Single Chromosomal Genome Assemblies on the Sequel System with Circulomics High Molecular Weight DNA Extraction for Microbes

J. Wong[1], C. Heiner[1], M. Kim[2], H. Ferrao[1], V. J. Wallace[2], K. Eng[1], R. Fedak[2], D. Kilburn[2], M. Ashby[1], P. Baybayan[1], J. M. Burke[2], K. J. Liu[2]
[1]Pacific Biosciences, Inc., Menlo Park, CA, [2]Circulomics, Baltimore, MD

Abstract

Background: The Nanobind technology from Circulomics provides an elegant HMW DNA extraction solution for genome sequencing of Gram-positive and -negative microbes. Nanobind is a nanostructured magnetic disk that can be used for rapid extraction of high molecular weight (HMW) DNA from diverse sample types including cultured cells, blood, plant nuclei, and bacteria. Processing can be completed in <1 hour for most sample types and can be performed manually or automated with common instruments.

Methods: We have validated several critical steps for generating high-quality microbial genome assemblies in a streamlined microbial multiplexing workflow. This new workflow enables high-volume, cost-effective sequencing of up to 16 microbes totaling 30 Mb in genome size on a single SMRT Cell 1M using a target shear size of 10 kb. We also evaluated this method on a pool of four "class 3" microbes that contain >7 kb repeats. Fragment size was increased to ~14 kb, with some fragments >30 kb.

Results: Here we present a demonstration of these capabilities using isolates relevant to high-throughput sequencing applications, including common foodborne pathogens (*Shigella, Listeria*, *Salmonella*), and species often seen in hospital settings (*Klebsiella*, *Staphylococcus*). For nearly all microbes, including difficult-to-assemble class III microbes, we achieved complete *de novo* microbial assemblies of ≤5 chromosomal contigs with minimum quality scores of 40 (99.99% accuracy) using data from multiplexed SMRTbell libraries. Each library was sequenced on a single SMRT Cell 1M with the PacBio Sequel System and analyzed with streamlined SMRT Analysis assembly methods.

Conclusions: We achieved high-quality, closed microbial genomes using a combination of Circulomics Nanobind extraction and PacBio SMRT Sequencing, along with a newly streamlined workflow that includes automated demultiplexing and push-button assembly.

