# Low-Input Long-Read Sequencing for Complete Microbial Genomes and Metagenomic Community Analysis

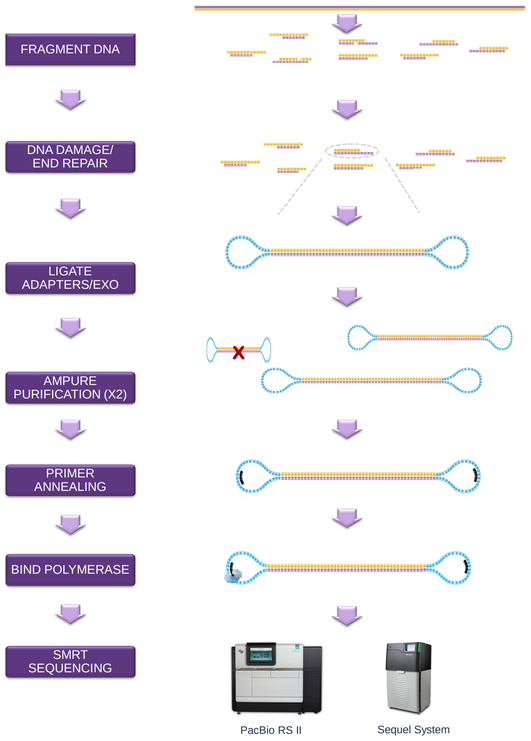
Cheryl Heiner, Steve Oh, and Richard Hall
PacBio, 1380 Willow Road, Menlo Park, CA 94025

## 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 microbiomes, 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.

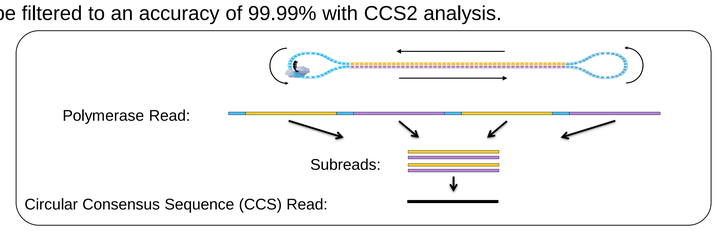
## SMRTbell Library Prep Workflow




## Highly Accurate Single-Molecule Sequencing

### MULTIPLE READS FROM A SINGLE MOLECULE

As a function of the SMRTbell adapters, multiple single-pass reads are generated from an individual molecule. Combining these subreads corrects for random errors and results in a highly accurate single-molecule consensus sequence. Data can be filtered to an accuracy of 99.99% with CCS2 analysis.




## Library Prep Options for Low-Input Sequencing

| Library Size | Protocol | Input | SMRT Cells | Total Bases |
|---|---|---|---|---|
| 10 kb | Standard | 1 µg (minimum) | ≥150 | ≥90 Gb |
| 10 kb | Low Input | 100 ng | ≥4 | ≥2.4 Gb |
| 2 kb | Standard | 500 ng (minimum) | ≥400 | ≥240 Gb |
| 2 kb | Very Low Input | 10 ng | ≥4 | ≥2.4 Gb |

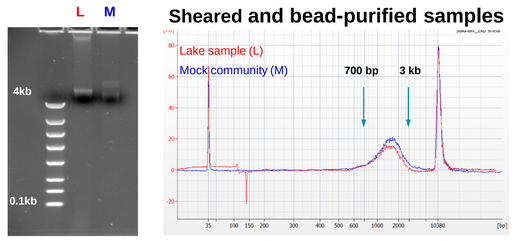
Results for the PacBio RS II, with conservative estimates of library and SMRT Cell yield.

## 2 kb SMRTbell Libraries from 10 ng Input DNA

### MICROBIOME PROFILING

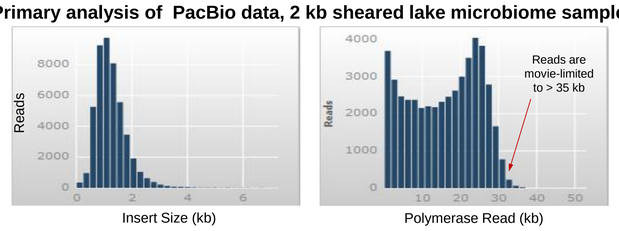
DNA was purified from an environmental (lake) sample and prepared for sequencing using the 2 kb, very low input SMRTbell Library Prep Protocol. Data was used to determine genes, which were aligned to a reference database.

**A**




Lonza FlashGel and Bioanalyzer QC of sheared microbiome DNA prior to SMRTbell library prep

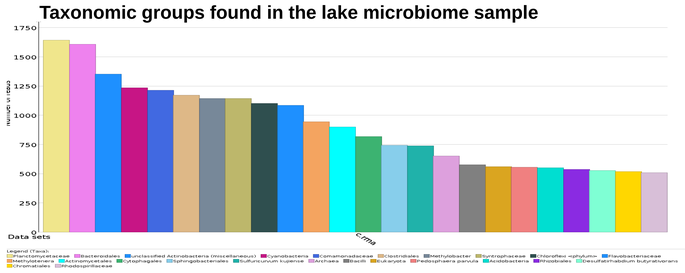
**B** Primary analysis of PacBio data, 2 kb sheared lake microbiome samples




**C** Circular consensus Reads of Insert yield, SMRT Analysis

| Sample | Primary | 90% Accuracy | | 99% Accuracy | |
|---|---|---|---|---|---|
| | P1 Reads | Total Bases | # of Reads | Total Bases | # of Reads |
| Lake Microbiome | 90 K | 74 Mb | 64 K | 56 Mb | 48 K |
| Mock Community | 114 K | 90 Mb | 82 K | 66 Mb | 60 K |

