illumina®

Food safety surveillance with the MiSeq™ i100 Series

Accurate detection of foodborne pathogens

Reduced hands-on time with simplified library prep

Faster sequencing run times and higher outputs

Reliable assembly of high-quality bacterial genomes

M-GL-02912 v1.0

Introduction

Managing food safety is an essential part of public health. Foodborne illness related to microbial contamination from bacteria, viruses, and parasites represents a substantial threat to human health. Approximately 1 in 10 people worldwide become ill with contaminated food. In addition, food safety has a significant economic impact, resulting in $110 billion lost each year to productivity and health care costs from unsafe foods in low- and middle-income countries.[1,2]

The primary goal of foodborne disease surveillance is the early detection of clusters, facilitating early interruption of transmission, thereby preventing outbreak expansion and associated health care costs.[3] The advent of molecular surveillance via next-generation sequencing (NGS) offers higher resolution than historical methods for cluster detection and source attribution to specific products.[4] It also provides vastly more information in a single workflow, including species identification, type prediction, and detection of alleles or genes associated with antimicrobial resistance (AMR).[5] NGS significantly shortens the time to results, enabling a faster public health response.

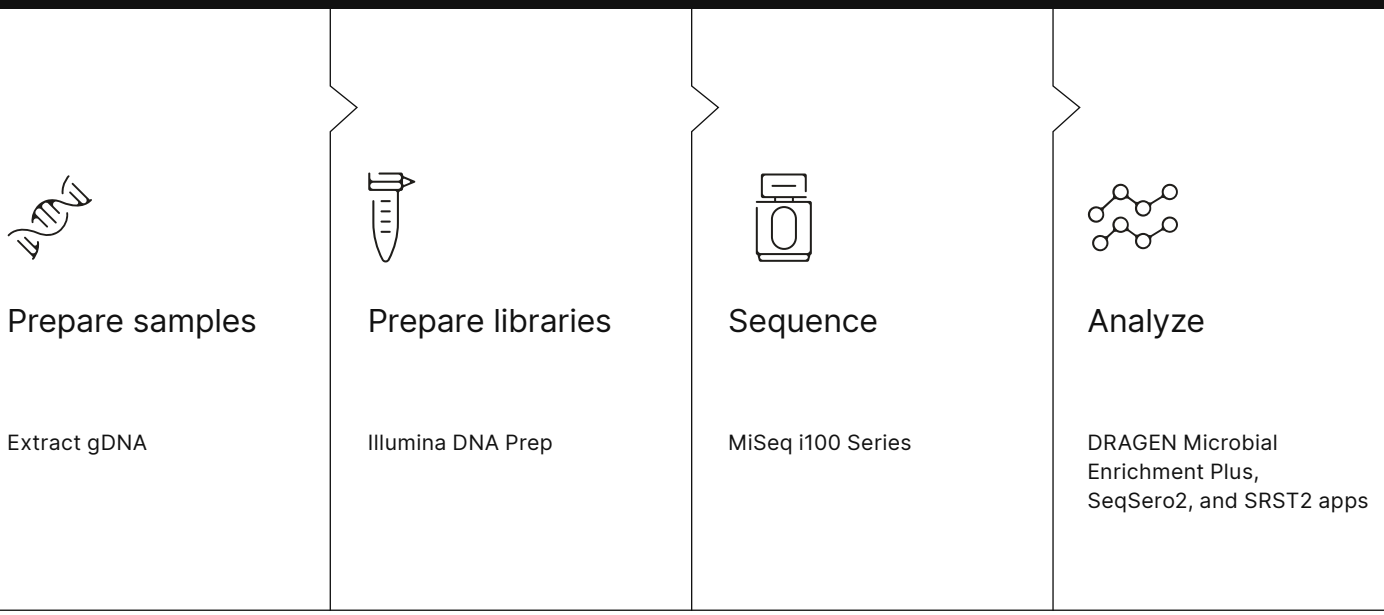
This application note demonstrates highly accurate detection and characterization of the foodborne pathogen *Salmonella* in real-world samples using an NGS workflow that integrates Illumina DNA Prep, the MiSeq i100 Series, and DRAGEN™ secondary analysis (Figure 1).

