illumina®

APPLICATION NOTE

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

Introduction

Using laboratory methods to subtype enteric bacteria, including *Salmonella*, has been critical in identifying potential outbreaks and linking bacteria-causing illnesses to potential outbreak sources. As the resolution of laboratory subtyping methods has increased, the precision in connecting illness with a source has also grown. In the United States, linking of potential outbreak cases is done through PulseNet, a national molecular surveillance network for foodborne, waterborne, and One Health–related enteric illnesses found in over 80 state, local, and federal laboratories. The PulseNet network prevents over 250,000 illnesses from enteric bacteria every year and saves half a billion dollars in medical costs and lost productivity.[1,2]

PulseNet began implementing whole-genome sequencing (WGS) using Illumina instruments in the early 2010s with federal partners at the US Food and Drug Administration (FDA) and the US Department of Agriculture Food Safety and Inspection Service (USDA-FSIS).[3] Using WGS, public health scientists can identify potential outbreaks by identifying strains from cases that are highly genetically similar and associated with illnesses occurring within the same time or geographic region. WGS is also used to identify trends in bacterial strains that were not previously characterized by lower resolution molecular methods.[4,5] These strains reoccurred over time periods and caused outbreaks, emerged and caused increasing illness, or persisted causing illnesses over months or years. The Centers for Disease Control and Prevention (CDC) began describing these strains as reoccurring, emerging, and persisting (REP) strains. REP strains can be identified by genetic relatedness and other defining characteristics, including multidrug resistance, virulence, or other genetic markers. Because only approximately 10% of illness-causing bacteria in PulseNet are associated with an outbreak, using WGS to identify REP strains may provide an opportunity to prevent further illnesses by linking additional strains that may share a common source.

This application note demonstrates accurate detection and characterization of a persisting strain of multidrug-resistant *Salmonella enterica* serotype Infantis (REP JFX01) using an efficient NGS workflow that integrates Illumina DNA Prep, the MiSeq i100 Series, and third-party analysis tools (Figure 1).

