*	25,000*	335 ± 76
Vibrio	4	61.5	77	208,000	303 ± 62

* Sub-par assembly metrics possibly due to poor quality of reference sequences in our database

Table 1 summarizes the key sample-level sequencing metrics for 376 samples across the six PulseNet pathogens of interest from 54 sequencing runs performed on the Clear Dx™ Microbial Surveillance platform. Most of these metrics comfortably meet the acceptance criteria established by PulseNet.

Figure 5: Reproducibility of Sequencing Results

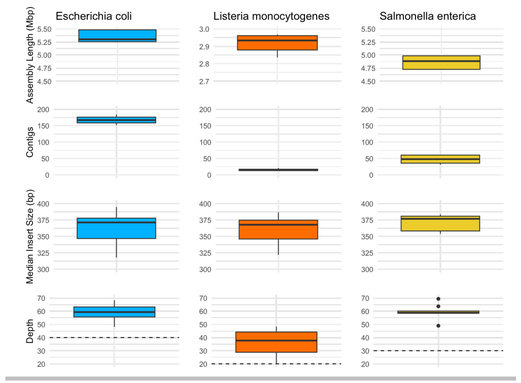


Figure 5 shows the Mean Coverage Depth, Median Insert Size (bp), # of Contigs and Assembly Length (Mbp) of 12 samples (3 *Escherichia coli* strains, 6 *Listeria monocytogenes* strains and 3 *Salmonella enterica* strains) across four replicate runs. The dashed lines represent the PulseNet minimum coverage depth thresholds for the respective pathogens.

All the samples in this study met the PulseNet acceptance criteria and the tightness of the data distributions illustrate the high reproducibility of sequencing performance and results produced by the fully automated workflow on the platform.

Table 2: Serotyping and MLST Results

Pathogen	Serotype	# Rep.	Serotyping Approach		MLST Approach	
			Signature Markers	Concordance	Signature Markers	Concordance
Salmonella	Enteritidis	4	O:9; fliC:g,m	4/4	aroC:5; dnaN:2; hemD:3; hisD:7; purE:6; sucA:6; thrA:11	4/4
	Newport	4	O:8; fliC:e,h; fljB:1,2	4/4	aroC:10; dnaN:7; hemD:12; hisD:9; purE:5; sucA:9; thrA:2	4/4
	Typhimurium	4	O:4; fliC:i; fljB:1,2	4/4	aroC:10; dnaN:7; hemD:21; hisD:14; purE:15; sucA:12; thrA:12	4/4
Escherichia	O157:H7	4	wzx:1; wzy:118; fliC:1	4/4	adk:12; fumC:12; gyrB:8; icd:12; mdh:15; purA:2; recA:2	4/4
	O26:H11	4	wzx:1; wzy:2; fliC:4	4/4	adk:6; fumC:4; gyrB:3; icd:17; mdh:7; purA:7; recA:6	4/4
	O45:H2	4	wzx:1; wzy:18; fliC:2	4/4	adk:16; fumC:4; gyrB:12; icd:16; mdh:9; purA:7; recA:7	4/4
Listeria	1/2a, 3a	4	PRS:1; LMO0737:1	4/4	abcZ:5; bglA:8; cat:5; dapE:7; dat:6; ldh:22; lhkA:1	4/4
	1/2c, 3c	4	PRS:1; LMO0737:1; LMO1118:1	4/4	abcZ:6; bglA:5; cat:6; dapE:4; dat:1; ldh:62; lhkA:1	4/4
	4b, 4d, 4e	4	PRS:1; ORF2110:1; ORF2819:1	4/4	abcZ:1; bglA:1; cat:11; dapE:11; dat:2; ldh:12; lhkA:5	4/4

Table 2 highlights the reproducibility of the platform for various pathogen subtyping applications (serotyping and MLST) with 100% concordance. Three serotypes each of *Salmonella*, *Escherichia* and *Listeria* were tested across four replicates. Common surface antigen profiles for all replicates tested were successfully identified by serotyping. Similarly, pathogen-specific allelic profiles (same sequence type) were successfully assigned to all replicates tested using the MLST approach[1-3].

CONCLUSION

With growing demand for WGS for pathogen surveillance and outbreak investigations, timely processing of bacterial isolates on a large scale is critical. The Clear Dx™ Microbial Surveillance platform from Clear Labs leverages the well-established tagmentation-based shotgun WGS workflow from Illumina and the faster sequencing speed of the compact iSeq sequencer. This fully automated iWGS platform will enable many public health labs to achieve accurate and consistent results with reduced cost and hands-on time (from more than 6 hours down to less than 30 minutes). The power of this technology is not limited to only foodborne pathogens but can also be applied to many other clinical use cases such as healthcare-associated infections (HAI) outbreak investigations and pathogen surveillance programs.

REFRENCES

1. Achtman M, Zhou Z, Alikhan NF, et al. (2020) “Genomic diversity of Salmonella enterica -The UoWUCC 10K genomes project”, Wellcome Open Research 5, 223, https://doi.org/10.12688/wellcomeopenres.16291.1
2. Zhou Z., Alikhan NF, Mohamed K, the Agama Study Group, Achtman M (2020), “The EnteroBase user’s guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny and Escherichia core genomic diversity”, Genome Res. 2020. 30: 138-152, https://doi.org/10.1101/gr.251678.119
3. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol. 2012 Apr;50(4):1355-61. doi: 10.1128/JCM.06094-11. Epub 2012 Jan 11. PMID: 22238442; PMCID: PMC3318499.

For more general information, please contact us at inquiries@clearlabs.com
For specific information regarding the poster, please contact rick.kim@clearlabs.com

1559 Industrial Road
San Carlos, CA 94070

Disclaimer:
The Clear Dx instrument is for Research Use Only (RUO) and is not intended for diagnostic purposes.

Presented at