Evaluation of Early Access NextSeq 2000 2x300 Cycle Sequencing Chemistry Utilizing Datasets of Importance to Food Safety

Christopher Grim1, Phillip Curry1, Ai Kataoka1, Jose Roberto Guzman2, Padmini Ramachandran3, Tina Y. Huang3, Samuel S. Hunter1, Adrienne Hall1, Kelly Hoon3, Sandra M. Tallent3
1Center for Food Safety and Applied Nutrition, Food and Drug Administration, USA; 2Goldbelt G6, LLC., Chesapeake, VA, USA; 3Illuminca, Inc., San Diego, CA

Abstract
Background: Illumina-based short-read genomics has become the community standard for whole genome sequencing (WGS). Over the last five years, read-length has increased from 75 to 150 bp to currently up to 250 bp and the sequencing runs are typically 1500 cycles in length. This has limited our ability to detect and monitor emerging pathogens, however, recently only up to 600 bp paired-end sequencing is now available.
Methods: To assess the application of 600 cycle sequencing on the NextSeq 2000 platform, both bacterial isolate WGS and shotgun metagenomic libraries were used. The WGS sample set was composed of (1) Salmonella enterica serovar Typhi, (2) S. Typhimurium, (3) Mycobacterium tuberculosis, and (4) a direct validation (homogenate) sample. Two runs of NS200 600 cycle sequencing data were generated on the NextSeq 2000 (NS600c and NS600b runs). The MiSeq was run with 300 cycles (NS300) and analyzed for small and small sample assemblies quality metrics of contig and N50. The metagenomic sample (heat and taste samples) were run on NS 100-2x250 cycles and compared with 300 cycles (NS300). Metagenomic datasets were composed according to % checked reads and % reads identified with a target genome.
Results: The read of the average read quality scores for read 1 and 2 was 38 and 29 for NS600 c and 34 and 26 for NS600 b cycle, respectively. Moreover, MiSeq and MiSeq 300 cycle, respectively, better. Moreover, MiSeq was the only run of a 250 cycles to achieve more than 80% of the overall reads being classified. This may be due to the read level for the draft assemblers from the NS600 flow cell run. For the optimal percent of reads being identified, the MiSeq was the only platform that was able to achieve 100% with 100 cycles of sequencing. Regardless of the organism, the N50 value was higher and the observed improved resolution at the read level for the NS600 flow cell run.
Conclusions: The NextSeq 2000 NS 600 c flow cell produced 2x longer base in sequencing quality, which allowed better SNP assemblies and better determination with fewer reads depth for metagenomic datasets. The improved resolution enables higher throughput support applications for food safety.

Experimental Design
- Illumina provided 3: P1 600 cycle kits and 3: P2 600 cycle kits for Early Access testing on NextSeq 2000 (NS200).
- Objective(s): To evaluate NSX 600 cycle kit performance compared to MiX 300 cycle kit performance (2x500 cycles and 2x300 cycles) and performance (2x500 cycles and 2x300 cycles) and performance 6 trials (and evaluation parameters):
  - Isolate WGS - assembly N50 and % contigs
  - 1. P1 run 1: S1 STC1, S1 Salmonella, N5600 and M5500
  - 2. P1 run 1: S1 Salmonella, 75 %; psa5600, NS600, M5500
  - 3. P1 run 2: S1 STC1, S1 Salmonella, N5600 and M5500
  - 4. P1 run 1: S1 Salmonella, CFS1, NS6500, N5600, M5500
  - 5. P2 run 1: S1 STC1, S1 Salmonella, N5600 and M5500
  - 6. P2 run 1: S1 Salmonella, 75 %; psa5600, NS600, M5500
  - Shotgun Metagenomics – taxonomic classification, select target resolution
    - 1. P2 run 1: S1 STC1, S1 Salmonella, N5600 and M5500
    - 2. P2 run 2: S2 scat and soil samples; unenriched, and primary and selective (RF or TT) enrichments for Salmonella
    - 6. P2 run 3: Gene reads

WGS Assembly Data Quality
- Isoate whole genome sequencing de novo assembly metrics by organism for each platform and cycle combination

Sequencing Read Quality
- Representative distribution plots of N50 base-call quality score by cycle number for NextSeq and MiSeq platforms. Plots enclosed in boxes represent runs in which the same DNA Prep libraries were run on each instrument and flow cell.

Conclusions
- NextSeq 600 cycle kits yielded sequencing accuracy quality scores equivalent or better than MiSeq 600 cycle V3 chemistry, especially in late R1 and R2 cycles, with less variability.
- NextSeq 600 cycle kits yielded genome assemblies that were equivalent or better than MiSeq 600 cycle V3 chemistry and were a considerable improvement over NS 300 cycle kits.
- For metagenomic samples, longer read lengths improve taxonomic classification – fewer unclassified reads, and improve diagnostic power – more specific with lower read depth.
- For ICLR, initial libraries were run on NextSeq 1500 platforms.