Evaluation of Whole Genome Sequencing for Rapid Identification and Molecular Characterization of *Escherichia coli* Isolates from Food and Environment

Nicolas Lopez¹, Claudia Diaz¹, Guodong Zhang², and Li Maria Ma¹

¹Oklahoma State University, Stillwater, OK, ²Food and Drug Administration, College Park, MD

Introduction: Whole genome sequencing is becoming the tool for the detection and characterization of foodborne pathogens for various applications.

Purpose: This study was conducted to evaluate the performance of Minion sequencing for rapid identification and molecular characterization of *E. coli* isolated from food and environment.

Methods: Eleven *E. coli* isolates obtained from pecan orchards were sequenced using Minion and Illumina NextSeq 500. As Minion allows real-time reads analysis, the reads were time-based subsampled to determine the earliest identification turnaround time for each isolate.

Results: Species level identification was achieved at 15 min of sequencing run. In 8 hours of sequencing run, O antigen prediction was available, whereas the antigen H could not be determined, not even after the whole sequencing process was completed (48 hours). Variants of the virulence genes eae and stx were detected in seven isolates at 24 hours of sequencing run. The presence of these virulence genes was confirmed by multiplex PCR. Also, assemblies obtained from the subsampled reads were compared with assemblies obtained from the full reads dataset, as well as hybrid assemblies from the reads from Minion and Illumina. The results showed that the best values of continuity (N50, size of the longer contig, and the entire number of contigs) (p < 0.05) were obtained in just 4 hours of sequencing run. Finally, from a stringent BLASTp search (percentage of identity of 95 % and query coverage of 85 %) against the Comprehensive Antibiotic Resistance Database (CARD) and Virulence Factor Database (VFDB) against the proteins annotated from each of the assembled genomes, we could determine that neither the complete reads datasets were sufficient to generate results significantly similar to those obtained from the hybrid assemblies, nevertheless, it was possible to obtain an average of 87.25% and 75.51% of the hits acquired from the hybrid assemblies using the CARD and VFDB, respectively, where no significant changes were observed after 4 hours compared to the complete datasets (p < 0.05).

Significance: These results demonstrated that Minion can offer an effective tool for the rapid identification of E. *coli* isolates with pathogenic potential, with certain capabilities for their characterization.

Key words: Whole genome sequencing, Minion, Illumina, Shiga toxin-producing E. coli