Multiplexed Complete Microbial Genomes on the Sequel System

Cheryl Heiner, Steve Oh, James Drake, and Joan Wilson
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

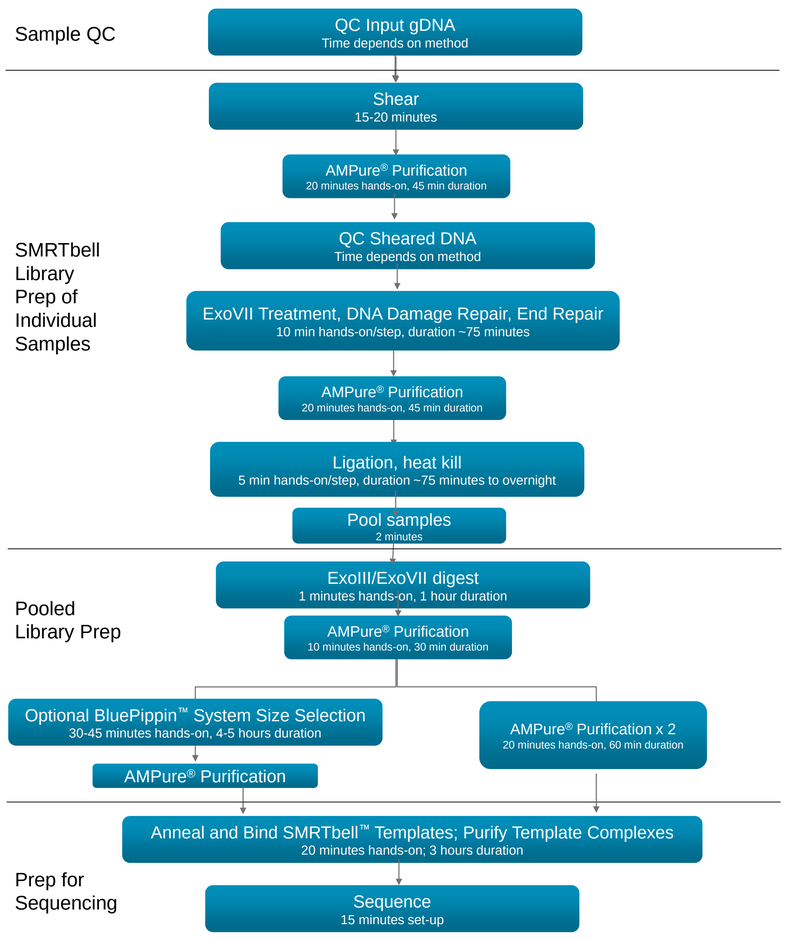
Abstract

Microbes play an important role in nearly every part of our world, as they affect human health, our environment, agriculture, and aid in waste management. Complete closed genome sequences, which have become the gold standard with PacBio long-read sequencing, can be key to understanding microbial functional characteristics. However, input requirements, consumables costs, and the labor required to prepare and sequence a microbial genome have in the past put PacBio sequencing out of reach for some larger projects. We have developed a multiplexed library prep approach that is simple, fast, and cost-effective, and can produce 4 to 16 closed bacterial genomes from one Sequel SMRT Cell.

Additionally, we are introducing a streamlined analysis pipeline for processing multiplexed genome sequence data through *de novo* HGAP assembly, making the entire process easy for lab personnel to perform.

Here we present the entire workflow from shearing through assembly, with times for each step. We show HGAP assembly results with single or very few contigs from bacteria from different size genomes, sequenced without or with size selection. These data illustrate the benefits and potential of the PacBio multiplexed library prep and the Sequel System for sequencing large numbers of microbial genomes.

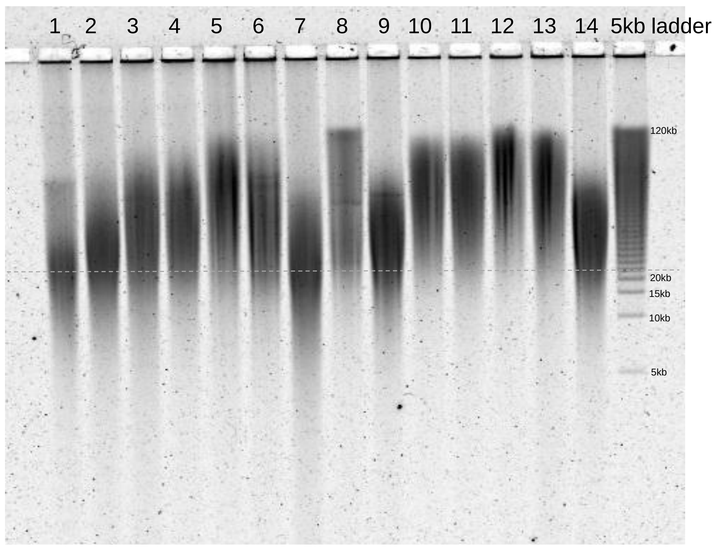
Barcoded Library Prep Workflow



Steps in microbial multiplex barcoded library prep from input QC through sequencing set-up, including approximate hands-on and duration times. The entire process can be done in < 2 days, excluding size selection.