**D** Taxonomic groups found in the lake microbiome sample




### Bioinformatic Workflow for Profiling Microbial Communities

**CCS** → **Pro**karyotic **D**ynamic Programming **G**enefinding **Al**gorithm (**Prodigal**) → **BLASTp** → **MEGAN** Taxonomy assigned

### 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

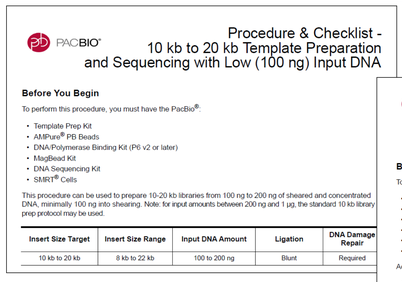
| | Input DNA (ng) | Bar-code | # Cells | Total Bases | P1 Reads | CCS Reads (≥99% accuracy) |
|---|---|---|---|---|---|---|
| Sample 1 | 3.9 | 34 | 3 | 2.8 Gb | 195,876 | 104,631 |
| Sample 2 | Un-detectable* | 51 | | | | 71 |

*With Qubit dsDNA High Sensitivity Assay

**Profiling results**
- Sample 1 and Sample 2 communities were quite different
- Sample 2 reads included perfect to near-perfect kb-long alignments


**Primary analysis results** for one PacBio RS II SMRT Cell; all 3 SMRT Cells had similar results.

## 2 - 20 kb Low-Input SMRTbell Library Protocols



PACBIO®
Procedure & Checklist - 10 kb to 20 kb Template Preparation and Sequencing with Low (100 ng) Input DNA

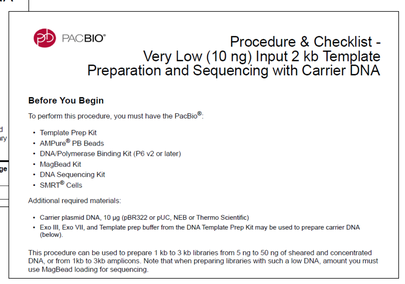
**Before You Begin**

To perform this procedure, you must have the PacBio®:

- Template Prep Kit
- AMPure® PB Beads
- DNA/Polymerase Binding Kit (P6 v2 or later)
- MagBead Kit
- DNA Sequencing Kit
- SMRT® Cells

This procedure can be used to prepare 10-20 kb libraries from 100 ng to 200 ng of sheared and concentrated DNA, minimally 100 ng into shearing. Note: for input amounts between 200 ng and 1 µg, the standard 10 kb library prep protocol may be used.

| Insert Size Target | Insert Size Range | Input DNA Amount | Ligation | DNA Damage Repair |
|---|---|---|---|---|
| 10 kb to 20 kb | 8 kb to 22 kb | 100 to 200 ng | Blunt | Required |



PACBIO®
Procedure & Checklist - Very Low (10 ng) Input 2 kb Template Preparation and Sequencing with Carrier DNA

**Before You Begin**

To perform this procedure, you must have the PacBio®:

- Template Prep Kit
- AMPure® PB Beads
- DNA/Polymerase Binding Kit (P6 v2 or later)
- MagBead Kit
- DNA Sequencing Kit
- SMRT® Cells

Additional required materials:

- Carrier plasmid DNA, 10 µg (pBR322 or pUC, NEB or Thermo Scientific)
- Exo III, Exo VII, and Template prep buffer from the DNA Template Prep Kit may be used to prepare carrier DNA (below).

This procedure can be used to prepare 1 kb to 3 kb libraries from 5 ng to 50 ng of sheared and concentrated DNA, or from 1kb to 3kb amplicons. Note that when preparing libraries with such a low DNA, amount you must use MagBead loading for sequencing.

For the full protocols, visit
www.pacb.com/support/documentation

## 10 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. SMRTbell libraries were prepared using the 10-20 kb Low-Input Shared Protocol (125 ng input) or the standard 10 kb Template Preparation and Sequencing protocol (1.25 µg input).

**A. QC of sheared biofilm DNA prior to SMRTbell library prep**




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




**C. PacBio RS II sequencing yield**

| Sample | Number of SMRT Cells[1] | Number of P1[2] Reads | Total Bases | Avg Insert Length | Avg Polymerase Read Length |
|---|---|---|---|---|---|
| Low (125 ng) Input | 2 | 97 K | 1.3 Gb | 4.6 kb | 13.7 kb |
| Std (1.25 µg) Input | 7 | 579 K | 9.2 Gb | 5.2 kb | 15.9 kb |

[1]Did not sequence entire library [2]P1 reads contain usable sequence information

**D. Genome assembly of the two most prevalent species**







**Far left, species 1 genome:** a complete circular genome; coverage wave indicates active replication, with origin of replication at peak[1].

**Left, plasmid**: also circular; plasmid base modification is identical to species 1 genome.




**Left, species 2**: a second circular genome with flat coverage, indicating much slower growth.

**Reference**
[1] 1. Korem, T., Zeevi, D., Suez, J., et al. (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomics sample. *Science* doi: 10.1126/science.aac4812

## Results

- Microbial genomes have been assembled from low input amounts (100 ng) of DNA using the 10 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 protocol. 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-Glaven, Research Biologist at the Naval Research Laboratory, for the biocathode samples.
- Dr. Tanja Woyke, the Microbial Genomics Program Lead at the DOE Joint Genome Institute, for the lake, plant and mock metagenomic samples.