Circulomics DNA Extraction and SMRTbell Library Preparation Workflows

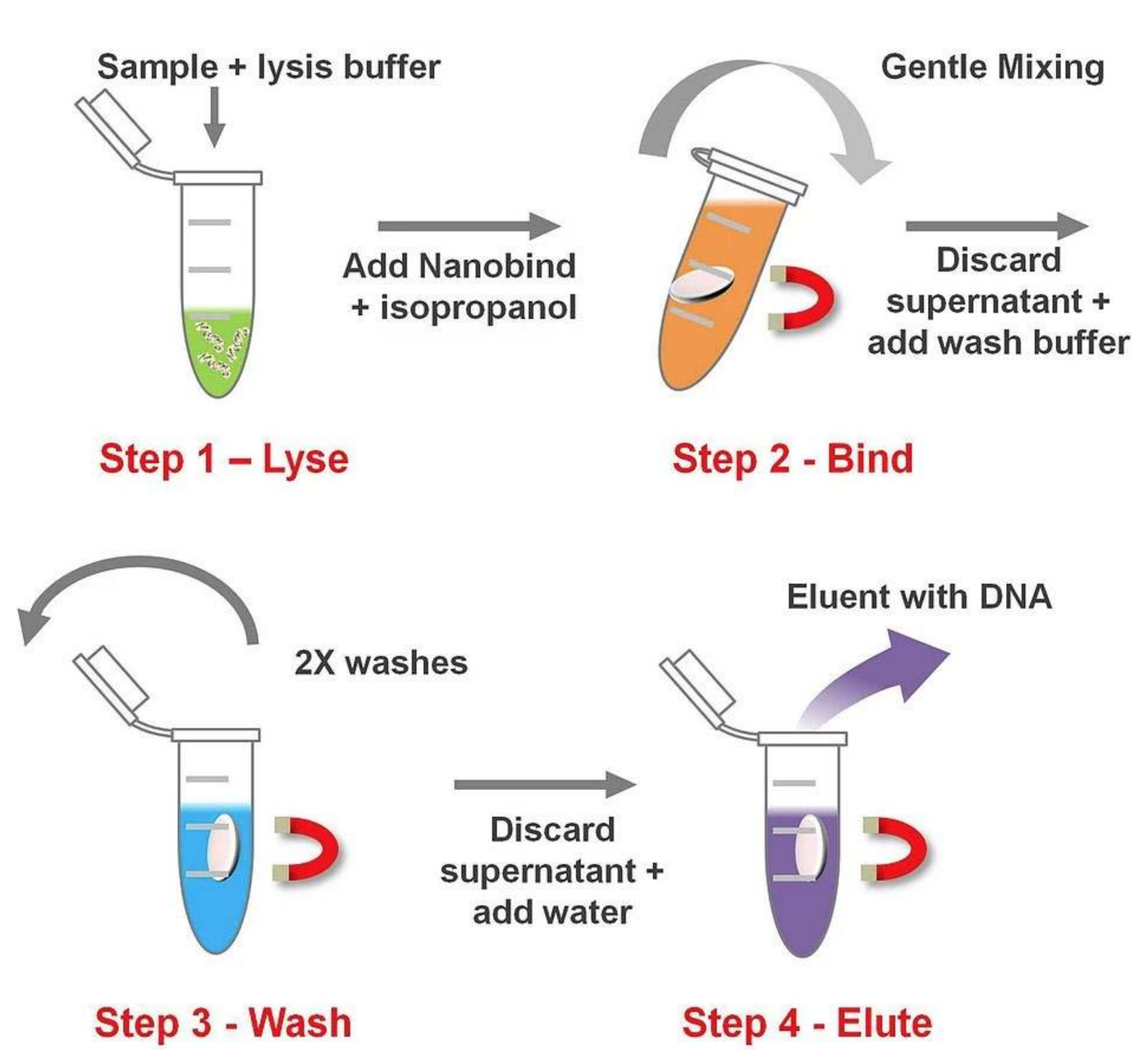

Figure 1. Circulomics high-quality DNA extraction. Successful DNA extraction is achieved using Nanobind, a thermoplastic disk that contains a high density of micro- and nanostructured silica. The unique structure of Nanobind enables vast amounts of genomic DNA (gDNA) to bind and elute without inducing damage. Processing comprises a rapid bind, wash, and elute process that parallels magnetic beads and is easily automated. Nanobind extraction produces DNA molecules of 50 to 300+ kb in length with ultra-low damage and high purity in under an hour.