Library Prep

Step 1: Input gDNA QC. Quality assessment by pulsed field gel, field inversion gel, or capillary electrophoresis is recommended. An example of a CHEF Mapper® gel is shown below. To generate the best assemblies, gDNA should be predominately >20-25 kb (dotted line below). 1.5 µg of gDNA is required per sample.



Step 2: SMRTbell library construction through ligation. Pre-mixes are calculated for the number of samples in each library.

Pre-mix calculator – 8plex

Table 2. DNA Damage Repair Master Mix	Vol. (µL)
1x Elution Buffer	13.0
10x DNA Damage Repair Buffer (µL)	5.0
ATP high (µL)	20.0
dNTPs (µL)	2.0
NAD (µL)	2.0
DNA damage repair mix (µL)	8.0

Add 5.0 uL of damage repair master mix to µL aliquot of ExoVII-treated, sheared gDNA, for a final volume of 20 uL

Step 3: Pooling of samples after ligating barcoded adapters. Equimolar representation of all samples can be achieved through the use of a pooling calculator, which takes into account genome and shear sizes.

Pooling calculator (User input values)

Plex Level	8	Pool vol (µl)	40

Sample Description	Expected Genome Size (Bases)	Average Shear Size (Bases)	Pooling Volume (µL)
1	4,600,000	9,000	5.1
2	4,000,000	10,000	4.9
3	5,500,000	8,500	5.8
4	2,900,000	9,500	3.4
5	4,600,000	10,200	5.8
6	4,000,000	8,800	4.3
7	5,500,000	10,500	7.1
8	2,900,000	10,000	3.6
Average	4,250,000	9,563	

Sequencing and Analysis Workflow

Streamlined software workflow in SMRT Link v5.1 to demultiplex and assemble multiplexed microbial data

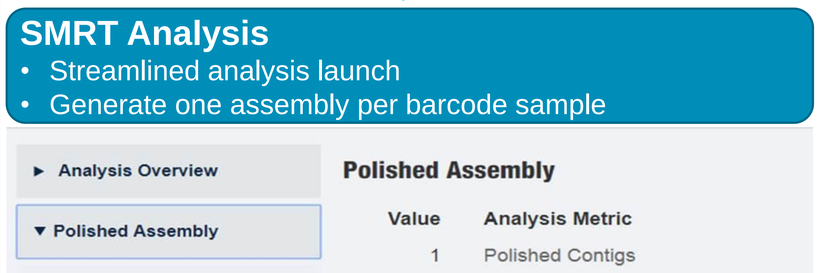
Run Design
- User inputs Bio Sample Names per barcode

	A	B
1	Barcode Name	Bio Sample Name
2	bc1001--bc1001	E. coli
3	bc1002--bc1002	S. aureus
4	bc1003--bc1003	B. subtilis

Data Management
- Barcoded samples are automatically demultiplexed

Data Set Name	Bio Sample Name	Barcode Name
My microbial samples	[multiple]	[multiple]
My microbial samples (E. coli)	E. coli	bc1001--bc1001
My microbial samples (S. aureus)	S. aureus	bc1002--bc1002
My microbial samples (B. subtilis)	B. subtilis	bc1003--bc1003

SMRT Analysis
- Streamlined analysis launch
- Generate one assembly per barcode sample



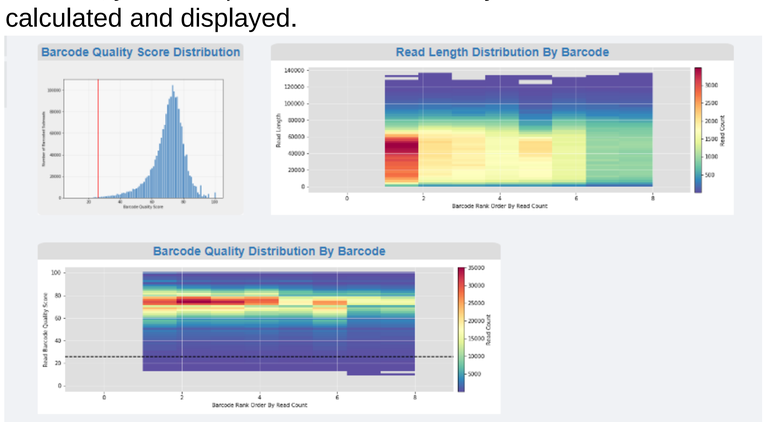
Sequencing Primary Results

Well	Total Gbases	Polymerase Read RL		Longest Subread RL		Productive (P1) Reads
		Mean	N50	Mean	N50	
A01	9.7	16,531	34,750	5,324	7,750	587,142
B01	9.2	12,437	28,250	4,687	7,250	737,521
C01	9.1	13,296	31,250	6,067	9,750	684,147
D01	11.4	16,475	36,250	6,754	9,750	690,303

Primary sequencing results from 4 cells, multiplexed microbial samples, run with SMRT Link v5.1. Data collection was for 10 hours.