Methods

Samples

Bacterial isolates from human specimens collected during a foodborne illness investigation in the United States and were cultured on BBL Blood Agar Base (without the addition of blood) (BD, Catalog no. 211037). Genomic DNA (gDNA) was extracted using the Wizard Genomic DNA Purification Kit (Promega, Catalog no. A1120) with slight modifications to the protocol for the isolation of gDNA from gram-negative bacteria. Purified gDNA was quantified with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Catalog no. Q32851) before library preparation.

Figure 1: Comprehensive NGS workflow for foodborne pathogen detection



Combine Illumina library preparation with sequencing on the MiSeq i100 Series and DRAGEN secondary analysis for accurate foodborne pathogen surveillance.

Library preparation

Sequencing-ready libraries were prepared manually from 100 ng extracted gDNA using Illumina DNA Prep (Illumina, Catalog no. 20060060) with modifications to the protocol to improve performance on the MiSeq i100 Series, including an additional round of bead purification with a bead to sample ratio of 0.5× to remove short inserts.[6] The quality and concentration of PCR-amplified libraries were assessed using the 4200 TapeStation System (Agilent Technologies, Catalog no. G2991BA) and the Qubit 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, Catalog no. Q33231) before pooling.

Sequencing

Prepared libraries were pooled and diluted to a loading concentration of 60 pM (32 libraries/run), then sequenced on the MiSeq i100 Plus System using a 25M flow cell with a run configuration of 2 × 301 bp. For larger studies, sequencing runs can be scaled up to the NextSeq™ 1000, NextSeq 2000, NovaSeq™ 6000, or NovaSeq X Systems.

Data analysis

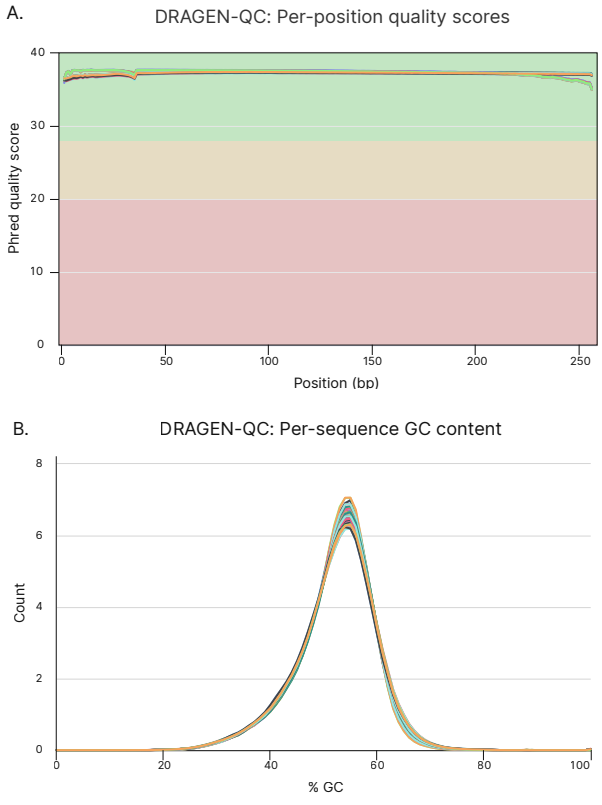
After sequencing was complete, data was streamed to BaseSpace™ Sequence Hub and analyzed with the DRAGEN FastQC + MultiQC app to provide general statistics and metrics, sequence mapping, and alignment. The DRAGEN Microbial Enrichment Plus pipeline was used for secondary analysis of sequencing data and to perform pathogen detection and quantification and AMR marker profiling. The third-party, off-instrument analysis pipeline SeqSero2 was used specifically for *Salmonella* serotype prediction from sequencing data.[7] The SRST2 (Short Read Sequence Typing for Bacterial Pathogens) app in BaseSpace Sequence Hub was used to perform rapid molecular typing of bacterial pathogens using whole-genome sequencing (WGS) data and report the presence of sequence types (ST) from a multilocus sequence typing (MLST) database and/or reference genes from a sequence database of virulence genes, resistance genes, and plasmid replicons.[8] Seven MLST loci (ST, aroC, dnaN, hemD, hisD, purE, sucA, thrA) were used for analysis of *Salmonella enterica* samples.[9]

Results

Sequencing metrics

Demultiplexing data was generated after 20 cycles with a total run time of 14 hr and 20 min, confirming libraries were quantified and pooled correctly. The MiSeq i100 Plus System generated sequencing data with an average of 1.88M reads per sample at a read length of 2 × 301 bp with ≥ 97.5% of bases above Q30 (Figure 2A). The observed percent GC content of reads followed a roughly normal distribution, indicating that library preparation chemistry and sequencing were not heavily biased toward higher or lower GC regions (Figure 2B).

