#651 (Friday, June 17 2016)

WGS SMRT Sequencing of Patient Samples from a Fecal Microbiota Transplant Trial

C. Heiner[1], S. Oh[1], R. Hall[1], C. Staley[2], M. Hamilton[2], A. Khoruts[3], C. Kelly[4], L. Brandt[5] and M. Sadowsky[2]
[1]PacBio, Menlo Park, CA; [2]BioTechnology Institute, University of Minnesota, St. Paul, MN; [3]Department of Medicine, University of Minnesota, Minneapolis, MN; [4]Department of Medicine, The Warren Alpert Medical School of Brown University, Providence, RI; [5]Department of Medicine, Division of Gastroenterology, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, NY

Introduction

- Fecal samples were obtained from human subjects in the first blinded, placebo-controlled trial to evaluate the efficacy and safety of fecal microbiota transplant (FMT) for treatment of recurrent *C. difficile* infection. Samples included pre-and post-FMT transplant, post-placebo transplant, and the donor control; samples were taken at 2 and 8 week post-FMT.
- Samples were randomly sheared to 2-3 kb fragments for long-read metagenomic profiling, which is described more fully in Richard Hall's poster[1]. Standard SMRTbell libraries were prepared and then sequenced on the new PacBio Sequel System. The resulting kb-long high-quality sequences are used to identify whole genes or gene clusters, with the goal of better understanding the relationship between the composition and functional capabilities of intestinal microbiomes and patient health.
- Here we characterize samples and describe library prep methods and sequencing parameters, including Sequel System run conditions. Descriptive statistics of data output (primary analysis) are presented, along with SMRT Analysis reports on circular consensus sequence (CCS) reads generated using an updated algorithm (CCS2). Final sequencing yields are filtered at various levels of predicted accuracy from 90% to 99.9%.
- These results demonstrate that the Sequel System is well-suited for characterization of complex microbial communities, with the ability for high-throughput generation of extremely accurate single-molecule sequences greater than 1 kb in length. These sequences are a truly random sampling of the community, due to very low sequence-context bias in SMRT Sequencing. With highly accurate kb-long reads, taxonomic identification can be obtained from a few or even a single read that matches a reference sequence. Furthermore, the full-length gene sequences provide functional insight into the community that may not be obtained from 16S sequences, including metabolic pathways present in the community.

[1] Hall, Richard et al., "Workflow for Processing High Throughput Single Molecule Real Time Sequencing Data, for Analyzing the Microbiome of Patients Undergoing Fecal Microbiota Transplantation" Poster #719

Workflow: Library Prep to Analysis

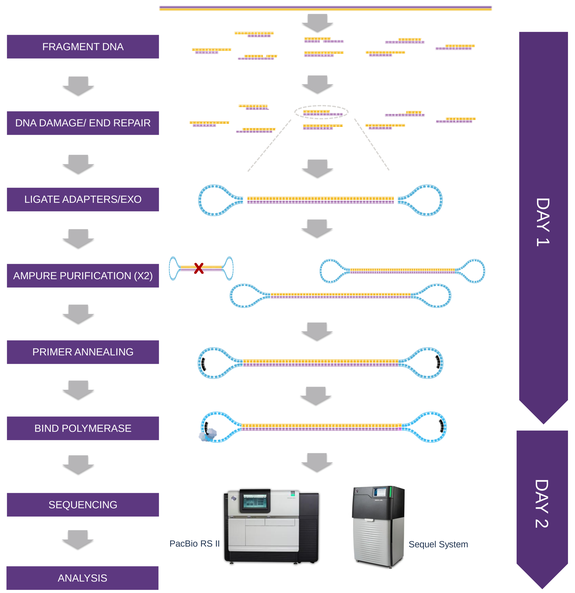


Fig 1. Workflow for long-read metagenomic profiling on PacBio Systems. The entire process from shearing and library prep through sequencing and CCS analysis can be completed in less than 48 hours.

Highly Accurate, Single-Molecule Sequencing

As a function of the SMRTbell adapters, multiple single-pass reads are generated from an individual molecule. Combining subreads corrects for random errors and results in a highly accurate single molecule consensus sequence, with filtering to 99.99%+ predicted accuracy.

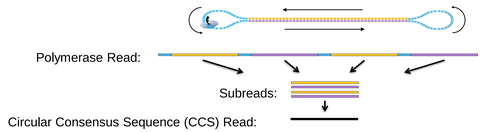
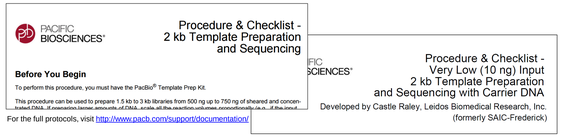


Fig 2. Multiple reads generated from a single molecule.

