012

Comparison of Targeted Amplicon Sequencing Using the MiSeq and GridION Next Generation Sequencing Platforms for Detection of Foodborne Pathogens

Isha Patel, Mark Mammel, Jayanthi Gangiredla

Division of Molecular Biology, Office of Applied Safety and Research Assessment, Center for Food Safety and Applied Nutrition

Introduction

The use of Whole Genome Sequencing (WGS) for pathogen detection has increased the accuracy and reduced the time for traceback and source attribution in foodborne outbreaks. However, low level contaminants may go undetected due to challenges in detecting them from food matrices or due to a high background of other microbial flora. We have previously shown using the Illumina platform that a custom targeted amplicon sequencing (TAS) panel offers increased sensitivity and specificity for pathogen detection in "spike-in" experiments. The turnaround time from isolation of DNA to obtaining data is two days. Nanopore GridION platform offers a sequencing approach that enables direct real-time sequencing thus it is expected to save time to get results.

Objective: The objective of this work is to provide a rapid and sensitive method using targeted detection of low-level pathogen contamination in complex samples thereby positively impacting the use of targeted amplicon sequencing as a screening method for pathogen detection.

Materials and Methods

Primer3 software was used to design primers from alignments of multiple sequences of ten core genes for each of the 266 species that included 135 pathogens. The desired amplicon size was set to approximately 600bp. The custom primer panel was designed in collaboration with Swift Biosciences (Ann Arbor, MI).

DNA reference material from 3 strains (NIST RM 8376) was used to compare the limit of pathogen detection using the TAS panel. The strains used were *Listeria monocytogenes* ATCC 19115, *Escherichia coli* BAA 2309 and *Salmonella enterica* ATCC 12324. The reference material stock solution was at 40ng/ul. Three serial dilutions of the DNA (100ng, 10ng and 1ng) from the three strains were used to set up the PCR amplification. Equal volume of the three strains were also mixed in one tube at the same concentrations for the pool sample.

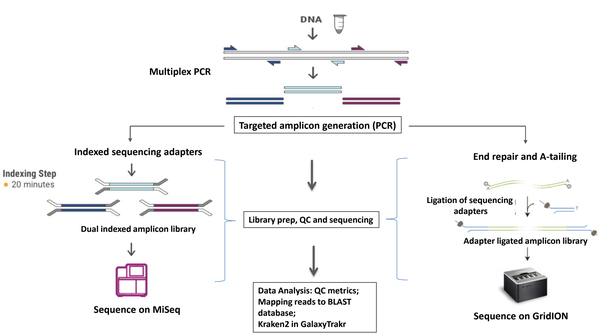

Figure 1. Library preparation workflow

FDA Mission Relevance

The use of a targeted approach for detecting low number of bacterial pathogens without the need of sample enrichment may provide an efficient and effective tool for the FDA to identify foodborne pathogens such as *Escherichia coli, Listeria monocytogenes* and *Salmonella enterica*. Using a real-time sequencing platform like the GridION has potential to significantly reduce labor and time for compliance testing of samples to assure a safe food supply chain particularly for products with a short shelf-life.

Materials and Methods cont.

The amplicons generated after the multiplex PCR step were then used for respective library preparation for the MiSeq and GridION sequencers using the manufacturer's protocol as shown in Figure 1. For the MiSeq library preparation, Ampure XP beads were used to cleanup the reaction followed by indexing and final elution. The library was then quantified, and library size and quality were determined using an Agilent Tapestation. The indexed amplicon library (12 picomolar) was sequenced using the Illumina MiSeq Platform. For the GridION library preparation, DNA ends were repaired and dA-tailed using the NEBNext End Repair/dA-tailing module. This was followed by a ligation step to ligate sequencing adapters. The final library was cleaned using Ampure XP beads and washed using short fragment buffer and eluted. The library quality and size were determined using the Agilent Tapestation. The flow cell was then primed and loaded with 50 femtomoles of DNA library and sequenced for 1 hour.