Demultiplexing

With SMRT Link 5.1.0, demultiplexing occurs automatically at the end of a sequencing run, if sample and barcode information is included in the run design. In addition to binning the data by barcode, yield and performance metrics by barcode are calculated and displayed.



Bio Sample Name	Barcode	# Polym Reads	# Subreads	# of Bases	Mean Polym Read Length	Mean Barcode Quality	Rank Order (Num. Reads)
Bacillus subtilis	bc1001--bc1001	41,813	245,594	1,029,358,501	24,618	67	6
Ecoli K12	bc1002--bc1002	51,893	284,396	1,244,590,689	23,984	70	3
Rhodpseduomonus palustris	bc1003--bc1003	46,200	195,150	1,022,333,660	22,128	69	5
Staphylococcus aureus	bc1004--bc1004	28,105	150,965	747,761,521	26,606	70	7
Bacillus subtilis	bc1008--bc1008	46,766	270,980	1,140,250,893	24,382	70	4
Ecoli K12	bc1009--bc1009	54,283	300,537	1,312,323,207	24,176	71	2
Rhodpseduomonus palustris	bc1010--bc1010	75,055	315,169	1,678,994,881	22,370	70	1
Staphylococcus aureus	bc1012--bc1012	27,416	150,242	736,876,105	26,878	69	8
No Name	Not Barcoded	215,575	236,904	673,642,026	3,125	0	NA

Demultiplexing results from Experiment 1 below. Although only 63% of polymerase reads were barcoded, 93% of the bases were barcoded; unbarcoded reads were short.

Assembly

Assembly results from three 8-plex experiments as described below. Consensus QV was >50 for all control samples.

Experiment	Microbe	Size Select?	Avg Pre-Assembled Sub Read Length	# of Polished Contigs	Longest Contig	Sum of Contigs
Experiment 1 8-plex: 4 microbes, 2 barcodes each, no size selection	Bacillus subtilis	no	5,336	1	4,044,361	4,044,361
	Ecoli K12	no	5,800	1	4,642,487	4,642,487
	Bacillus subtilis	no	5,477	1	4,051,597	4,051,597
	Staphylococcus aureus	no	5,608	2*	2,885,313	2,912,379
	Rhodpseduomonus palustris	no	7,370	2	5,459,211	5,476,173
	Rhodpseduomonus palustris	no	6,776	1	5,459,215	5,459,215
	Staphylococcus aureus	no	5,650	3*	2,878,891	2,921,302
	Ecoli K12	no	5,779	1	4,642,497	4,642,497
Experiment 2 8-plex: 4 microbes, 2 barcodes each, with size selection	Bacillus subtilis	yes	5,986	1	4,045,573	4,045,573
	Rhodpseduomonus palustris	yes	7,150	1	5,459,213	5,459,213
	Bacillus subtilis	yes	6,133	2	4,054,881	4,065,461
	Rhodpseduomonus palustris	yes	7,820	1	5,459,214	5,459,214
	Ecoli K12	yes	6,336	1	4,642,498	4,642,498
	Staphylococcus aureus	yes	6,138	2*	2,878,895	2,905,963
	Ecoli K12	yes	6,263	1	4,642,493	4,642,493
	Staphylococcus aureus	yes	6,122	2*	2,878,898	2,905,966
Experiment 3 8-plex: 1 control microbe, compare 8 barcodes, no size selection	Ecoli K12	no	5,770	2	4,469,964	4,665,771
	Ecoli K12	no	5,554	1	4,642,492	4,642,492
	Ecoli K12	no	5,371	1	4,642,498	4,642,498
	Ecoli K12	no	5,870	1	4,642,494	4,642,494
	Ecoli K12	no	5,625	1	4,637,574	4,637,574
	Ecoli K12	no	5,606	1	4,642,484	4,642,484
	Ecoli K12	no	5,921	1	4,638,784	4,638,784
	Ecoli K12	no	5,337	1	4,642,491	4,642,491

*Contains 1 or 2 plasmids

Conclusion

We have demonstrated that 8 microbes of ~4 Mb can be multiplexed on one Sequel SMRT Cell using barcoded adapters. Resulting assemblies produce high quality complete genomes, often 1 contig plus plasmids.