#716 June 19th, 2016

Multiplexing Strategies for Microbial Whole Genome SMRT Sequencing

Christine Lambert, John Harting, Walter Lee, and Primo Baybayan
PacBio, 1380 Willow Road, Menlo Park, CA 94025

Introduction

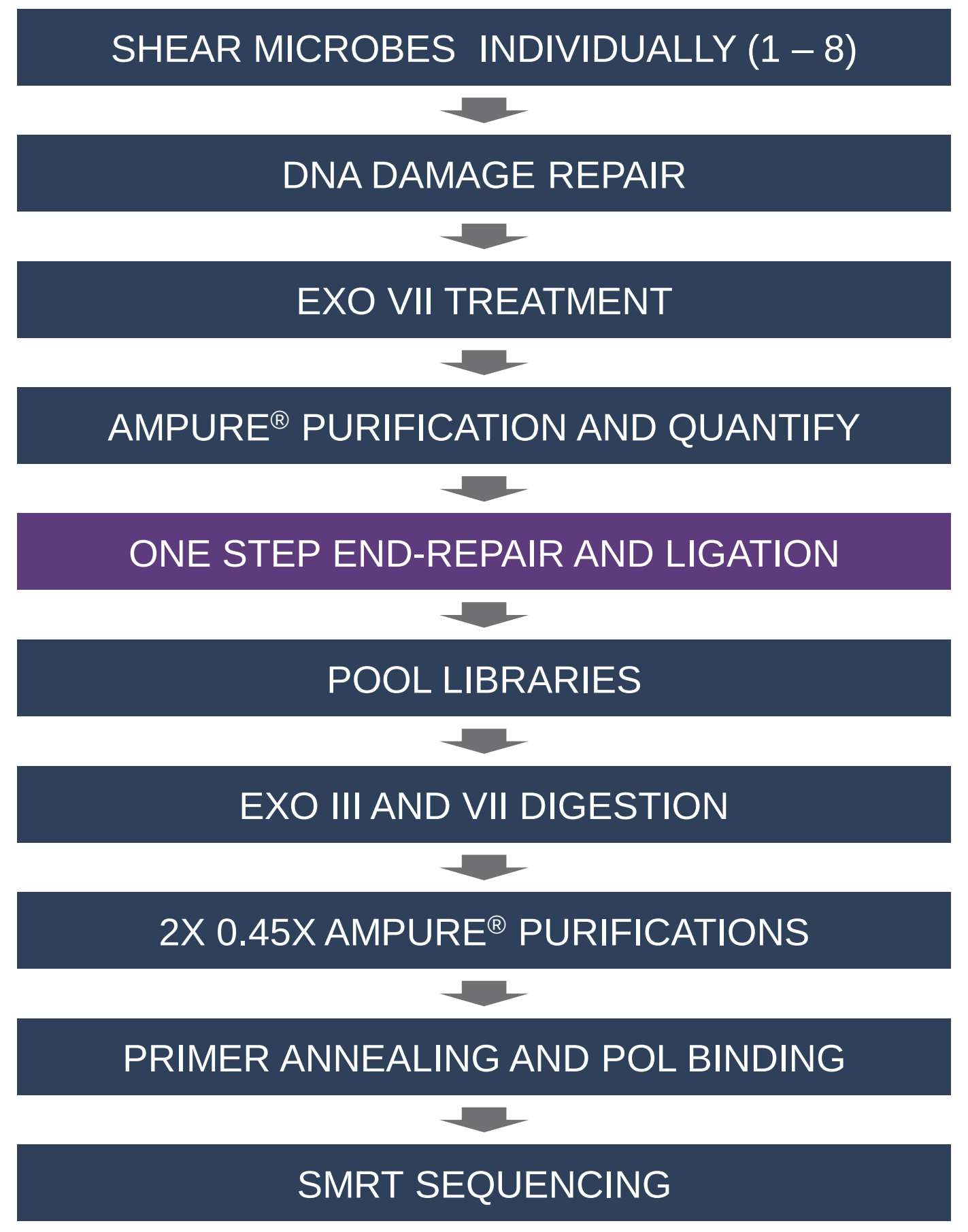
As the throughput of the PacBio Systems continues to increase, so has the desire to fully utilize SMRT Cell sequencing capacity to multiplex microbes for whole genome sequencing. Multiplexing is readily achieved by incorporating a unique barcode for each microbe into the SMRTbell adapters and using a streamlined library preparation process. Incorporating barcodes without PCR amplification prevents the loss of epigenetic information and the generation of chimeric sequences, while eliminating the need to generate separate SMRTbell libraries.