Sequence Analysis: GalaxyTrakr and BLAST matching of the amplicons were used for data analysis. Kraken2 was used in GalaxyTrakr to identify and classify reads. Additionally, our in-house bioinformatic pipeline was used for identification and quantification of the targeted organisms from the sequence reads datasets. To quantify the number of genes present from each species, all reads were matched by BLAST to a database of MLST genes for each species. The database contains multiple sequence types for each of the 10 genes for each species. The top BLAST match for each read was taken, and the number of matching bases in that read was added to a tally for that gene. The tally for each gene was then normalized by dividing by the amplicon length that was represented in the database. A total count of genes present in each species was obtained by summing the tally for all genes belonging to that species. Additionally, a simpler count for each species was obtained by adding 1 to the count for each species based on the top hit of each read, instead of normalizing by gene length.

Results

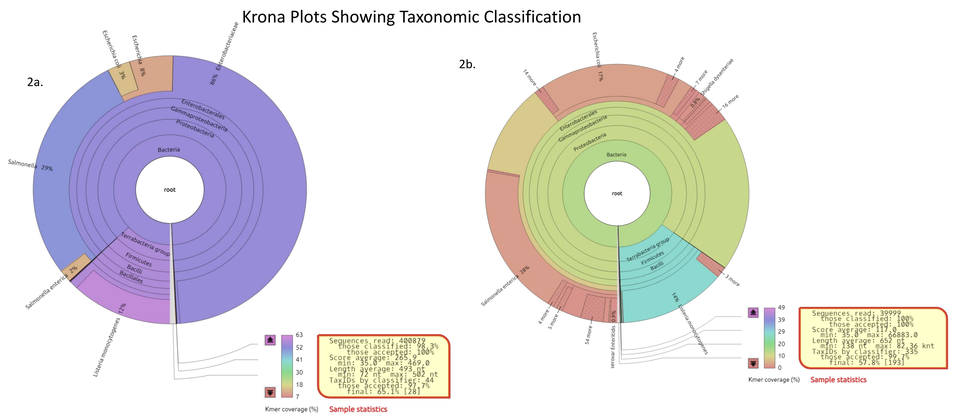

Figure 2
Krona Plots were generated using Kraken2 and Recentrifuge
a. MiSeq data from the pooled sample was used to generate the plot that shows the sample classification summary.
b. GridION data from the pooled sample was used to generate the plot that shows the sample classification summary.

Strain	*aggR*	*stx2A*	LMOf2365_2530 (autolysin, amidase)	*inlA*	*inlB*	*rluA*	*prfA*	*clpC*	*clpP*	*hly*	*spvC*	*sopE*	*pagN*	*misL*	*sinH*	*lpfA*	*invA*	*hilA*	*avrA*
Listeria monocytogenes ATCC 19115			+	+	+	+	+	+	+	+									
1:10			+	+	+	+	+	+	+	+									
1:100			+	+	+	+	+	+	+	+									
Escherichia coli BAA 2309	+	+	+																
1:10	+	+	+																
1:100	+	+	+																
Salmonella enterica ATCC 12324											+	+	+	+	+	+	+	+	+
1:10											+	+	+	+	+	+	+	+	+
1:100											+	+	+	+	+	+	+	+	+
POOL of all 3 strain	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
1:10	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
1:100	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 2. MiSeq and GridION data using Kraken2 in GalaxyTrakr gave similar results for detection of virulence genes that were targeted

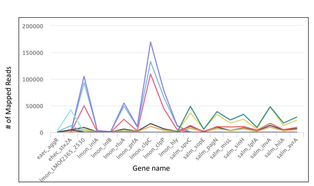

Figure 3. Mapped reads for virulence genes

Results cont.

Strain	Total reads	Average Read length	Mean Quality Score	*L. monocytogenes*	%	*E. coli*	%	*S. enterica*	%
Listeria monocytogenes ATCC 19115	870004	245.7	36.3	809452	93.0				
1:10	1070758	241.5	36.0	939289	87.7				
1:100	739772	221.1	35.4	573359	77.5				
Escherichia coli BAA 2309	2904412	250.5	35.2			2706404	93.2		
1:10	900806	245.4	35.1			862228	95.7		
1:100	683358	236.4	34.9			616794	90.3		
Salmonella enterica ATCC 12324	762622	247.9	35.4					706199	92.6
1:10	936010	250.5	35.4					786215	84.0
1:100	413436	249.4	35.3					298473	72.2
POOL of all 3 strains	801760	243.5	35.2	88491	11.0	393127	49.0	176239	22.0
1:10	632772	244.3	35.4	62021	9.8	320056	50.6	148164	23.4
1:100	1059196	234.3	35.1	52174	4.9	438917	41.4	170326	16.1