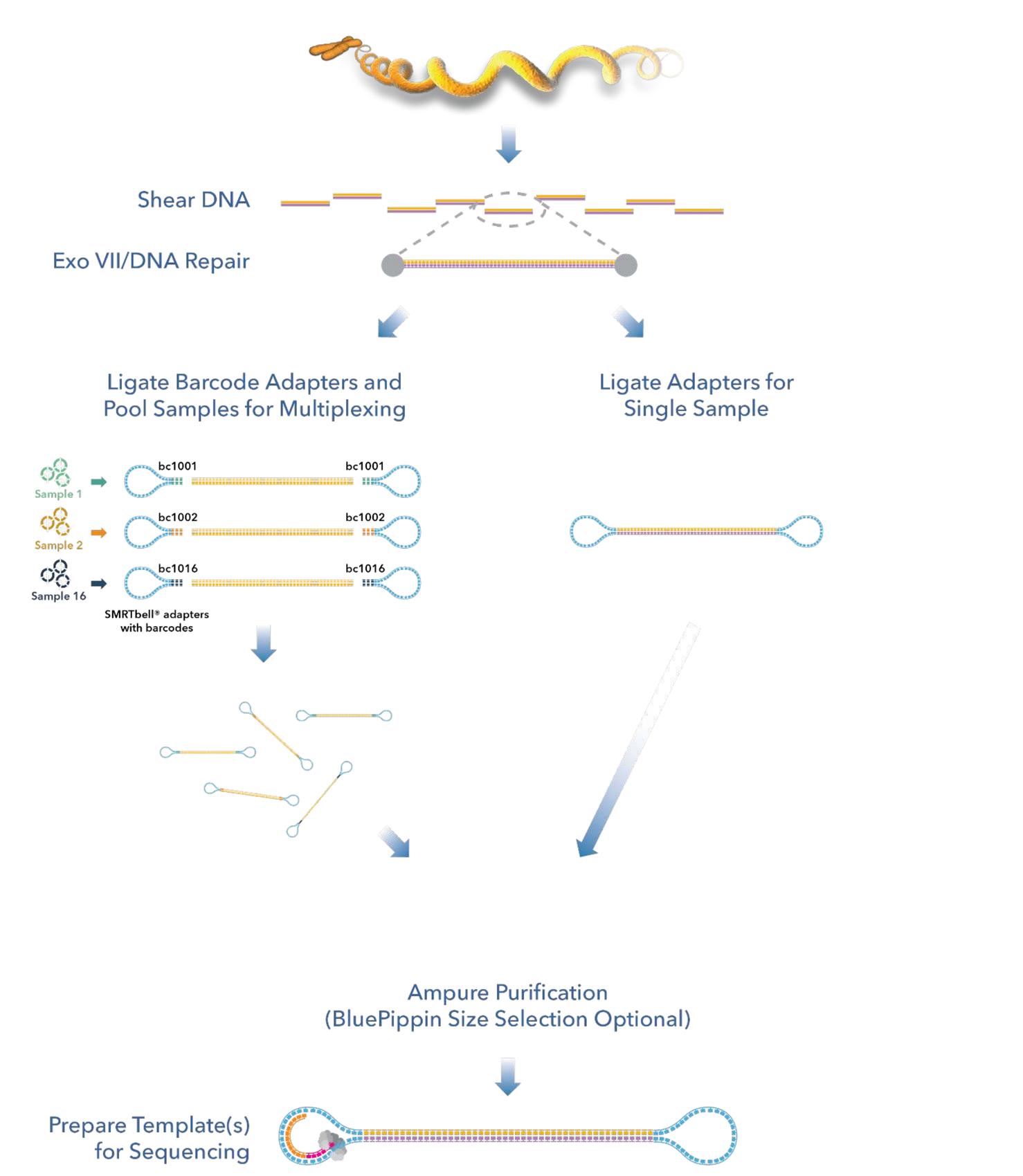

Figure 2. SMRTbell library preparation. High-quality gDNA samples are sheared to a 10-kb distribution and ligated with barcoded adapters. Two new kits are available, each with 8 barcoded SMRTbell adapters specifically validated for the microbial multiplexing application on the Sequel System.

Microbial DNA Extraction and Library Preparation

Well	Microbe	Genome Size, Mb	# of Plasmids
1	*Bacillus cereus* Strain 971[a]	5.3	1
2	*Enterococcus faecalis* OG1RF[b]	2.7	0
3	*Klebsiella pneumoniae* ATCC BAA-2146[b]	4.6	4
4	*Listeria monocytogenes* CFSAN008100[a]	5.8	1
5	*Escherichia coli* K12[c]	3.0	0
6	*Salmonella enterica* subsp. *enterica* serovar Heidelberg[a]	5.0	3
7	*Shigella sonnei* CFSAN030807[a]	4.8	8
8	*Staphylococcus aureus* subsp. *aureus* ATCC 25923[b]	2.8	1
	Sum of genomes	**34.1**	

Table 1. Microbes sequenced on one Sequel SMRT Cell 1M.
The Circulomics Nanobind CBB Big DNA Kit was used to extract high-quality DNA from:
[a] common foodborne pathogens
[b] species often seen in hospital settings
[c] the control microbe