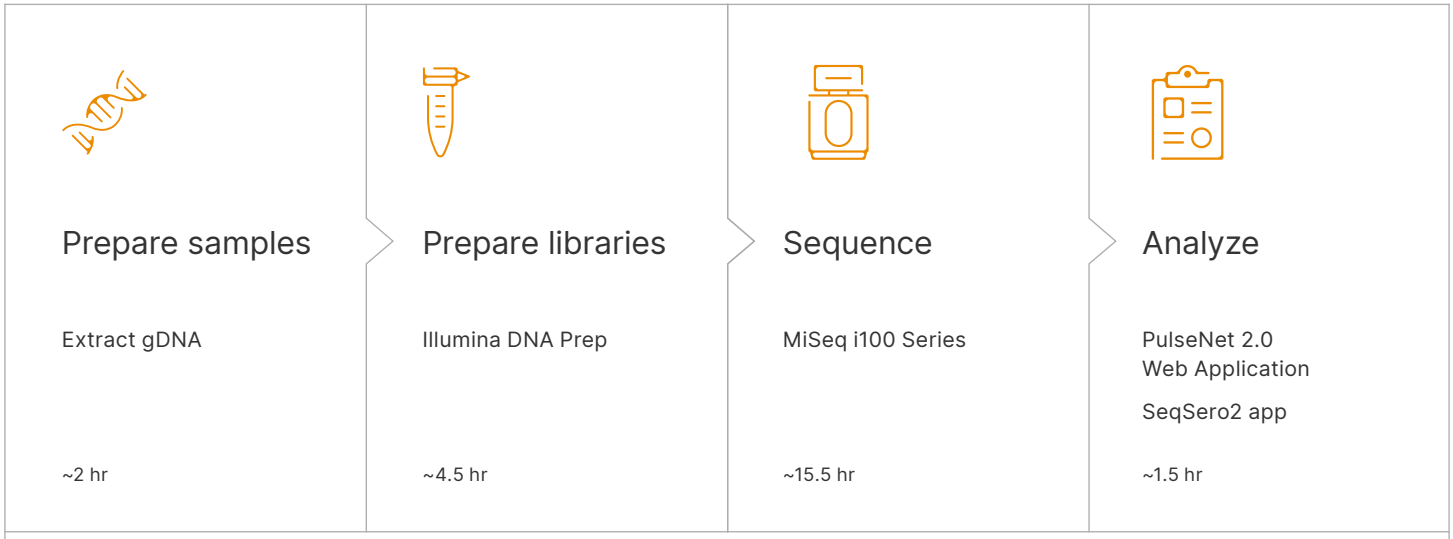

Figure 1: Comprehensive NGS workflow for *Salmonella* detection
Combine Illumina library preparation with sequencing on the MiSeq i100 Series and DRAGEN secondary analysis for accurate surveillance of the foodborne pathogen *Salmonella*.

Methods

Samples

Bacterial isolates collected as part of PulseNet routine surveillance 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) following a previously described protocol for the isolation of gDNA from Gram-negative bacteria.[6] Purified gDNA was quantified with the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Catalog no. Q32853) before library preparation.

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.[7] 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 by volume following a previously described protocol[6] with a slight modification that included an additional round of bead purification at 0.5× SPRI performed after quantification and equimolar pooling of each library. Pooled libraries were diluted to a loading concentration of 60 pM (32 libraries/run). Sequencing was performed 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. Analysis was performed in the PulseNet 2.0 Web Application. Within this application, read- and assembly-based quality control (QC) metrics were calculated,[5] core genome multilocus sequence typing (cgMLST) was performed using the PulseNet 2.0 allele caller,[6] serotyping was performed using SeqSero2,[8] AMR profiling was performed using AMRFinderPlus,[9] and plasmids were identified from the PlasmidFinder databases.[10] The cgMLST tree and associated genotypes were visualized using the iTOL v7 web application.[11]

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).

The 32 bacterial isolates that were sequenced demonstrated consistently high sequencing quality, with a median Q-score of 37.1 (Table 1). The estimated depth of coverage ranged from 46.1× to 109.9×, with an average of 71.9×. The percent of core loci of the PulseNet 2.0 *Salmonella* cgMLST scheme was > 99% for all samples, indicating sufficient genome completeness for clustering by cgMLST.

Salmonella serotype prediction

FASTQ files were analyzed using SeqSero2 software for *Salmonella* serotype prediction. Thirty-one samples were correctly predicted as *Salmonella* Infantis (I 7:r:1,5) and one was correctly predicted as *Salmonella* Enteritidis (I 9:g,m:-) (Figure 3).

Molecular typing of pathogens

Cluster analysis of 3002 cgMLST loci in the PulseNet 2.0 *Salmonella* scheme inferred the degree of genomic relatedness between the isolates in the sample set.

Samples clustered as expected, with samples typed as REPJFX01 clustering together. Additional *Salmonella* Infantis samples and one *Salmonella* Enteritidis sample that were not typed as REPJFX01 cluster separately as expected (Figure 3). Seven gene multilocus sequence types (ST) were found using *loci aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*. The sequence types (ST) matched the expected result for all 32 samples. Predicted AMR matched the expected results for all 32 samples and, notably, the gyrA_D87Y mutation that is characteristic of REPJFX01 strains was detected in all REP samples of the data set (Figure 3). Plasmid content matched the expected results for all 32 samples, and the IncFIB (pN55391) plasmid was detected in all REPJFX01 strains, which is also a defining characteristic of this REP strain (Figure 3).