Table 1a. MiSeq data analyzed using target specific BLAST

Strain	# Total reads	Average Read length	Mean Quality Score	*L. monocytogenes*	%	*E. coli*	%	*S. enterica*	%
Listeria monocytogenes ATCC 19115	462572	633	21	191525	41.4				
1:10	6183055	681	19	1498510	24.2				
1:100	3012219	670	20	871495	28.9				
Escherichia coli BAA 2309	487438	583	20			117394	24.1		
1:10	568182	607	21			168703	29.7		
1:100	64214	1252	20			17814	27.7		
Salmonella enterica ATCC 12324	7733052	637	19					2081187	26.9
1:10	1646200	1171	20					435586	26.5
1:100	2407312	1557	21					409481	17
POOL of all 3 strains	7494366	624	20	493052	6.6	1231149	16.4	790054	10.5
1:10	597519	606	20	36957	6.2	97811	16.4	62442	10.5
1:100	1458457	1707	20	36093	2.5	82684	5.7	57292	3.9

Table 1b. GridION data analyzed using target specific BLAST

Strain	Total reads	Average Read length	Mean Quality Score	*L. monocytogenes*	%	*E. coli*	%	*S. enterica*	%
Listeria monocytogenes ATCC 19115	435001	245.7	36.3	415396	99.9				
1:10	535378	241.5	36.0	495320	99.9				
1:100	369885	221.1	35.4	299966	99.9				
Escherichia coli BAA 2309	1452205	250.5	35.2			85371	99.9		
1:10	450402	245.4	35.1			32567	99.9		
1:100	341678	236.4	34.9			16870	99.2		
Salmonella enterica ATCC 12324	381310	247.9	35.4					23871	100.0
1:10	468004	250.5	35.4					29896	100.0
1:100	206717	249.4	35.3					18694	99.9
POOL of all 3 strains	400879	243.5	35.2	47050	73.7	10646	16.7	5980	9.4
1:10	316385	244.3	35.4	32048	67.3	9787	20.6	5613	11.8
1:100	529597	234.3	35.1	30250	47.1	16383	25.5	17049	26.6

Table 1c. MiSeq data analyzed using Kraken2 in GalaxyTrakr

Strain	# of raw reads	Average Read length	Mean Quality Score	*L. monocytogenes*	%	*E. coli*	%	*S. enterica*	%
Listeria monocytogenes ATCC 19115	40,000	775.3	21	39366	99.36				
1:10	40,000	718.7	19	38250	96.91				
1:100	36,223	647.1	20	35759	99.65				
Escherichia coli BAA 2309	40,000	1018.7	20			14958	73.3		
1:10	36,183	1286.3	21			12997	73.1		
1:100	36,216	1797.7	20			17579	83.5		
Salmonella enterica ATCC 12324	40,000	600.3	19					30363	98.8
1:10	40,000	598.1	20					22710	70.7
1:100	40,000	1204.3	21					18067	54.1
POOL of all 3 strains	40,000	1021.8	20	5785	22.37	6821	26.4	11078	42.8
1:10	40,000	640.9	20	5553	21.66	6467	25.2	11272	44
1:100	40,000	1819.6	20	3024	9.58	20600	65.3	6438	20.4

Table 1d. GridION data (using only 10 output files) analyzed using Kraken2 in GalaxyTrakr

Results show that both sequencing platforms using their respective analysis pipelines detect specific pathogens and their associated virulence genes at similar levels. The MiSeq took ~39 hours to complete a sequencing run using a 500 cycles cartridge. The sequencing run on the GridION was stopped after 1 hour and the results (Table 1b and 1d) show that we were able to accurately detect the respective pathogenic strain tested.

This shows that GridION offers a real-time sequencing advantage resulting in obtaining sufficient data in about an hour to accurately identify a pathogenic bacterial strain.

Conclusion

- This study shows that targeted sequencing to detect pathogens yields similar results with either GridION or MiSeq, but GridION is faster as it provides data in real time.
- Integrating NGS-based TAS with a high-resolution bioinformatic analytical workflow allows for a simple, reproducible, and rapid detection assay for bacterial pathogens from samples relevant to food-safety.
- TAS integrated with real-time detection using the GridION sequencing platform establishes a new threshold of resolution and specificity to identify various pathogens by integrating NGS, genome-wide target design, and an adaptive bioinformatic analytical workflow.