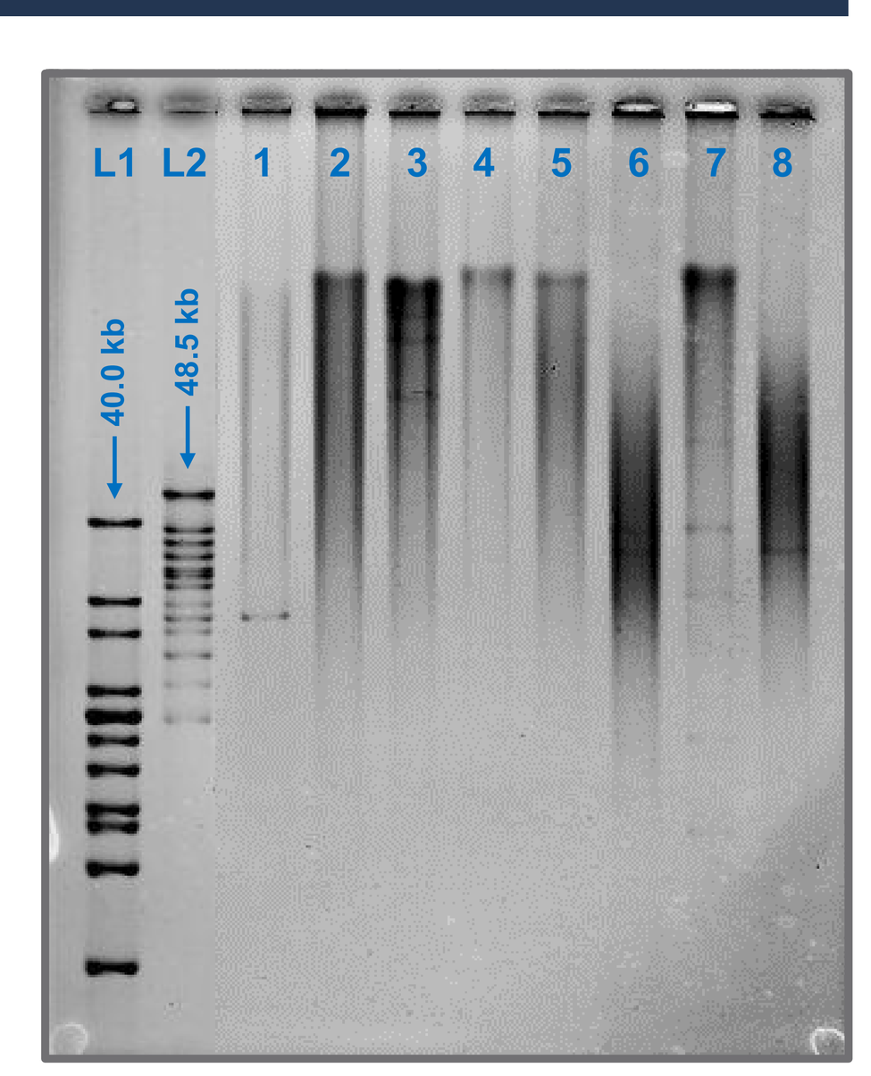

Figure 3. CHEF Mapper QC of microbes in Table 1.

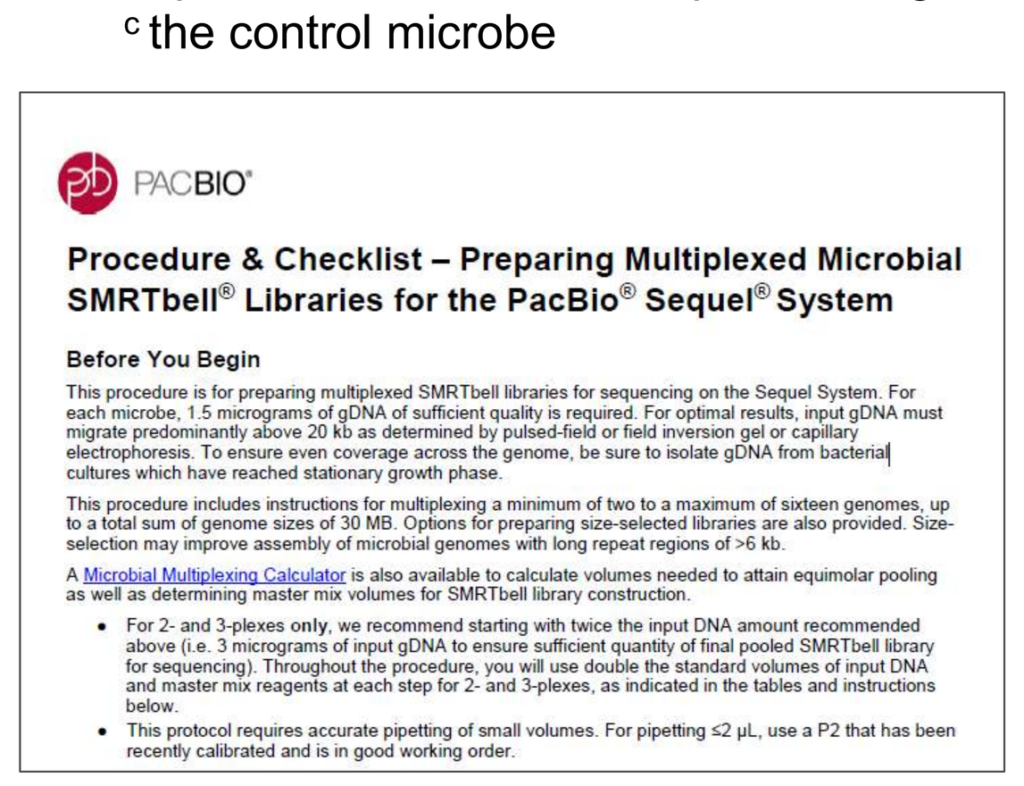

Figure 4. SMRTbell library preparation procedure and checklist.