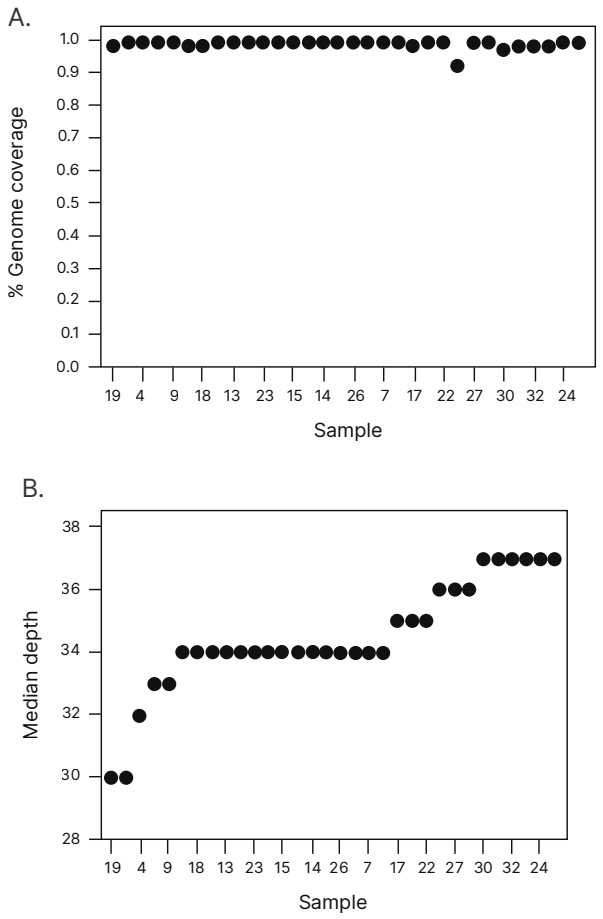
Figure 2: High-quality data for foodborne pathogen detection



(A) The MiSeq i100 Plus System generates high-quality data with ≥ 97.5% of bases above Q30 (green area) for all samples analyzed (colored lines). (B) The percent GC content distribution for sequenced reads on the MiSeq i100 Plus System met expected levels for all 32 samples analyzed (colored lines). Data was processed using the DRAGEN FastQC + MultiQC app.

Analysis with the DRAGEN Microbial Enrichment Plus pipeline reported all samples had genome coverage of ≥ 92% at a depth threshold of 1× (Figure 3A). The median depth for all samples was ≥ 30 (Figure 3B).

Figure 3: Coverage of *Salmonella* genome



Analysis with the DRAGEN Microbial Enrichment Plus pipeline resulted in (A) coverage across the *Salmonella* genome with a mean fraction of genome covered at 0.99 and (B) median genome depth for all samples ranging from 30–37.

Salmonella serotype prediction

FASTQ files were analyzed using SeqSero2 software for *Salmonella* serotype prediction. Thirty-one samples were correctly predicted as *Salmonella infantis* (7:r:1,5) and one was correctly predicted as *Salmonella enteritidis* (9:g,m:-) (Table 1).