2 kb SMRTbell Library Prep and Sequencing

- DNA was extracted from fecal samples at various time points before, during, and after FMT, as described in the introduction. At least 300 ng was obtained per sample.



- 2 kb SMRTbell libraries were prepared according to the standard protocol (above left), with the following changes from the 2 kb low input protocol (above right):
  - Half of each sample was sheared to 2 kb and the other half to 3 kb using the Covaris® S2 Focused-ultrasonicator to obtain a broad size distribution. The two shears were pooled in the library.

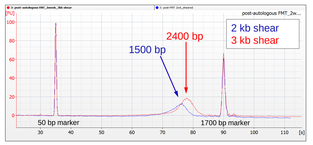


Figure 3. Bioanalyzer® traces of 2 kb and 3 kb shears, sample 3. Both shears were combined in the library for sequencing.

  - DNA Damage Repair was done for 30 minutes; ligation was for 60 minutes.
- Libraries were sequenced on the Sequel System, using Sequel Sequencing Kit v1.1, Mag Bead loading, and 240 or 360 minute data collection.
- One Sequel SMRT Cell 1M was collected per sample.

#	SAMPLE DESCRIPTION	INPUT DNA	AMT POST SHEAR & CLEANUP	FINAL LIBRARY AMOUNT	LIBRARY YIELD, AFTER SHEAR	READS IN 1 SEQUEL SMRT CELL 1M
2	Pre-FMT sample	300 ng	175 ng	35 ng	20.0%	431,714
3	2 weeks following autologous FMT	1.15 µg	810 ng	199 ng	24.6%	617,185
5	2 weeks following open-label FMT	780 ng	498 ng	142 ng	28.5%	472,480
6	8 weeks following open-label FMT	1.62 µg	513 ng	117 ng	22.8%	480,535

Table 1. Amount of input DNA and yields and library prep with corresponding primary analysis yields.
The loss after shearing and AMPure® Bead clean up was presumably due to lower sample quality, typical for fecal samples.

Sequel System Sequencing Results

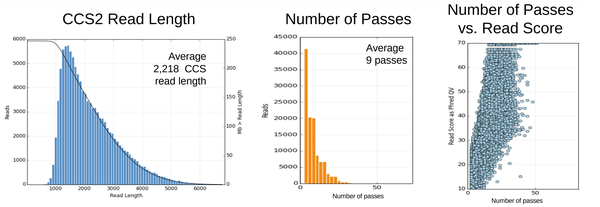
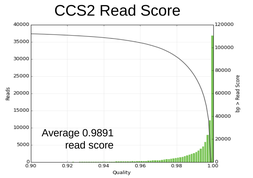


Fig. 4. CCS2 Analysis results from sample 3
After running on Sequel System, SMRT Analysis v3.1.0 was used to filter reads for >3 passes and >90% predicted single-molecule accuracy. Filtering can also be done at higher levels of predicted accuracy, as shown in Table 2 (all with >3 passes). These long, highly accurate reads can be used for profiling down to the species level, sometimes even strain level. With >1 kb reads, 1 or more full-length genes may be included, enabling functional comparison of the samples[1].



	CCS FILTERING CRITERIA (# OF READS / MEAN ACCURACY)		
SAMPLE	90% ACCURATE	99% ACCURATE	99.9% ACCURATE
2	76,996 / 98.83%	52,410 / 99.75%	21,965 / 99.96%
3	113,489 / 98.91%	78,140 / 99.75%	33,252 / 99.97%
5	78,457 / 98.38%	41,668 / 99.66%	11,245 / 99.96%
6	75,124 / 98.55%	42,435 / 99.66%	11,981 / 99.96%

Table 2. CCS2 yields and average predicted accuracies after filtering at various accuracies for each sample.

Conclusions

- Fecal microbiome DNA of average quality may be used to create ~2 kb sheared libraries. With a few hundred ng of input DNA, several hundred thousand reads or more may be generated per sample.
- Using CCS2 analysis, single molecule sequences of >99% accuracy can be obtained, which can be used for taxonomic profiling to the species level, and full-length gene identification for functional analyses.
- With higher throughput than the PacBio RSII, the PacBio Sequel System is ideal for this application, generating hundreds of thousands of highly accurate single molecule sequences per cell. The entire process can be done in just a few days.

The authors would like to thank everyone who helped generate samples and data for the poster