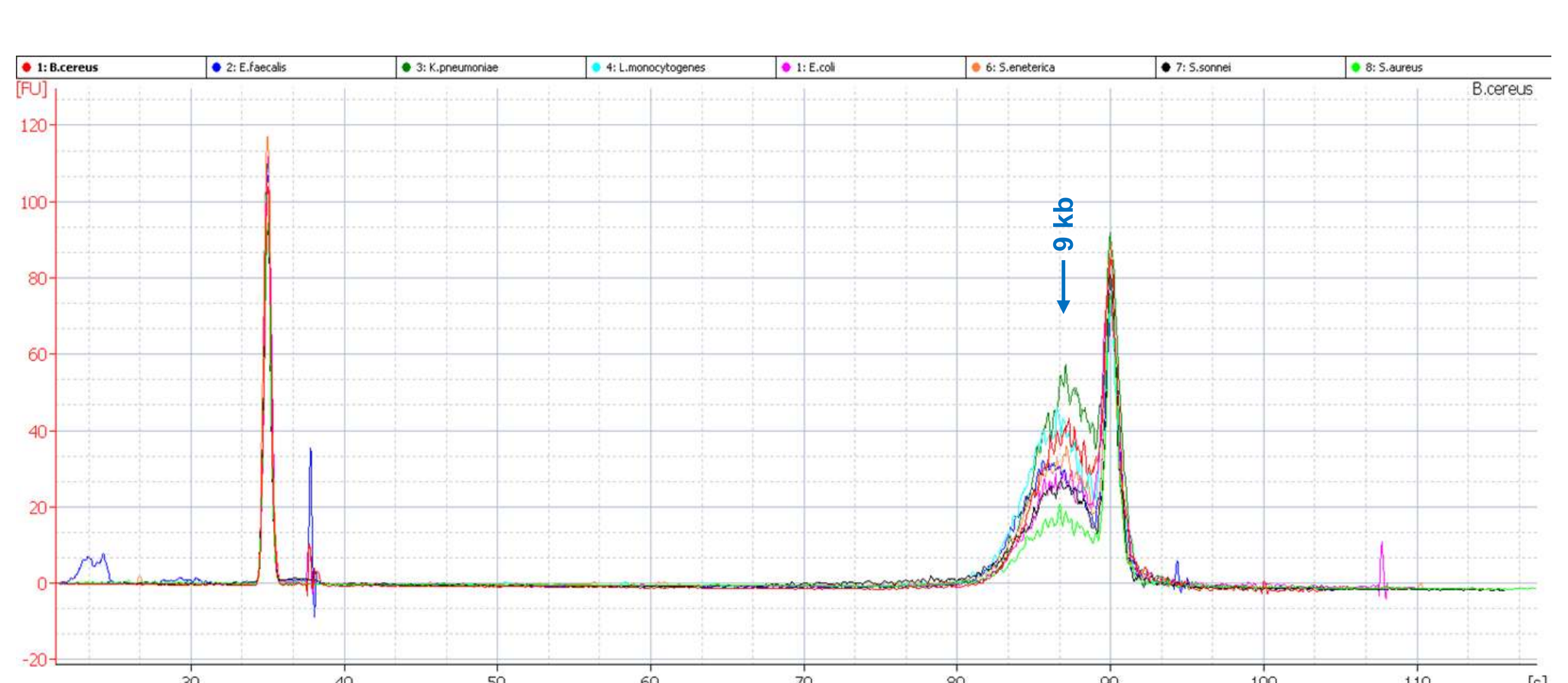

Figure 5. Size distribution of sheared samples. The mode is ~9 kb.

