SINGULAR GENOMICS

Performance Evaluation of the G4™ Sequencing Platform for Microbiome Community Analysis

Ann Tong[1], Chrystal Day[1], Elanor Sievers[1], Sabrina Shore[1], Bubba Brooks[1], Jiangwei Yao[1], Martin M Fabani[1], Timothy Looney[1]
[1]Singular Genomics, 3010 Science Park, Rd, San Diego, California

Introduction

Next-generation sequencing (NGS) has become a central component of modern microbiome basic and translational research by enabling high resolution assessment of species diversity and *de novo* genome assembly at a far lower cost than traditional methods. A key challenge to NGS-based microbiome analysis is the presence of high and low GC content organisms, which are traditionally underrepresented in NGS data. Here, we evaluate the performance of the novel Singular Genomics G4™ Sequencing Platform for microbial community analysis by sequencing of a National Institute for Standards and Technology (NIST) microbial reference material, comparing results to those from the Illumina® NextSeq 2000.

Methods

G4™ Sequencing Platform

The G4 Sequencing Platform is a highly versatile benchtop sequencer that is well suited for demanding research applications. The G4 Platform leverages a novel, 4-color Rapid sequencing by synthesis (SBS) chemistry to deliver highly accurate reads (single or paired-read format with optional index reads) with a single-day turnaround. To maximize flexibility, the G4 Platform enables users to load up to four flow cells at a time, with each flow cell comprising four fluidically independent lanes.

Figure 1 Microbiome workflow and G4 Platform specifications. Library preparation involves incorporation of G4 flow cell adapter sequences and optional sample indices.

G4™ Sequencing Configurations for Metagenomics

	F2 Flow Cell	F3 Flow Cell
Reads Delivered / Flow Cell	200 M	400 M
Reads Delivered / Run	800 M	1,600 M
# Samples / Flow Cell[a]	20	40
# Samples / Run[a]	80	160
Run Time[b]	19-24 hours	
Quality (%Q30)	80-90%	

Throughputs listed are approximations and not guaranteed above kit specifications. Results may vary based on experimental design and sample type. Contact Support for more details.

a. Assumes 10M reads per sample and 2x150 run configuration for shotgun metagenomics[1]
b. Run time includes clustering, sequencing and instrument wash for non-indexed reads

Methods

Library Preparation, Sequencing, and Analysis

Whole genome sequencing libraries were prepared in triplicate for G4 and NextSeq 2000 platforms from 200 ng Covaris-sheared gDNA comprising an equimolar pool of 19 microbial taxa spanning broad GC and phylogenetic diversity (NIST Reference Material 8376; QuantaBio sparQ DNA Library Prep and SG PCR-Free or std Illumina adaptors). With 2x150 bp read format, libraries were sequenced via the G4 Platform and the NextSeq 2000, with the F2 flow cell and P1 flow cell, respectively. Data was downsampled to 1M reads prior to analysis via Kraken[2] v2.1.2, Bracken[3] v2.6.2 and metaphlan4[4]. The F2 flow cell yielded 190M reads with 89.5% and 92.7% of base calls > Q30 for Read 1 and Read 2, respectively.

Results

Accurate and Reproducible Species Quantitation

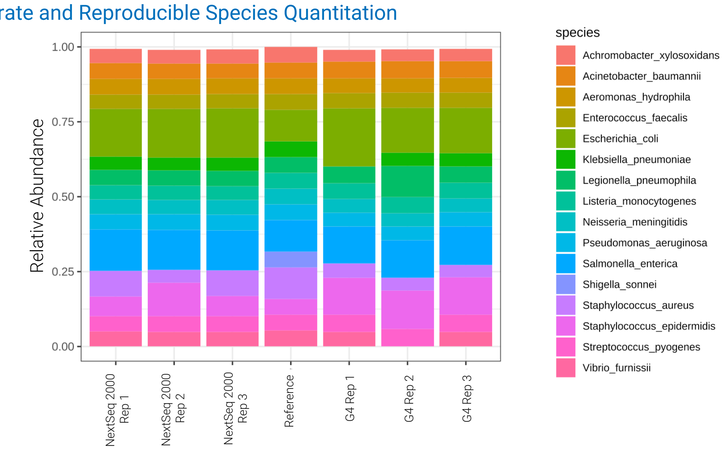

Figure 4 Estimated species abundance for each of three replicate libraries prepared for the G4 and NextSeq 2000. Abundance values derive from Bracken.