We multiplexed the genomes of up to 8 unique strains of *H. pylori*. Each genome was sheared and processed through adapter ligation in a single, addition-only reaction. The barcoded samples were pooled in equimolar quantities and a single SMRTbell library was prepared.

We demonstrate successful *de novo* microbial assembly from all multiplexes tested (2- through 8-plex) using data generated from a single SMRTbell library, run on a single SMRT Cell with the PacBio RS II, and analyzed with standard SMRT Analysis assembly methods.

This strategy was successful using both small (1.6 Mb, *H. pylori*) and medium (5 Mb, *E. coli*) genomes. This protocol facilitates the sequencing of multiple microbial genomes in a single run, greatly increasing throughput and reducing costs per genome.

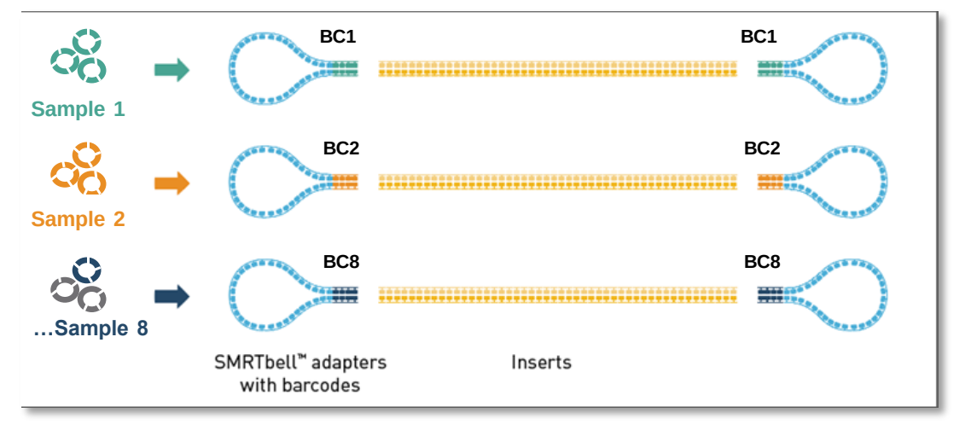
Workflow for Multiplexing Microbial Genomes



SHEAR MICROBES INDIVIDUALLY (1 – 8)
DNA DAMAGE REPAIR
EXO VII TREATMENT
AMPURE® PURIFICATION AND QUANTIFY
ONE STEP END-REPAIR AND LIGATION
POOL LIBRARIES
EXO III AND VII DIGESTION
2X 0.45X AMPURE® PURIFICATIONS
PRIMER ANNEALING AND POL BINDING
SMRT SEQUENCING

Fig. 1. Workflow for microbial multiplexing SMRTbell library preparation

A



B

lbc1--lbc1	lbc48--lbc48
lbc17--lbc17	lbc52--lbc52
lbc29--lbc29	lbc54--lbc54
lbc34--lbc34	lbc62--lbc62
lbc38--lbc38	lbc70--lbc70
lbc40--lbc40	lbc9--lbc9

Fig. 2. SMRTbell barcode workflow and recommended adapters

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

http://www.pacb.com/wp-content/uploads/2015/09/User-Bulletin-Barcode-Plate-Mapping.pdf
http://www.pacb.com/wp-content/uploads/2015/09/Package-Insert-Barcoded-Adapters-Plate-96.pdf