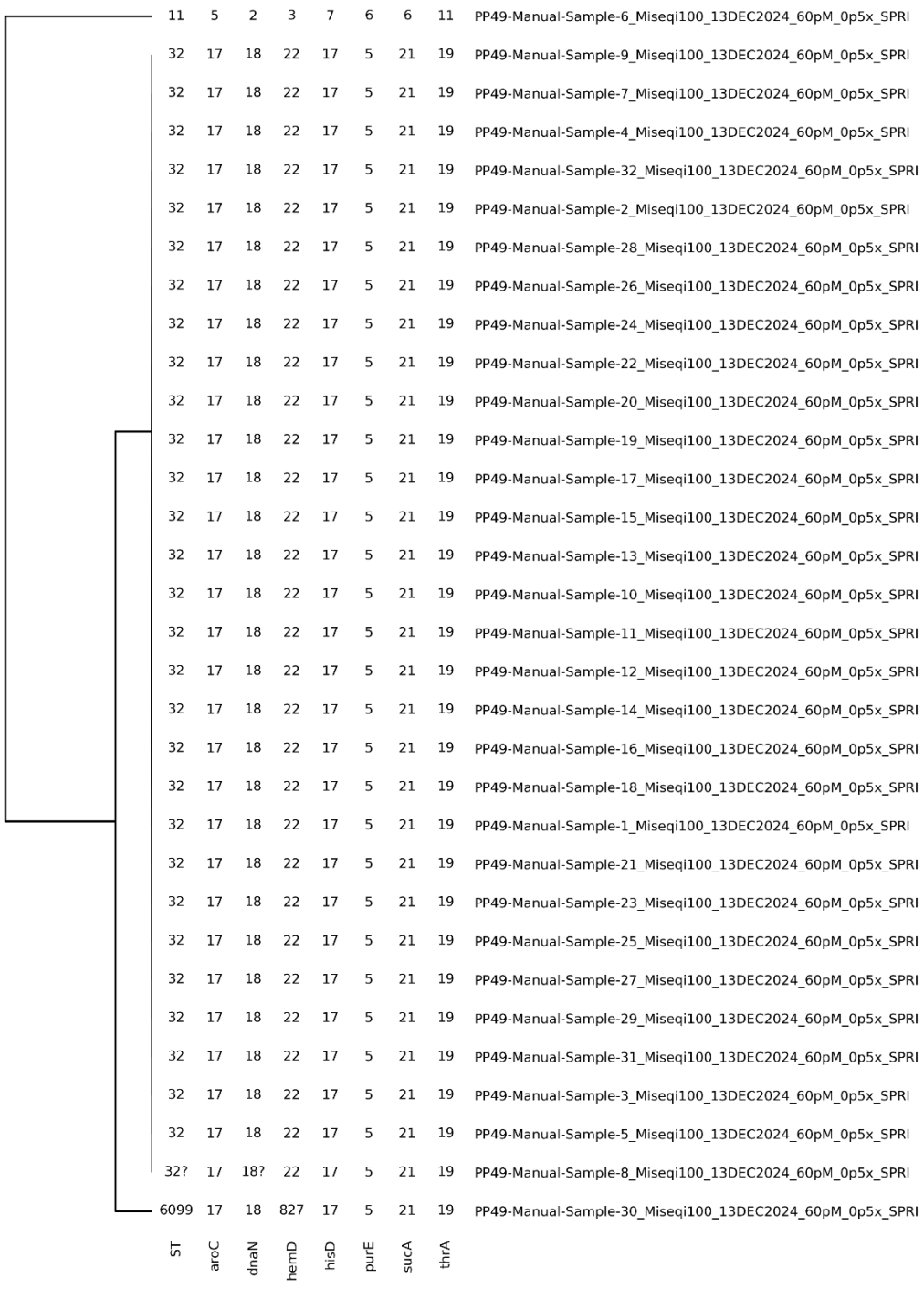
Molecular typing of pathogens

Cluster analysis of seven MLST loci in *Salmonella enteritidis* inferred the degree of genomic relatedness between the isolates in the sample set.[9] Samples clustered in three clades (Figure 3). Samples were also compared by 258 Virulence Factor Genes (Virulence Factors of Pathogenic Bacteria, updated May 10, 2022), 13 Resistance Gene (CARD DataBase), and Plasmid (PlasmidFinder). Samples were aggregated and the Jaccard similarity coefficient of detected alleles was computed for every pair of samples, demonstrating high similarity across all 32 samples analyzed (data not shown).

