Introduction

There are many sequencing-based approaches to understanding complex metagenomic communities spanning targeted amplification to whole-sample shotgun sequencing. While targeted approaches provide valuable data at low sequencing depth, they are not limited by primer design and PCR amplification. Whole-sample shotgun experiments generally use short-read, second-generation sequencing, which results in data processing difficulties. For example, reads less than 1 Kbp in length will likely not cover a complete gene or region of interest, and will require assembly. We found the possibility of incompletely combining sequence from different community members, it requires a high depth of coverage. As such, rare community members may not be represented in the resulting assembly.

Circul-consensus, single molecule, real-time (SMRT®) Sequencing reads in the 1-2 kb range, with >99% accuracy can be efficiently generated for low amounts of input DNA. 10 ng of input DNA sequenced in 4hr collection times. Libraries were sequenced using P6/C4 chemistry with 3-hr collection times.

SMRT®Titanium libraries were prepared from whole-sample metagenomes that were each sheared to target size of 2 kb using the SMRTbell template, generating high quality circular consensus sequence (CCS) reads.

Multiple sequencing passes were generated from each SMRT®Titanium template, generating high quality circular consensus sequence (CCS) reads.

Prodigal (Prokaryotic Dynamic Programming Genotyping Algorithm)1, is used to predict genes from individual CCS reads and impulse amino acid sequences.

SEED is a tool to assign to the putative protein sequence to the SEED database.

SEED results are imported into MEGAN® and a Lowest Common Ancestor (LCA) algorithm is used to predict genes from individual CCS reads and impulse amino acid sequences. This allows species level taxonomic classification and functional studies.

Figure 1. Comparison of CCS-gene prediction and other mapping strategies

(a) Distribution and Topology of Taxonomic hits
(b) Mapping Nucleotide Data

(a) Class level comparison with published results (not same sample)

• Using primary samples we show comparable results to both 16S and microarray data, while allowing finer level.

• Using a mock community, we demonstrate concentration comparable to 16S quantification.

• Metagenomic CCS profiling offers distinct advantages over both 16S and shotgun assembly methods.

Colonel FMT

(a) BEI Mock Staggered Community
(b) Stool Sample from a 4 Day Old Healthy Infant

Conclusions

• Metagenomic CCS profiling offers distinct advantages over both 16S and shotgun assembly methods.

• While having a high tolerance for sample input quantities such as low input quantities and fragmented DNA, CCS profiling offers species level taxonomic classification and functional studies.

• Using a mock community, we demonstrate abundance measurements comparable to 16S quantification.

• Using primary samples we show comparable results to both 16S and microarray data, allowing finer grasping species level classification and meaningful functional insight.

References


[5] Hyatt D, et al. Candidate gene discovery: Genes that are seen that are not present post-FMT.

For Research Use Only. Not for use in diagnostic procedures. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell and Ion-Seq are trademarks of Pacific Biosciences of California, Inc. All other trademarks are the property of their respective owners. © 2015 Pacific Biosciences of California, Inc. All rights reserved.