Materials and Methods

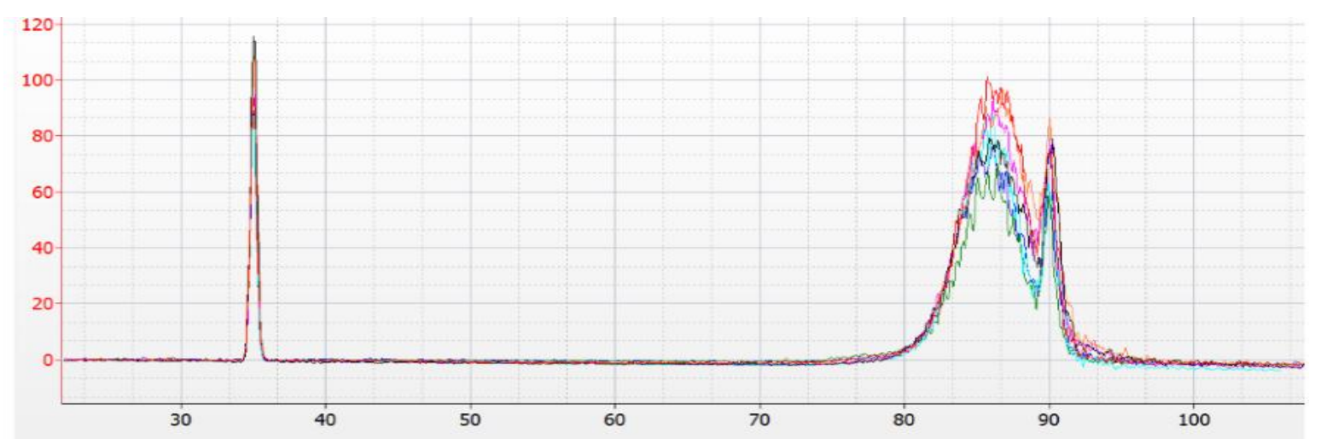


Fig. 3. Bioanalyzer traces of eight strains of sheared *H.pylori* gDNA used in an 8-plex library

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

- Recommend shearing 1 µg of DNA per sample to account for sample loss during shearing, Exo VII treatment, and DNA damage repair
- Sample input DNA for end-repair and ligation 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 END-REPAIR AND LIGATION
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*

*Assuming fragment distribution for each microbe is the same

Table 1. Multiplexing DNA input recommendations

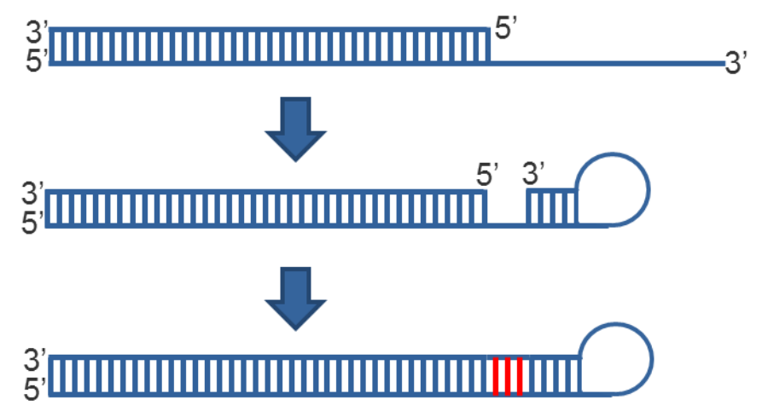


Fig. 4. Remove 3' overhangs by treating the sheared DNA with Exo VII

SAMPLES	TOTAL SUBREADS	PALINDROMIC SUBREADS	% PALINDROMES	# CONTIGS
B.sub (-) Exo VII	205684	23689	11.5	204
***B.sub* (+) Exo VII**	**300300**	**4217**	**1.4**	**1**
E.coli (-) Exo VII	145000	15866	11.0	101
***E.Coli* (+) Exo VII**	**347387**	**7101**	**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

1 SMRT SEQUENCING
2 SCORE/ALIGN READS PER MOLECULE
3 FILTER BY BARCODE
4 PERFORM HGAP ASSEMBLY BY BARCODE

Fig. 5. Informatics pipeline for microbial multiplexing (SMRT Analysis v2.3*)

1. Samples were sequenced on the PacBio RS II with P6-C4 chemistry and 360 minute movies.
2. To score and align each read with their barcodes, SMRT Cells were processed with *pbbarcode labelZmws*, a custom tool available in PacBio DevNet. This command line tool scores and aligns barcodes against all potential barcodes and return the highest scoring for each molecule.
3. FASTQ files were then exported for each barcode using *pbbarcode emitFastq* and then the headers were extracted to use as a "whitelist" of molecules for filtering input data for HGAP analysis.
4. HGAP *de novo* assembly was run using each barcode's whitelist in the P_Filter module.