Table 1: *Salmonella* serotype prediction with SeroSeq2 software

Sample	O antigen	H1 antigen	H2 antigen	Predicted species	Sample	O antigen	H1 antigen	H2 antigen	Predicted species
1	7	r	1,5	*Salmonella infantis*	17	7	r	1,5	*Salmonella infantis*
2	7	r	1,5	*Salmonella infantis*	18	7	r	1,5	*Salmonella infantis*
3	7	r	1,5	*Salmonella infantis*	19	7	r	1,5	*Salmonella infantis*
4	7	r	1,5	*Salmonella infantis*	20	7	r	1,5	*Salmonella infantis*
5	7	r	1,5	*Salmonella infantis*	21	7	r	1,5	*Salmonella infantis*
6	9	g,m	—	*Salmonella enteritidis*	22	7	r	1,5	*Salmonella infantis*
7	7	r	1,5	*Salmonella infantis*	23	7	r	1,5	*Salmonella infantis*
8	7	r	1,5	*Salmonella infantis*	24	7	r	1,5	*Salmonella infantis*
9	7	r	1,5	*Salmonella infantis*	25	7	r	1,5	*Salmonella infantis*
10	7	r	1,5	*Salmonella infantis*	26	7	r	1,5	*Salmonella infantis*
11	7	r	1,5	*Salmonella infantis*	27	7	r	1,5	*Salmonella infantis*
12	7	r	1,5	*Salmonella infantis*	28	7	r	1,5	*Salmonella infantis*
13	7	r	1,5	*Salmonella infantis*	29	7	r	1,5	*Salmonella infantis*
14	7	r	1,5	*Salmonella infantis*	30	7	r	1,5	*Salmonella infantis*
15	7	r	1,5	*Salmonella infantis*	31	7	r	1,5	*Salmonella infantis*
16	7	r	1,5	*Salmonella infantis*	32	7	r	1,5	*Salmonella infantis*

Figure 4: Clustering of study isolates based on MLST



Seven MLST loci (ST, aroC, dnaN, hemD, hisD, purE, sucA, thrA) were used for analysis in *Salmonella enteritidis* samples. Samples clustered into three clades, showing the genomic similarity between isolates in the sample set. Data analysis was perfromed using the SRTS2 app, clustering of study isolates was based on MLST.

Summary

Foodborne illness related to microbial contamination from bacteria, viruses, and parasites represents a substantial threat to human health. The MiSeq i100 Series helps enable a fast, comprehensive NGS workflow that enables highly accurate detection of foodborne pathogens for effective surveillance as part of public health efforts.

Learn more

Illumina DNA Prep

MiSeq i100 Series

DRAGEN secondary analysis

References

1. CDC and Food Safety. cdc.gov/food-safety/about/what-cdc-is-doing.html?CDC_AAref_Val=https://www.cdc.gov/foodsafety/cdc-and-food-safety.html. Accessed February 5, 2025.
2. World Health Organization and Food Safety. who.int/news-room/fact-sheets/detail/food-safety. Accessed February 5, 2025.
3. Scharff RL, Besser J, Sharp DJ, Jones TF, Peter GS, Hedberg CW. An Economic Evaluation of PulseNet: A Network for Foodborne Disease Surveillance. *Am J Prev Med*. 2016;50(5 Suppl 1):S66-S73. doi:10.1016/j.amepre.2015.09.018
4. Ribot EM, Freeman M, Hise KB, Gerner-Smidt P. PulseNet: Entering the Age of Next-Generation Sequencing. *Foodborne Pathog Dis*. 2019;16(7):451-456. doi:10.1089/fpd.2019.2634
5. Leeper MM, Tolar BM, Griswold T, et al. Evaluation of whole and core genome multilocus sequence typing allele schemes for Salmonella enterica outbreak detection in a national surveillance network, PulseNet USA. *Front Microbiol*. 2023;14:1254777. Published 2023 Sep 21. doi:10.3389/fmicb.2023.1254777
6. Illumina. Maximizing performance on the MiSeq i100 Series. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/miseq-i100-library-optimization-tech-note-m-gl-03322/miseq-i100-library-optimization-tech-note-m-gl-03322.pdf. Published 2024. Accessed February 5, 2025.
7. Zhang S, den Bakker HC, Li S, et al. SeqSero2: Rapid and Improved Salmonella Serotype Determination Using Whole-Genome Sequencing Data. *Appl Environ Microbiol*. 2019;85(23):e01746-19. Published 2019 Nov 14. doi:10.1128/AEM.01746-19
8. Inouye, M., Dashnow, H., Raven, LA. et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med*. 6, 90 (2014). https://doi.org/10.1186/s13073-014-0090-6
9. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. 2018;3:124. Published 2018 Sep 24. doi:10.12688/wellcomeopenres.14826.1

illumina®

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

M-GL-02912 v1.0