Analysis of Full-Length Metagenomic 16S Genes by Single Molecule, Real-Time Sequencing

Anand Sethuraman, Brett Bowman, Kevin Eng, Cheryl Heiner, Richard Hall
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

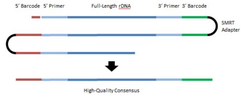
Introduction

- High-throughput sequencing of the complete 16S rRNA gene has become a valuable tool for characterizing microbial communities.
- However, the short reads produced by second-generation sequencing cannot provide taxonomic classification below the genus level.
- In this study, we demonstrate the capability of PacBio's Single Molecule, Real-Time (SMRT®) Sequencing to generate community profiles using mock microbial community samples from BEI Resources.
- We also evaluate multiplexing capabilities using PacBio® barcodes on pooled samples comprising heterogeneous 16S amplicon populations representing soil, fecal, and mock communities.

CCS Sequencing and Analysis Workflow

Circular Consensus Sequence (CCS) Reads

- PacBio's long reads and circularized templates provide high-quality consensus from multiple passes over the same molecule

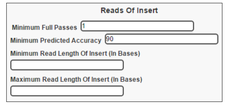


Analysis Workflow

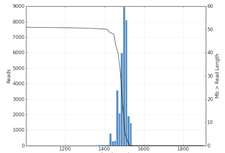
- CCS reads are exported, quality filtered, and where required binned by barcode using the "Reads of Insert" program in SMRT Analysis version 2.3.0.p4.
- Further 16S analysis is carried out using PacBio's rDNA tools pipeline, described at right. CCS sequences undergo further filtering and clustering to generate sequences for classification.
- Output sequences may be classified using a number of available databases.

Accurate Single Molecule 16S Sequences

The PacBio **SMRT Portal Reads of Insert protocol** generates highly accurate CCS reads. Filtering parameters are adjustable according to project needs.



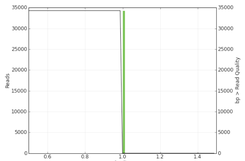
Output of 99.9% predicted accuracy filtering



Read Length of Insert

CCS yield depends on filtering criteria

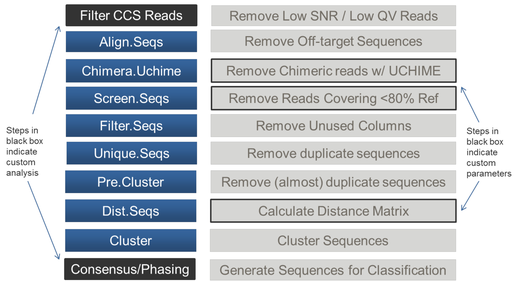
Predicted CCS Accuracy	Full-Length 16S Reads per SMRT Cell
90.0 %	50K
99.0 %	25K
99.9 %	10K



Read Quality of Insert

rDNA Tools Pipeline

Analysis of 16S CCS sequences may be carried out with standard tools from the Mothur package (http://www.mothur.org)[2] combined with Python custom scripts. The complete rDNA tools analysis pipeline with custom scripts is available for download on GitHub: https://github.com/PacificBiosciences/rDnaTools

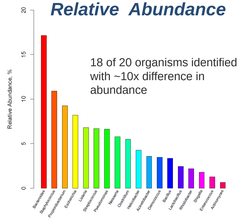


16S Analysis of BEI Mock Communities

- Full-length 16S amplification was performed on two mock microbial community samples from BEI Resources.
  - Microbial Mock Community B, Even, containing DNA from 20 bacterial strains at ~equimolar concentrations
  - Microbial Mock Community B, Staggered, containing DNA from 20 bacterial strains at up to 4 orders of magnitude difference in concentrations
- SMRTbell™ Libraries were prepared following standard protocols.
- A single SMRT Cell was run for each BEI Mock Community using standard P6-C4 chemistry and protocols, with a three-hour collection time.

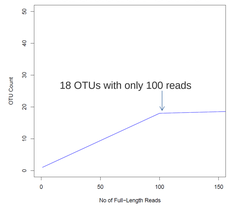
BEI Even Mock Community Analysis from 1 SMRT Cell

Relative Abundance



18 of 20 organisms identified with ~10x difference in abundance

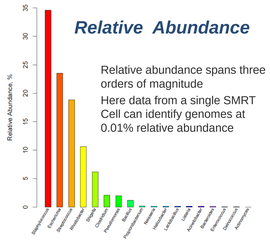
Rarefaction Curve



18 OTUs with only 100 reads

BEI Staggered Mock Community Analysis from One SMRT Cell

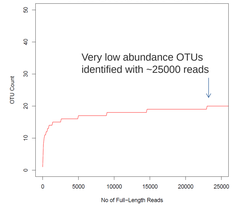
Relative Abundance



Relative abundance spans three orders of magnitude

Here data from a single SMRT Cell can identify genomes at 0.01% relative abundance

Rarefaction Curve

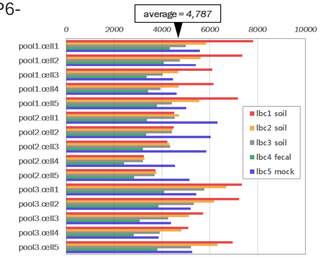


Very low abundance OTUs identified with ~25000 reads

Multiplexing 16S Samples with Barcodes

- 16S amplification was carried out on five metagenomic samples, three soil, one fecal, and one mock community, using PCR primers tailed with PacBio barcodes.
- Equimolar amounts of the five barcoded amplicons were pooled for SMRTbell library preparation using standard protocols.
- Five SMRT Cells were run for each library using standard P6-C4 chemistry and protocols, with 3-hour collection times.

Results of demultiplexing using SMRT Portal Reads of Insert protocol. All samples were represented by >2000 reads/cell with stringent filtering criteria.



Conclusions

- SMRT Sequencing provides accurate, full-length 16S sequences for identification of community constituents to the species level.
- PacBio provides tools for 16S data analysis: Reads of Insert analysis for generation of accurate single-molecule sequences and barcode demultiplexing, and rDNA tools pipeline for 16S-specific filtering and clustering.
- With the PacBio RS II system, genomes at 0.01% relative abundance can be accurately identified with a single SMRT Cell.
- Multiplexing with PacBio barcodes offers cost-effective solutions for 16S sequencing.

References

1. rDnaTools - https://github.com/PacificBiosciences/rDnaTools
2. Schloss, P.D., *et al.*, Introducing mothur: Open-source, platform-independent, community- supported software for describing and comparing microbial communities. Appl Environ Microbiol, 2009. 75(23):7537-41(http://www.mothur.org).
3. Bowman, *et al.*, [2013] Analysis of Full Length Metagenomic 16S Genes by SMRT Sequencing.

Acknowledgements

The authors would like to thank Dr. Jongsik Chun (ChunLab Inc, Seoul National University, Seoul, Korea) and the members of him team for sharing the results of their barcoding experiment.

We also thank Dr . Tanja Woyke, the Microbial Genomics Program Lead at the DOE Joint Genome Institute, for sharing her full-length 16S PCR conditions, which were used for the BEI samples