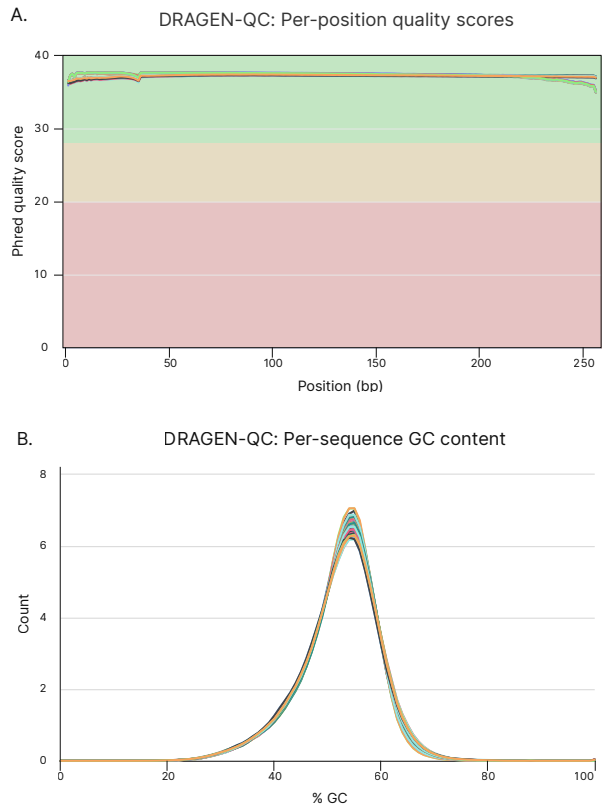

Figure 2: High-quality data for foodborne pathogen detection

(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.

Table 1: Select quality metrics

Sample	Average quality (Q-score)	Estimated depth of coverage	Percent core loci called
1	37.1	76.7×	99.3%
2	37.1	93.0×	99.3%
3	37.1	68.3×	99.3%
4	37.1	67.2×	99.2%
5	37.1	65.5×	99.3%
6	37.0	96.9×	99.4%
7	37.1	77.3×	99.3%
8	37.0	62.1×	99.3%
9	37.1	59.9×	99.2%
10	37.0	57.7×	99.2%
11	37.0	63.4×	99.3%
12	37.0	52.8×	99.3%
13	37.1	86.2×	99.3%
14	37.0	55.7×	99.3%
15	37.0	66.8×	99.2%
16	37.0	55.6×	99.4%
17	37.1	109.9×	99.5%
18	37.1	70.5×	99.5%
19	37.0	54.8×	99.5%
20	37.0	46.1×	99.5%
21	37.1	77.0×	99.3%
22	37.1	78.2×	99.2%
23	37.1	76.6×	99.2%
24	37.1	100.6×	99.2%
25	37.1	80.1×	99.2%
26	37.2	78.0×	99.2%
27	37.1	63.0×	99.3%
28	37.0	73.4×	99.3%
29	37.1	88.8×	99.2%
30	37.0	61.6×	99.5%
31	37.1	71.3×	99.5%
32	37.1	66.8×	99.4%