Library Pooling and Sequencing

Microbial Multiplexing Calculator

Sample Information
Bacterial Multiplex

Calculate

Plex (Calculated): 8

Total Sample Pool Volume Post Ligation (µl): 40.0

Sample Name	Barcode	Expected Genome Size (bases)	Avg Shear Size (bases)	Optional Sample Conc (ng/µl)	Calculated Volumes (µl)
Microbe Sample 1	BC1001	5400000	10000		6.4
Microbe Sample 2	BC1002	2700000	10000		3.2
Microbe Sample 3	BC1004	5700000	10000		6.7
Microbe Sample 4	BC1008	3100000	10000		3.6
Microbe Sample 5	BC1009	4600000	10000		5.4
Microbe Sample 6	BC1010	4900000	10000		5.8
Microbe Sample 7	BC1012	4800000	10000		5.6
Microbe Sample 8	BC1014	2800000	10000		3.3

Table 2 (left). Pooling calculations. Following ligation of barcoded adapters, libraries were pooled at equimolar representation with the help of the Microbial Multiplexing Calculator, which takes into account both genome size and shear size to normalize genome coverage. Up to 16 microbes totaling 30 Mb can be multiplexed on one SMRT Cell 1M.

Total Gb	Polymerase Read Length (mean)	Polymerase Read Length (N50)	Insert Read Length (mean)	Insert Read Length (N50)	# of Primary Reads
8.03	10,283	25,750	4,828	7,750	789,620

Table 3. Primary analysis results from 8-plex library, sequenced with v2.1 chemistry.
The pooled library was sequenced on one Sequel SMRT Cell 1M, with 10-hour data collection.

Sequencing and Analysis Workflow

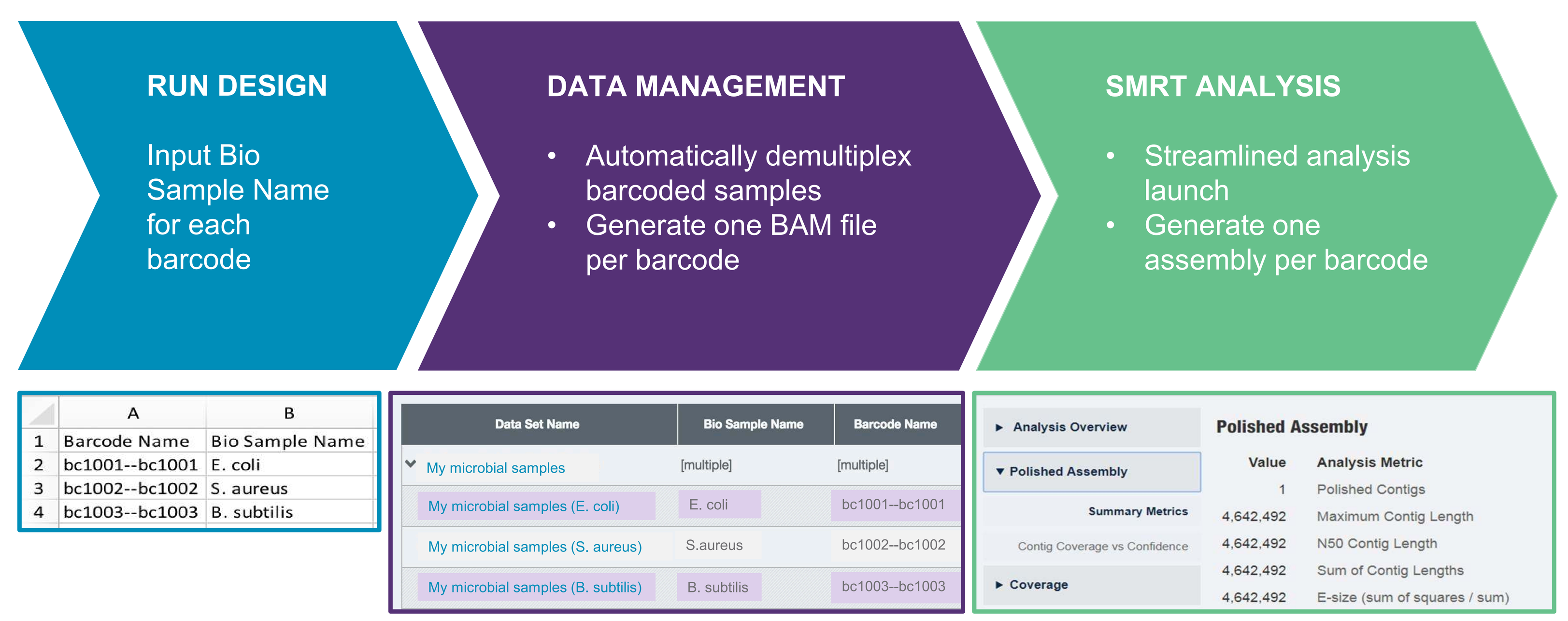

Figure 6. Streamlined analysis workflow in SMRT Link v5.1 to demultiplex and assemble microbial data.

