Unbiased characterization of metagenome composition and function using HiFi sequencing on the PacBio Sequel II System

Meredith Ashby, Shreyasee Chakraborty, Janet Ziegle, Joan Wong, Primo Baybayan, Richard Hall
PacBio, 1305 O’Brien Drive, Menlo Park, CA 94025

Abstract

Recent work comparing metagenomic sequencing methods indicates that a comprehensive picture of the taxonomic and functional diversity of complex communities will be difficult to achieve with short-read technology alone. While the lower cost of short reads has enabled greater sequencing depth, the greater contiguity of long-read assemblies and lack of GC bias in SMRT Sequencing has enabled better gene finding. However, since long-read assembly requires high coverage for error correction, the benefits of unbiased coverage have in the past been lost for low abundance species.

SMRT Sequencing performance improvements and the introduction of the Sequel II System has enabled a new, high throughput data type uniquely suited to metagenome characterization: HiFi reads. HiFi reads combine high accuracy with read lengths up to 15 kb, eliminating the need for assembly for most microbiome applications, including functional profiling, gene discovery, and metabolic pathway reconstruction. Here we present the application of the HiFi data type to enable a new method of analyzing metagenomes that does not require assembly.

HiFi Reads on the Sequel II System

Methods and Sequencing Performance

Table 1. Full-length 16S sequencing and shotgun profiling data were collected for mock communities (ATCC® 20 strain staggered (MSA-1002™) and even (MSA-1003™) mixes). For 16S sequencing, V1-V9 amplicons were sequenced on a single cell SMRT Cell 8M at 48 or 96-plex using either a barcoded universal primer / 2-step PCR approach (MSA-1002) or a barcoded 16S primer / 1-step PCR approach (MSA-1003).

<table>
<thead>
<tr>
<th>Sample</th>
<th># contigs</th>
<th># bases</th>
<th>N50 (bp)</th>
<th>SMRT Cell Avg reads / BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human fecal 1</td>
<td>7,043</td>
<td>370,834,836</td>
<td>221,615</td>
<td>11</td>
</tr>
<tr>
<td>Human fecal 2</td>
<td>11,993</td>
<td>583,831,665</td>
<td>123,296</td>
<td>21</td>
</tr>
<tr>
<td>Human fecal 3</td>
<td>4,945</td>
<td>275,125,396</td>
<td>182,086</td>
<td>10</td>
</tr>
</tbody>
</table>

Assembly-Free Gene Finding with FragGeneScan

Table 3. The long read length and high accuracy of HiFi reads means gene discovery can be done efficiently on unassembled metagenome sequences. As a result intact, error-free genes can be found even from species with too little coverage for assembly.

Table 4. Assembly of human fecal microbiome samples with Caru outperforms alternative metagenome assembly approaches.

Table 5. High quality RAST binning results for human fecal sample 2.

Conclusions

- There is high correspondence between 16S and shotgun profiling data and expected mock community compositions, reflecting low context bias of SMRT sequencing technology.
- HiFi shotgun profiling enables the economical recovery of intact genes, operons, and predicted proteins, without the need for assembly.
- HiFi data can be analyzed with standard bioinformatic tools without modification.
- HiFi sequencing on the Sequel II platform provides a new option for functional profiling of microbiome samples.