SATURDAY- #503 (June 3, 2017)

Multiplexing Strategies for Microbial Whole Genome Sequencing Using the Sequel System

Christine Lambert, John Harting, and Primo Baybayan
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Introduction

For microbial sequencing on the PacBio Sequel System, the current yield per SMRT Cell is in excess relative to project requirements. Multiplexing offers a viable solution; greatly increasing throughput, efficiency, and reducing costs per genome. This approach is achieved by incorporating a unique barcode for each microbial sample into the SMRTbell adapters and using a streamlined library preparation process.

To demonstrate performance,12 unique barcodes assigned to *B. subtilis* and sequenced on a single SMRT Cell. To further demonstrate the applicability of this method, we multiplexed the genomes of 16 strains of *H. pylori*. Each DNA was sheared to 10 kb, end-repaired and ligated with a barcoded adapter in a single-tube reaction. The barcoded samples were pooled in equimolar quantities and a single SMRTbell library was prepared.

Successful *de novo* microbial assemblies were achieved from all multiplexes tested (12-, and 16-plex) using data generated from a single SMRTbell library, run on a single SMRT Cell 1M with the PacBio Sequel System, and analyzed with standard SMRT Analysis assembly methods.

Here, we describe a protocol that facilitated the multiplexing up to 12-plex of microbial genomes in one SMRT Cell 1M on the Sequel System that produced near-complete microbial *de novo* assemblies of <10 contigs for genomes <5 Mb in size.

Workflow for Multiplexing Microbial Genomes

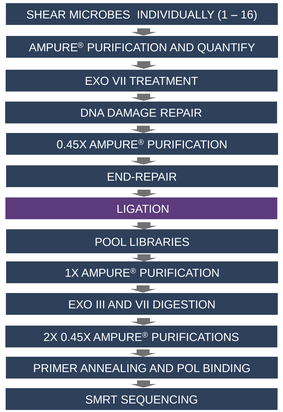
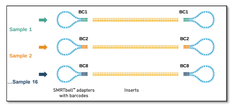


Fig. 1. Workflow for microbial multiplexing SMRTbell library preparation

A



B

Recommended Barcoded Adapters	
BC1002	BC1070
BC1004	BC1080
BC1016	BC1093
BC1032	BC1100
BC1048	BC1101
BC1054	BC1109
BC1055	BC1115
BC1063	BC1118

Fig. 2. SMRTbell barcode workflow and recommended adapters

(A) Barcodes are added to the SMRTbell adapters and incorporated during the Ligation reaction. **(B)** Recommended barcoded adapters

Materials and Methods

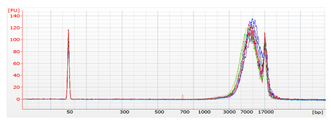


Fig. 3. Bioanalyzer traces of twelve strains of sheared *B. subtilis* gDNA used in an 12-plex library

- For most microbial genomes, recommend shearing genomic DNA to 10 kb
- To minimize bias, genomic DNA shears should have similar size distribution
- Shear genomic DNA to 10 kb with Covaris g-TUBEs
- Genomes with large repeats regions may not be suitable for multiplexing approach and may require 20 kb or 30 kb library preparation with size selection

- Recommend shearing 1 µg of DNA per sample to account for sample loss during shearing and purification
- Sample input DNA for Exo VII depends on the multiplex level
- For example: ~125 ng per microbe for 8-plex for post-shearing steps, assuming similar fragment distribution
- Use equimolar amounts if shear distribution is variable

MULTIPLEX	DNA INPUT INTO SHEARING	DNA INPUT INTO EXO VII*
2-plex	1.0 µg each	500 ng each
4-plex	1.0 µg each	250 ng each
6-plex	1.0 µg each	167 ng each
8-plex	1.0 µg each	125 ng each
10-plex	1.0 µg each	100 ng each
12-plex	1.0 µg each	83 ng each
16-plex	1.0 µg each	63 ng each

