Low-Input Long-Read Sequencing for Complete Microbial Genomes and Metagenomic Community Analysis
Cheryl Heiner, Steve Oh, and Richard Hall
PacBio, 1380 Willow Road, Menlo Park, CA 94025

Abstract
Microbial genome sequencing can be done quickly, easily, and efficiently with PacBio sequencing instruments, resulting in complete de novo assemblies. Alternative protocols have been developed to reduce the amount of purified DNA required for SMRT Sequencing to broaden applicability to lower-abundance samples. If 50-100 ng of microbial DNA is available, a 10-20 kb SMRTbell library can be made. The resulting library can be loaded onto one or more SMRT Cells, yielding more than enough data for complete assembly of microbial genomes using the SMRT Portal assembly pipeline HGAP, plus epigenetic analysis. The entire process can be done in 2 to 3 days by standard laboratory personnel.

This approach is particularly important for analysis of microbial communities, in which genomic DNA is often limited. For simple microorganisms, it may be possible to obtain complete or near complete genome sequences of abundant members with low-input 10 kb libraries. For more complex communities, or when only a few ng of input DNA are available, a 2 kb sheared library can be prepared to generate 1-2 kb reads with >99.9% accuracy with Circular Consensus Sequencing. These very accurate, single-molecule reads can be used to provide information about the microbial composition of a community, as well as biochemical pathways present in the sample.

Library Prep Options for Low-Input Sequencing

<table>
<thead>
<tr>
<th>Library Size</th>
<th>Protocol</th>
<th>Input</th>
<th>SMRT Cells</th>
<th>Total Bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 kb</td>
<td>Standard</td>
<td>1 µg (minimum)</td>
<td>≥150</td>
<td>&gt;90 Gb</td>
</tr>
<tr>
<td>10 kb</td>
<td>Low Input</td>
<td>100 ng</td>
<td>≥4</td>
<td>&gt;2 Gb</td>
</tr>
<tr>
<td>2 kb</td>
<td>Standard</td>
<td>500 ng (minimum)</td>
<td>≥400</td>
<td>&gt;240 Gb</td>
</tr>
<tr>
<td>2 kb</td>
<td>Very Low Input</td>
<td>10 ng</td>
<td>≥4</td>
<td>&gt;2 Gb</td>
</tr>
</tbody>
</table>

Results for the PacBio RS II, with conservative estimates of library and SMRT Cell yield.

2 kb SMRTbell Libraries from 10 ng Input DNA

MICROBIOME PROFILING
DNA was purified from an environmental (lake) sample and prepared for sequencing using the 2 kb, very low input SMRTbell Library Prep Protocol. Data was used to determine genomes, which were aligned to a reference database.

A. Sheared and bead-purified samples

B. Primary analysis of PacBio data, 2 kb sheared lake microbiome samples

C. Circular consensus Reads of Insert yield, SMRT Analysis

D. Taxonomic groups found in the lake microbiome sample

Bioinformatic Workflow for Profiling Microbial Communities

MULTIPLEXING EXTREMELY LOW INPUT SAMPLES
For sub-ng inputs, a small number of highly accurate 1-2 kb sequences may be obtained by pooling libraries with barcoded adapters. Here is an example of where an undetectable amount of DNA was pooled with a sample with ~4 ng input following ligation of barcoded adapters

Profile results for six PacBio RS II SMRT Cells; all 3 SMRT Cells had similar results.

10-20 kb SMRTell Libraries from 100 ng Input DNA

10-20 KB LOW-INPUT LIBRARY FROM A SIMPLE COMMUNITY: BIOCATHODE DNA
DNA was isolated from biofilm collected on an electrode. SMRTell libraries were prepared using the 10-20 kb Low-Input Protocol (125 ng input) or the 10-20 kb Template Preparation and Sequencing protocol (1.25 µg input).

A. QC of sheared biofilm DNA prior to SMRTell library prep

B. Primary analysis of PacBio data, 10 kb sheared biofilm samples

C. PacBio RS II sequencing yield

D. Genome assembly of the two most prevalent species

Results
• Microbial genomes have been assembled from low input amounts (100 ng) of DNA using the 10–20 kb–20 kb SMRTbell library prep protocol, sequenced on the PacBio RS II.
• PacBio community profiling can be done with very low amounts (10 ng or less) of DNA from microbiome samples prepared with the 2 kb, very low input SMRTbell library prep. Additionally, samples with sub-ng amounts of DNA may yield a small number of very accurate reads when multiplexed with samples with a few ng.

Acknowledgements
The authors would like to thank:
• Dr. Sarah Strycharz-Gray, Research Biologist at the Naval Research Laboratory, for the biofilm samples.
• Dr. Tanja Woyke, The Microbial Genomics Program Lead at the DOE Joint Genome Institute, for the lake, plant and mock metagenomic samples.