There are many sequencing-based approaches to understanding complex metagenomic communities. Our targeted amplification is whole-genome shotgun sequencing. While targeted approaches provide valuable data at a low sequencing depth, they are limited by primer design and PCR. Whole-genome shotgun sequencing allows the generation of high-quality metagenomes, which results in better assembly outcomes. For example, reads less than 500 bp in length rarely cover a complete gene or region of interest, and will require assembly. This not only introduces the possibility of incorrectly 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.

Circular, unaligned, Single-Read (SMRT) Sequencing reads in a 1-3 kb range, with 90% accuracy can be generated using Pacific Biosciences’ SMRT System. Within SMRT System, use of a consensus, short read sequencing methods, the reads are a true random sampling of the underlying community. SMRT Sequencing has been shown to have very low sequence-context bias. With single-molecule reads, it is highly consensual accuracy, it is reasonable to expect a high percentage of reads to include genes or gene fragments useful for analysis without the need for assembly.

Here we present the results of circular consensus sequencing for an individual’s microbiome, before and after undergoing fecal microbiota transplantation (FMT) in order to treat a chronic Crohn’s disease affecting the gut. We show that even with relatively low sequencing depth, the long-read assembly, free from sequence-context bias, aligns to profile-local abundance community members at the species level. We also show that using SMRT sequencing allows for assembly of functional insight into complex ecosystems and can be used to align the putative protein sequences to the bacterial protein database.

3.66 Blastn Full Recovery

Table 1. Throughout from multiple FMT microorganisms sequences on the Pacific RS II and the Sequel System. The Pre- and Post-FMT samples 3, 5, and 6 are all distinct FMT samples run on the higher throughput Sequel System for comparison. Note that each sample is unique only for the sequencing libraries, which allow for a more nuanced level of classification. The reads were aligned to the Greengenes database yielding more predicted genes. The new Sequel System and Galaxy with a split distribution of ~1.4 to 1.9 can yield >100,000 genes, with >250,000 being functional, as predicted by Prokka (v1.11), and Pseudo (v2.10). The Pacific RS II system yield ~20,000 genes and >50,000 being functional, as predicted by Prokka (v1.11), and Pseudo (v2.10). The Pacific RS II system yield ~20,000 genes and >50,000 being functional, as predicted by Prokka (v1.11), and Pseudo (v2.10).

Introduction

Long-read Metagenomic Profiling Workflow

Figure 1. Analysis workflow for long-read metagenomic profiling. (A) Shotgun genomic DNA with a mean length of > 5 kb is prepared and sequenced on the Pacific System. Multiple sequencing passes are made of the SMRTbilextm template, allowing the generation of high-quality circular consensus sequence reads. (B) PathoQuest (v1.0) is used to predict pathogenic microorganisms and the frequency of secretion pathways is compared to the RefSeq database. (C) BFAST is used to align the GeneSeq protein sequences to the RefSeq bacterial protein database. (D) Blastn is used to align the accurate GeneSeq reads to the RefSeq genomic database. (E) Post-SeqSeq, the reads are imported into MEGAN and a k-mer sequence is used to aid the classification to sequence. (F) Example blast hits for both nucleotide and protein alignments. In this case, the nucleotide classification has more power as the protein sequence is conserved across different species.

Conclusions

References


For Research Use Only. Not for use in diagnostic procedures. © Copyright 2012 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRTBell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BLUEPRINT and SageELF are trademarks of Sage Science. NGS-gp and NGSengine are trademarks of GenDx. All other trademarks are the sole property of their respective owners.