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

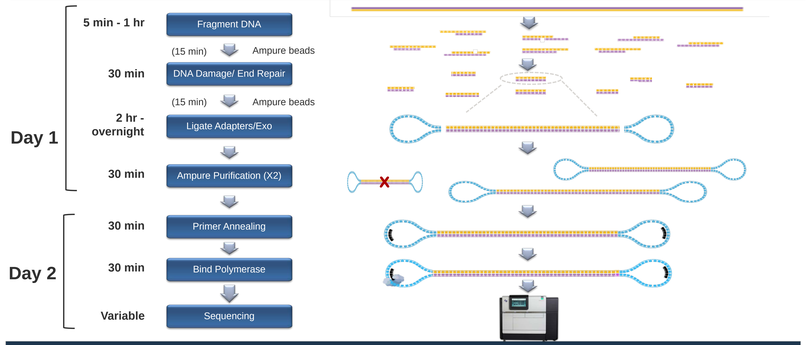
Cheryl Heiner, Steve Oh, Kevin Eng, and Richard Hall; Pacific Biosciences, Menlo Park, CA
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

Abstract

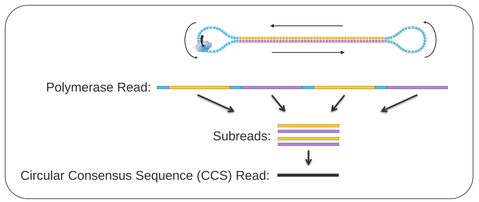
Microbial genome sequencing can be done quickly, easily, and efficiently with the PacBio® sequencing instruments, resulting in complete *de novo* assemblies. Alternative protocols have been developed to reduce the amount of purified DNA required for SMRT® Sequencing, to broaden applicability to lower-abundance samples. If 50-100 ng of microbial DNA is available, a 10-20 kb SMRTbell™ library can be made. A 2 kb SMRTbell library only requires a few ng of gDNA when carrier DNA is added to the library. The resulting libraries can be loaded onto multiple SMRT Cells, yielding more than enough data for complete assembly of microbial genomes using the SMRT Portal assembly program HGAP, plus base-modification analysis. The entire process can be done in less than 3 days by standard laboratory personnel.

This approach is particularly important for the analysis of metagenomic communities, in which genomic DNA is often limited. From these samples, full-length 16S amplicons can be generated, prepped with the standard SMRTbell library prep protocol, and sequenced. Alternatively, a 2 kb sheared library, made from a few ng of input DNA, can also be used to elucidate the microbial composition of a community, and may provide information about biochemical pathways present in the sample. In both these cases, 1-2 kb reads with >99% accuracy can be obtained from Circular Consensus Sequencing.

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 99.9%.

Library Prep Options for Low-Input Sequencing

Library Size	Input Requirement	# SMRT Cells	Total Bases*	Average Insert Size
2 kb	10 ng	2 cells	1.9 Gb	1.5 kb
10 kb	100 ng	4 cells	2.4 Gb	4.5 kb

* From Primary Analysis

2 kb SMRTbell Libraries from 10 ng Input DNA

A. 2 kb Low-Input Shared Protocol

PACIFIC BIOSCIENCES®

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

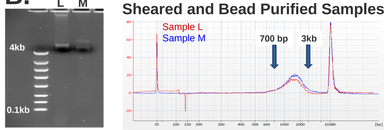
Developed by Castle Raley, Leidos Biomedical Research, Inc. (formerly SAIC-Frederick)

For the full protocol, visit https://pacbio.secure.force.com/Share/Protocol/List

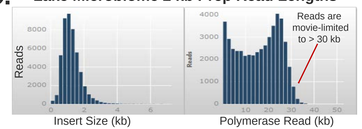
Microbiome Profiling

DNA was purified from an environmental (lake) sample and prepared for sequencing using this Shared Protocol. Data was used to determine genes in microbial constituents as described in poster 2544: "Profiling Metagenomic Communities Using Circular Consensus and Single Molecule, Real-Time Sequencing".

