Detection of Shiga Toxin-Producing *Escherichia coli* from Ready-to-Eat Romaine Lettuce Using Targeted Amplicon Sequencing Approach

Patel, Isha; Mammel, Mark; Gangiredla, Jayanthi.
Division of Molecular Biology, Office of Applied Safety and Research Assessment, Center for Food Safety and Applied Nutrition

Abstract

Next-generation sequencing (NGS) methods offer high throughput, resolution and sensitivity for both detection and identification of foodborne pathogens. Depending on what NGS methods is used, technical challenges remain in detecting low levels of unculturable pathogens in contaminated food. In this study, we evaluate the utility, specificity and limit of detection of targeted amplicon sequencing approach for detection and identification of Shiga toxin-producing *Escherichia coli* by spiking ready to eat romaine lettuce with different levels of inoculum of the bacterium. Using a modified version of the Bacteriological Analytical Methods (BAM), Shiga toxin-producing *Escherichia coli* at 3 different spike levels were inoculated in romaine lettuce. DNA was isolated from cells that were harvested at 0h, 5h and 20h post inoculation. DNA libraries were prepared for shotgun metagenomics using Swift Biosciences 2S library preparation kit. Amplicon libraries were prepared using a custom pathogen kit designed with Swift Biosciences. The libraries were sequenced using Illumina MiSeq Platform. Identification and quantification of the targeted organisms from the sequence reads datasets was done using GalaxyTrakr. The shotgun metagenomics data as well as the amplicon sequencing data for the 5h and 20h both had significant reads for *E. coli* when 30 and 300 colony forming units (CFU) were used to spike-in. At spike level of 3 CFU, the amplicon libraries were better at not only detecting the pathogen at species level, but also detecting the pathogenic markers such as *stx2A*, *eaeA* and *ehxA* at the 5h timepoint post inoculation. The targeted approach for detecting low levels of pathogens provides a rapid molecular approach for the FDA to identify foodborne bacterial such as *E. coli*.

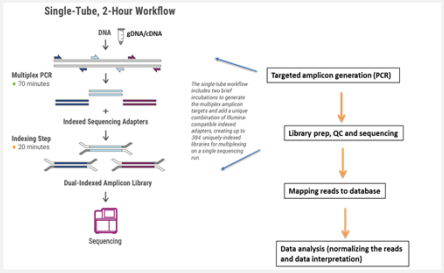
Introduction

High resolution molecular methods have allowed for higher sensitivity for detection and identification of foodborne pathogens. However, the technical challenges with detecting low levels and unculturable pathogens in contaminated food using either whole genome or metagenome sequencing methods still remain. One widely used and affordable NGS application is utilizing target gene-based PCR amplification of specific regions of bacterial 16S gene for community stratification of microbiome or metagenomic samples. We used an analogous approach to amplify conserved genes from bacterial species.
We have designed primers to allow amplification of multiple gene targets (specifically 10 core gene targets) for 266 species that included 135 pathogens in parallel, in one reaction. The advantage of doing it this way is it allows us to test many samples parallelly, thereby reducing cost and saving time when compared to running multiple independent assays.

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). For this study, an STEC strain was inoculated in broth overnight. Modified Bacteriological Analytical Manual was used. Dilutions were made from the overnight culture and 3 different dilutions were inoculated into Whirlpak bags containing 25 grams of romaine lettuce in mBPWp medium. To determine the respective colony forming units (cfu), the same volume of the dilutions was spread on respective Luria-Bertani plates and incubated overnight. The bags were incubated at 37°C with 185 rpm shaking for 30 minutes followed by static incubation for 5 hours. In the next step acriflavine, cefsulodin, and vancomycin were added to the medium followed by overnight incubation at 42°C for 18 h without shaking. Cells were harvested before pre-enrichment (30 minutes), at 5 h and at 24 h post inoculation and stored at −20°C prior to DNA extraction.
The shotgun metagenomics (SM) library preparation was performed following the manufacturer's protocol and the library was sequenced using the Illumina MiSeq Platform. For the TAS library, 10-100ng of DNA was used to set up the multiplex PCR step for 21 cycles. An indexing step followed cleanup with Ampure XP beads. The library was then quantified, and library size and quality were determined using Agilent Bioanalyzer. The PCR amplicons were sequenced using Illumina MiSeq Platform. Kraken2 was used in GalaxyTrakr to identify and classify reads which were then used to generate a Sankey diagram in Pavian. 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 contained 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 is added to a tally for that gene. The tally for each gene was then normalized by dividing by the amplicon length that is represented in the database. Next, 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.

Figure 1. Library preparation workflow



Results and Discussion

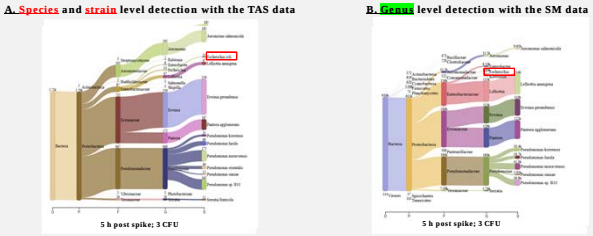
Table 1. Comparison *E. coli* detection using TAS and SM

TAS[a] (CFU/25g sample)			SM[b] (CFU/25g sample)		
3CFU	30CFU	300CFU	3CFU	30CFU	300CFU
30 min					
E. coli	*E. coli, stx2+*	*E. coli, stx2+*	*Escherichia*	*Escherichia*	*Escherichia*
5hr					
E. coli, stx2+	*E. coli, stx2+, ehxA+*	*E. coli, stx2+, ehxA+*	*Escherichia*	*Escherichia*	*E. coli, stx2+, ehxA+*

[a]TAS = Targeted Amplicon Sequencing
[b]SM = Shotgun Metagenomics Sequencing

TAS is at least 10-fold more sensitive at detecting *E. coli* at the strain level compared to SM.

Figure 2.
Sankey Diagram using Pavian shows the sample classification summary using both TAS and SM is similar



TAS is 10-fold more sensitive at detecting *E. coli* at the strain level compared to SM.

A. Sankey Diagram was generated using Pavian shows the sample classification summary; the width of the flow corresponds to the number of reads.
B. Hovering over a species node on the Sankey diagram brings up a barchart with the number of reads for the species across the sample set. *Escherichia coli* is shown here as an example.

Conclusion

- Integrating NGS-based TAS with high-resolution bioinformatic analytical workflow has resulted in a simple, reproducible and rapid assay for simultaneous detection of bacterial and viral pathogens from samples relevant to food-safety.
- When compared to widely used SM method, TAS had an advantage of being more sensitive at 30 min and 5 hours after post-spike to detect the Shiga toxin genes present in *E. coli*.
- TAS method establishes a new threshold of resolution and specificity to quantitate various pathogens in food safety-samples using NGS, genome-wide target design and a tailored bioinformatic analytical workflow.

FDA Mission Relevance
The use of a targeted approach for detecting low amounts of bacterial pathogens provides an efficient and effective tool for the FDA to identify foodborne pathogens such as *Escherichia coli* shown here. This technique may enhance detection of such foodborne pathogens in samples implicated in outbreaks and can be a significant means to reduce labor and time for compliance testing of samples to assure a safe food supply chain particularly for products with a short shelf-life.