*Multiplexed HGAP assembly in SMRT Analysis v3.1 is simplified by a new dataset abstraction model.

https://github.com/PacificBiosciences/pbbarcode
https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-Whitelisting-Tutorial

Results

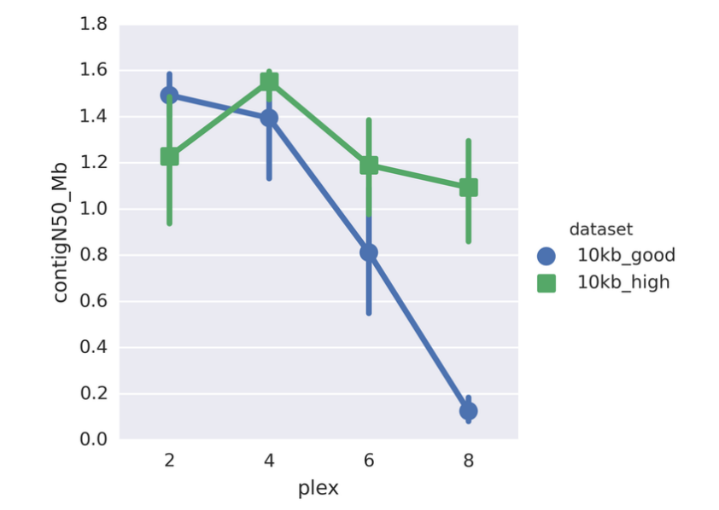


Fig. 6. Effect of increasing P1 productivity on multiplexed *H. pylori* (1.6 Mb) assembly contig N50s

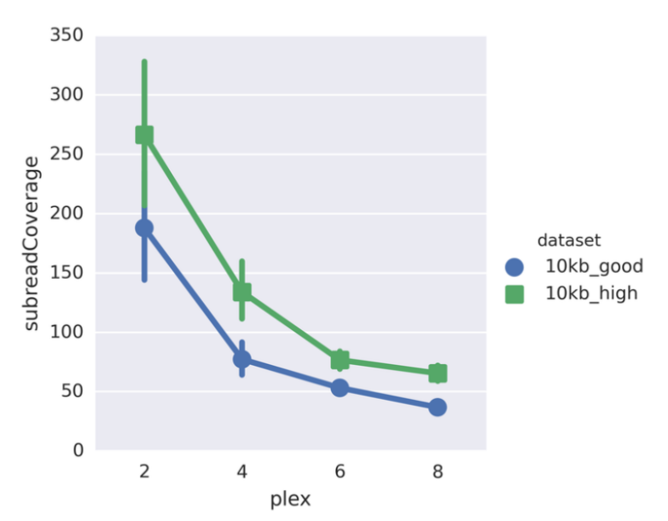


Fig. 7. Read coverage vs multiplex level

- Increasing single-loaded wells (P1%) to 50-60% provides better assembly for higher-multiplexed samples compared to 35-45% P1%
- Increased loading generates more subreads resulting in higher coverage for each *H. pylori* strain per multiplex level
- Note: if loading is too high, final consensus accuracy will be impacted (data not shown)