B. Sheared and Bead Purified Samples



C. Lake Microbiome 2 kb Prep Read Lengths



D. Sequencing Yield from 2 kb Libraries

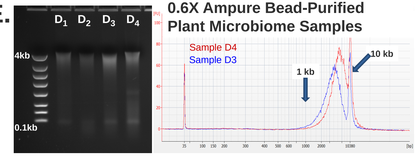
Sample	Primary	90% Accuracy		99% Accuracy	
	P1* Reads	Total Bases	# of Reads	Total Bases	# of Reads
Lake microbiome	90 K	74 Mb	64 K	56 Mb	48 K
Mock community	114 K	90 Mb	82 K	66 Mb	60 K

*P1 = Reads that contain usable sequence information

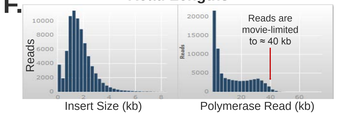
2 kb Libraries from Degraded Samples

- SMRTbell libraries from partially degraded samples can be successfully sequenced on PacBio instruments. Shearing is not necessary when input DNA is already fragmented to the desired size or smaller.
- Degraded samples are likely to contain many short fragments that can dominate loading. These fragments may be removed using an appropriate concentration of Ampure PB beads.

E. 0.6X Ampure Bead-Purified Plant Microbiome Samples



F. Read Lengths



A 2 kb library was prepared from 20 ng of degraded input DNA (sample D3), yielding 77,000 P1 reads from 1 SMRT Cell.

10 kb SMRTbell Libraries from 100 ng Input DNA

A. 10 kb Low-Input Shared Protocol

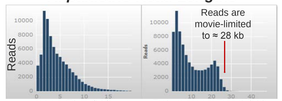
PACIFIC BIOSCIENCES® PacBio SampleNet – Shared Protocol

Please note: the shared protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® system. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

10 kb to 20 kb Template Preparation and Sequencing with Low-Input DNA

For the full protocol, visit https://pacbio.secure.force.com/Share/Protocol/List

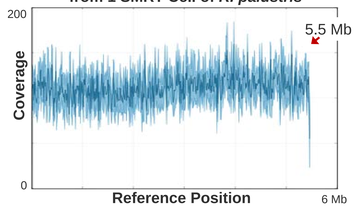
B. *R. palustris* Read Lengths



C. Sequencing Yield from 10 kb Prep of *R. palustris*

Library Size	Input	# SMRT Cells	Total Bases	Average Coverage
10 kb	100 ng	1 cell	813 Mb	110 X

D. Complete Genome Assembly from 1 SMRT Cell of *R. palustris*



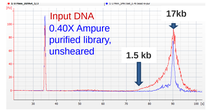
Genome Assembly

100 ng of *Rhodopseudomonas palustris* genomic DNA was prepared according to the 10 kb – 20 kb low-input protocol. Reads were assembled using PacBio RS_HGAP_Assembly3.

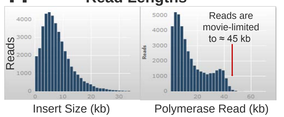
10 kb Libraries from Degraded Samples

- A 10 – 20 kb library was prepared from 500 ng of unsheared, degraded DNA and used for genome assembly

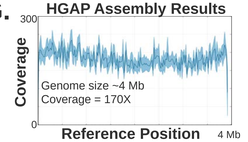
E. Iron Mine Microbiome



F. Read Lengths



G. HGAP Assembly Results



H. Sequencing Yield from 10 kb Prep of Iron Mine Microbiome

Input	# SMRT cells	Total Bases	Microbe	Average Coverage
500 ng	1	1 Gb	1	170X

Note: sufficient library was produced for >8 SMRT Cells at this loading

Conclusions

- Community profile information has been obtained from very low amounts of DNA of microbiome samples prepared with the 2 kb, Very Low Input (10 ng) shared protocol and sequenced on the PacBio® RS II
- Microbial genomes have been assembled from low inputs using the 10 kb – 20 kb shared protocol
- SMRTbell libraries can be constructed and sequenced from low inputs (20 – 500 ng) of degraded samples

Acknowledgements

The authors would like to thank Dr. Tanja Woyke, the Microbial Genomics Program Lead at the DOE Joint Genome Institute, for the lake, plant and mock metagenomic samples.
We also thank Dr. Jon Badalamenti, University of Minnesota, for the iron mine metagenomic sample.