*Assuming fragment distribution for each microbe is the same

Table 1. Multiplexing DNA input recommendations

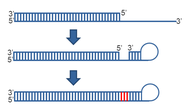


Fig. 4. Remove 3' overhangs by treating the sheared DNA with Exo VII

SAMPLES	TOTAL SUBREADS	PALINDROMIC SUBREADS	% PALINDROMES	# CONTIGS
B. subtilis(-) Exo VII	205,684	23,689	11.5	204
***B. subtilis*(+) Exo VII**	**300,300**	**4,217**	**1.4**	**1**
E.coli (-) Exo VII	145,000	15,866	11.0	101
***E.Coli* (+) Exo VII**	**347,387**	**7,101**	**2.0**	**5**

Table 2. Comparison of Exo VII treated and non-treated DNA shear

- The number of contigs is reduced for treated DNA shears
- Exo VII treatment of sheared DNA is necessary to reduce "missed" adapters sequence artefacts

Fig. 5. Informatics pipeline for microbial multiplexing (SMRT Analysis v4.0)

1. Samples were sequenced on the Sequel System with 2.0 chemistry and 600 minute movies.
2. To score and align each read with their barcodes, SMRT Cells were processed using Barcoding protocol in SMRT Link (GUI). This protocol scores and aligns barcodes against all potential barcodes and return the highest scoring for each molecule.
3. One subreadset xml per barcode was generated using the dataset split --barcodes tool (CL) and re-import each to SMRT Link.
4. *De novo* assembly was run in SMRT Link (GUI) using HGAP 4 for each imported subreadset from step 3.

http://www.pacb.com/wp-content/uploads/Analysis-Procedure-Multiplexed-Microbial-Assembly-SMRT-Link.pdf

Results

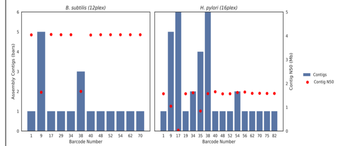
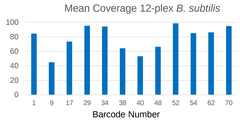


Fig. 6. Assembly contig N50s and count of assembly contigs from a 12-plex pool of 4 Mb *B. subtilis* (left) and a 16-plex pool of 1.6 Mb *H. pylori* (right)



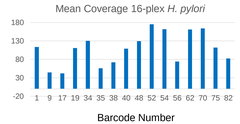


Fig. 7. Coverage per barcode of a 12-plex pool of *B. subtilis* and a 16-plex pool of *H. pylori*

- A single Sequel SMRT Cell 1M multiplexed 16 microbes (<2 Mb) or 12 microbes (<5 Mb) and achieved near-complete assemblies (<10 contigs)
- For HGAP assembly, recommend targeting 50-fold coverage for optimal assembly results
- If improved assembly performance is required:
  - Sequence additional SMRT Cells to increase coverage
  - Decrease multiplex level to increase per genome coverage
  - Sequence samples independently; Increase insert size and use size selection to increase likelihood of spanning repeat regions
- Base modification and motif analysis can be performed with multiplexed microbial libraries
- Kinetic detection distribution similar for single-plex and multiplexed libraries
- Target 100-fold coverage for base modification and motif analysis

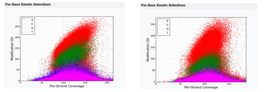


Fig. 8. Kinetic detection distribution comparison of multiplexed and single-plex libraries of *H. pylori* strain 76

Conclusion

- Multiplexing maximizes throughput and efficiency of microbial whole genome sequencing, while reducing cost per sample
  - Decrease template preparation cost and sample processing time by ~30%
  - Reduce sequencing cost by >50%
- Multiple near-complete genome assemblies from microbial samples can be achieved in 1 Sequel SMRT Cell 1M
- Recommend up to 12 microbes (<5 Mb) or 16 microbes (<2 Mb) per Sequel SMRT Cell 1M as a guideline for designing microbial WGS multiplexing experiments on the Sequel System
- Microbes with large repeat regions may need 20 kb or 30 kb library preparation with size selection; If the goal is complete genome assemblies, recommend sequencing genomes independently

Acknowledgements

- The authors would like to thank: Mutsuko Konno, JA Sapporo Hospital; Hirokazu Yano,University of Tokyo & Tsukuba University; Masaki Fukuyo, Chiba University; Ichizo Kobayashi, University of Tokyo, Kyorin University, & University of Paris – Saclay, for the *H.pylori* samples.