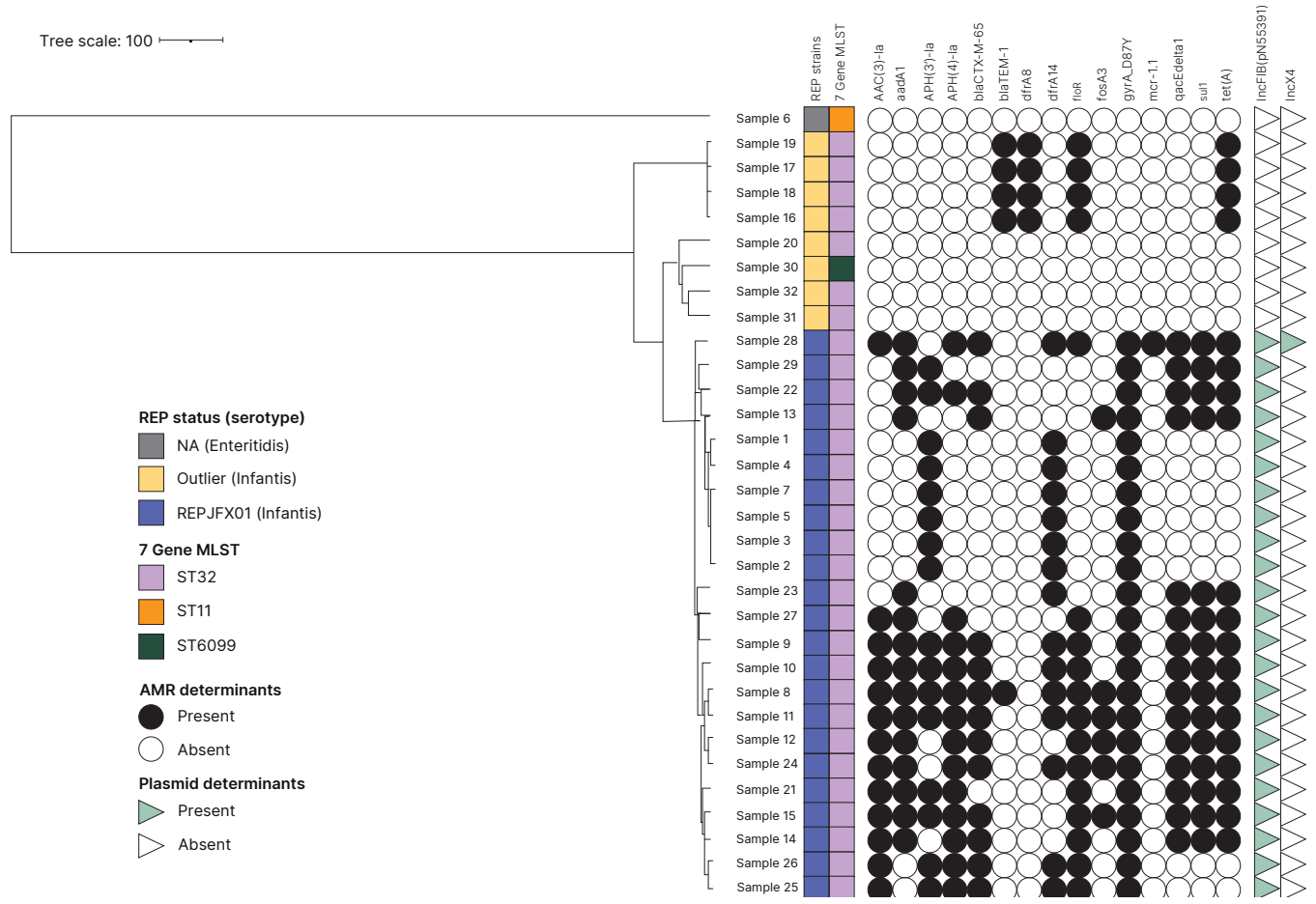

Figure 3: Clustering of study isolates based on cgMLST and associated genotype data

cgMLST based unweighted pair group method with arithmetic mean (UPGMA) tree of 32 *Salmonella enterica* samples with key genotype information; REP status and serotype Information, 7-gene MLST, AMR determinants, and plasmid determinants are to the right of the tree.

Summary

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

Learn more

Illumina DNA Prep

MiSeq i100 Series

References

1. 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
2. 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
3. Stevens EL, Carleton HA, Beal J, et al. Use of Whole Genome Sequencing by the Federal Interagency Collaboration for Genomics for Food and Feed Safety in the United States. *J Food Prot*. 2022;85(5):755-772. doi:10.4315/JFP-21-437
4. Tolar B, Joseph LA, Schroeder MN, et al. An Overview of PulseNet USA Databases. *Foodborne Pathog Dis*. 2019;16(7):457-462. doi:10.1089/fpd.2019.2637
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. Leeper MM, Schroeder MN, Griswold T, et al. Validation of Core and Whole-Genome Multi-Locus Sequence Typing Schemes for Shiga-Toxin-Producing E. coli (STEC) Outbreak Detection in a National Surveillance Network, PulseNet 2.0, USA. *Microorganisms*. 2025;13(6):1310. Published 2025 Jun 4. doi:10.3390/microorganisms13061310
7. Illumina. Maximizing performance on the MiSeq i100 Series. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/miseq-i100-library-optimization-technote-m-gl-03322/miseq-i100-library-optimization-technote-m-gl-03322.pdf. Published 2024. Accessed February 5, 2025.
8. 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
9. Feldgarden M, Brover V, Gonzalez-Escalona N, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep*. 2021;11(1):12728. Published 2021 Jun 16. doi:10.1038/s41598-021-91456-0
10. Carattoli A, Zankari E, García-Fernández A, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014;58(7):3895-3903. doi:10.1128/AAC.02412-14
11. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;52(W1):W78-W82. doi:10.1093/nar/gkae268

illumina®

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

M-GL-02912 v2.0