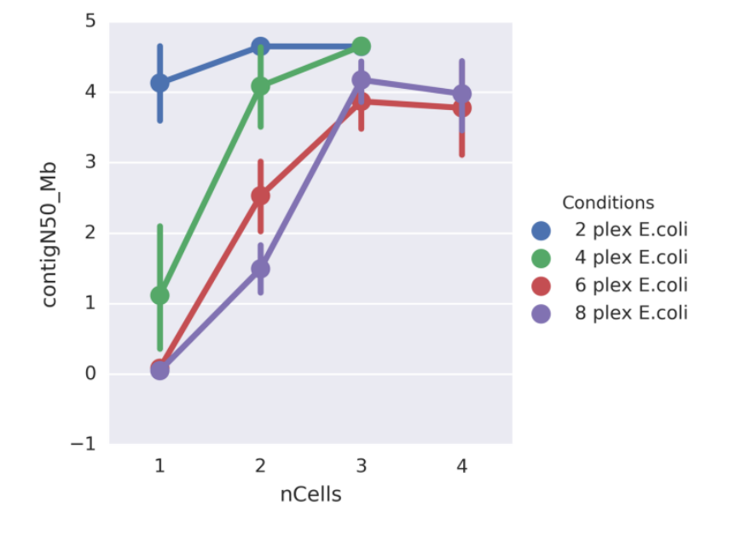


Fig. 8. Assembly contig N50s from multiplexed *E. coli* (5 Mb) strains

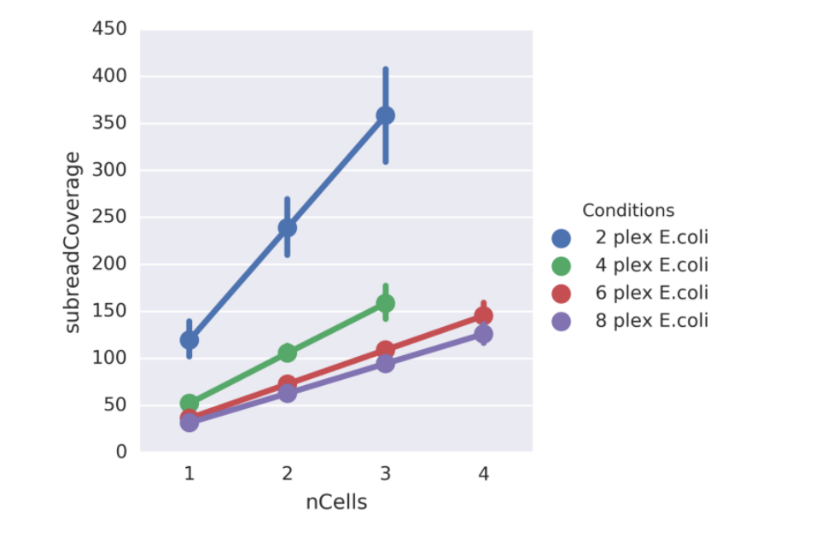


Fig. 9. Subread coverage vs number of SMRT Cells

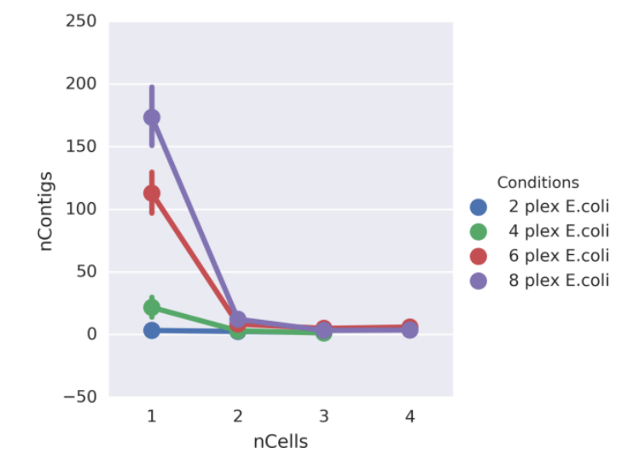


Fig. 10. *E. coli* assembly performance (contig number) by multiplex level and number of SMRT Cells

- Near-complete assemblies (<10 contigs) can be achieved with 50-150 fold coverage
- A single PacBio RS II SMRT Cell can multiplex two 5 Mb microbes and achieve a complete assembly
- 6- 8-plex experiments required running 3 and 4 SMRT Cells
- Recommend 'two microbes = 1 SMRT Cell' as a guideline for designing microbial WGS multiplexing experiments on the PacBio RS II
- Increase SMRT Cells if higher coverage and fewer contigs are required

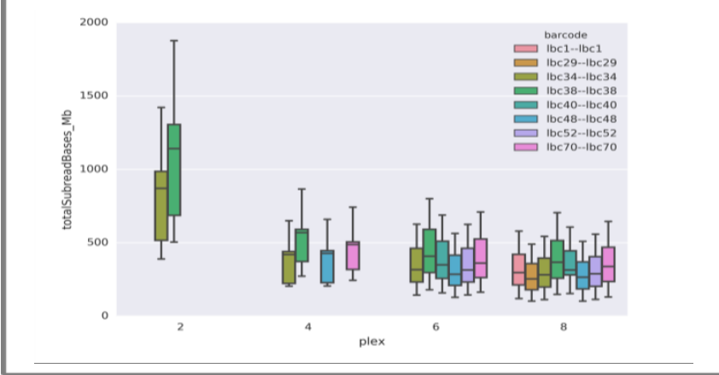


Fig. 11. Reads per microbes by multiplex level

- To achieve even number of subreads per sample, it is highly recommended to pool samples in equimolar amounts

Conclusions

- 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%-50%
  - Reduce sequencing cost by ~50%
  - Increase instrument sample processing capacity by 2-8 fold
- Optimizing % P1 productivity allows for increased multiplexing levels
- Recommend 'two microbes per 1 SMRT Cell' as a general guideline for PacBio RS II microbial WGS multiplexing experiments
- 5 Mb genomes can be pooled to 8-plex and generate near-complete genome assemblies (<10 contigs) in ~4 SMRT Cells
- Microbes with large repeat regions may need size selection or more SMRT Cell
- An automated informatics pipeline for HGAP assembly of multiplexed microbes is available in SMRT Analysis 3.1