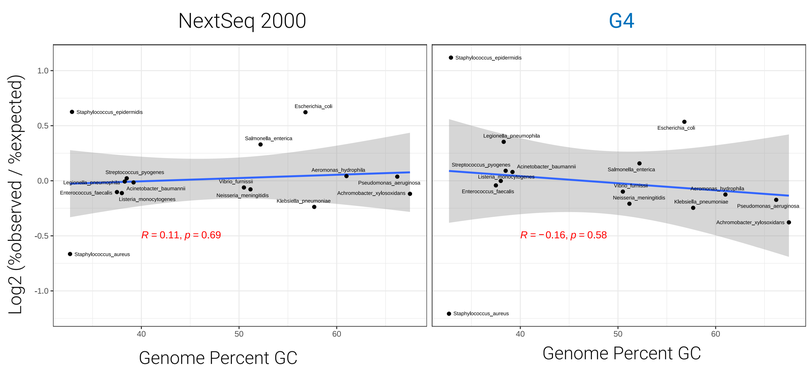

Figure 5 Species abundance as a function of GC content for a representative library sequenced on the G4 or NextSeq 2000; R and p-values derive from Pearson correlation. Neither platform shows a correlation between GC abundance and representation, and the difference in correlations is not significant per the Zou co-correlation test (R 'cocor' library, test='zou2007').

Results

Continued

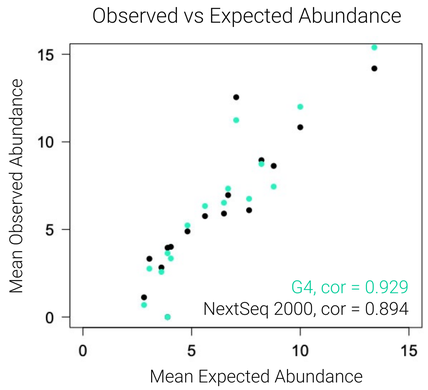

Figure 6 Mean observed vs expected abundance (percent) following sequencing of the 19 species reference material. Abundance values were obtained via metaphlan4. Pearson's correlations are indicated for each platform.

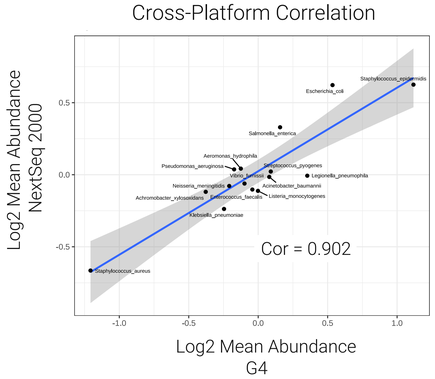

Figure 7 Cross-platform correlation of mean species abundance values following sequencing of the 19 species reference material. *Shigella sonnei* was not detected by either platform owing to the choice of reference database.

Conclusion

In this study we applied the G4 Sequencing Platform to accurately quantify the abundance of diverse bacterial species from a complex mock community sample using off the shelf bioinformatic tools built for the NextSeq 2000.

We observe accurate and reproducible quantification of low and high GC microorganisms using the G4 Platform, with results highly correlated to those from the NextSeq 2000 (cor = 0.984). Notably, the G4 accurately quantified the abundance of both high GC (A. xylosoxidans, 68%) and low GC (S. aureus, 33%) organisms using standard microbial bioinformatics tools without need for modification. We expect the speed and flexible throughput of the G4 Sequencing Platform will help reduce turnaround times in basic and translational microbiome research.

References

1. https://earthmicrobiome.org/protocols-and-standards/
2. Wood and Saltzberg. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15, R46 (2014).
3. Lu et al. Bracken: estimating species abundance in metagenomics data. PeerJ Comp Sci 3:e104. (2017)
4. Blanco-Miguez et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn4. bioRxiv 2022. doi: https://doi.org/10.1101/2022.08.22.504593