Assembly Results

Microbe	Genome Size, Mb	Mean Coverage	# Plasmids + Chromosomes	# Contigs	Contig N50
*Bacillus cereus**	5.3	159	2	2	5,416,535
Enterococcus faecalis	2.7	186	1	1	2,739,522
Escherichia coli K12	4.6	135	1	1	4,642,485
Klebsiella pneumoniae	5.8	128	4	4	5,435,721
Listeria monocytogenes	3.0	162	1	1	3,032,236
Salmonella enterica	5.0	191	4	8	4,795,026
Shigella sonnei	4.8	135	1	1	4,813,418
Staphylococcus aureus	2.8	226	2	2	2,778,840
Sum of genomes	**34.1**				

Table 4. Comparison of reference genomes and assembly results.
Analyses and assembly using HGAP and Circlator were carried out as described in: Analysis Procedure – Multiplexed Microbial Assembly with SMRT Link v5.1.0. *DNA extraction and sequencing were redone in a separate pool, resulting in improved contiguity.

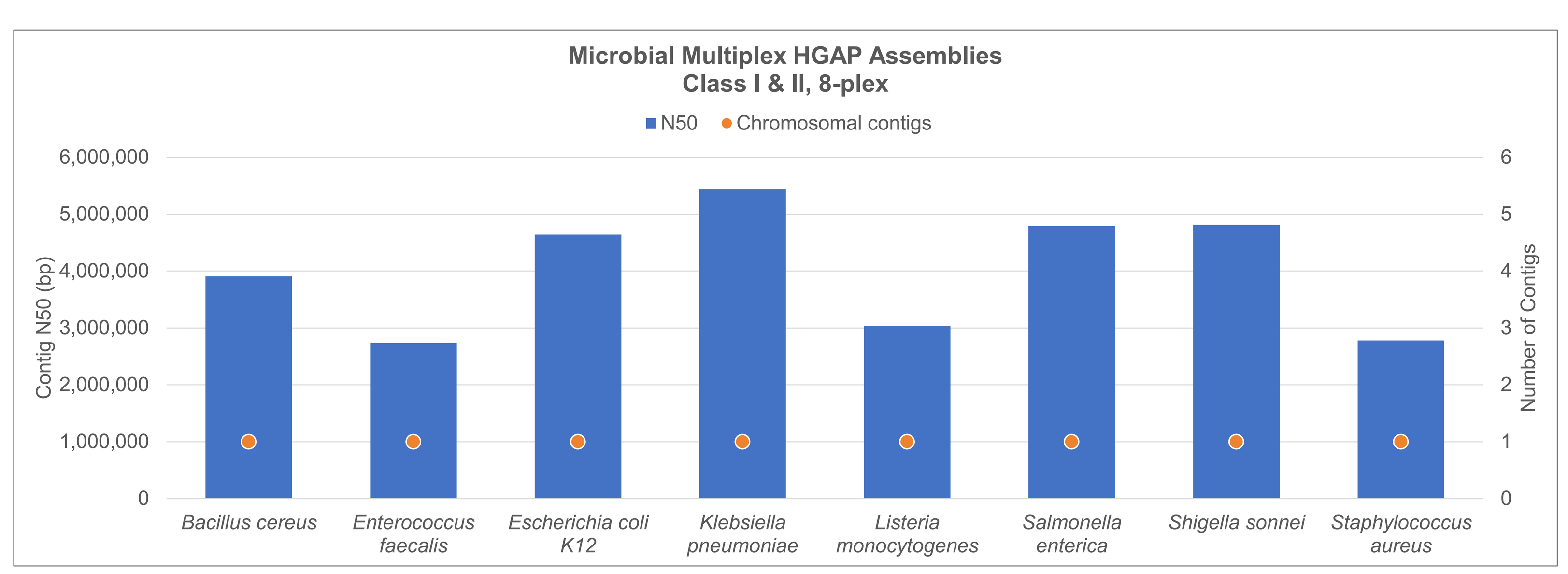

Figure 7. Number of chromosomal contigs and contig N50, excluding plasmids.
Chromosomes were assembled to single contigs for all eight microbes.

Coverage and Consensus Accuracy

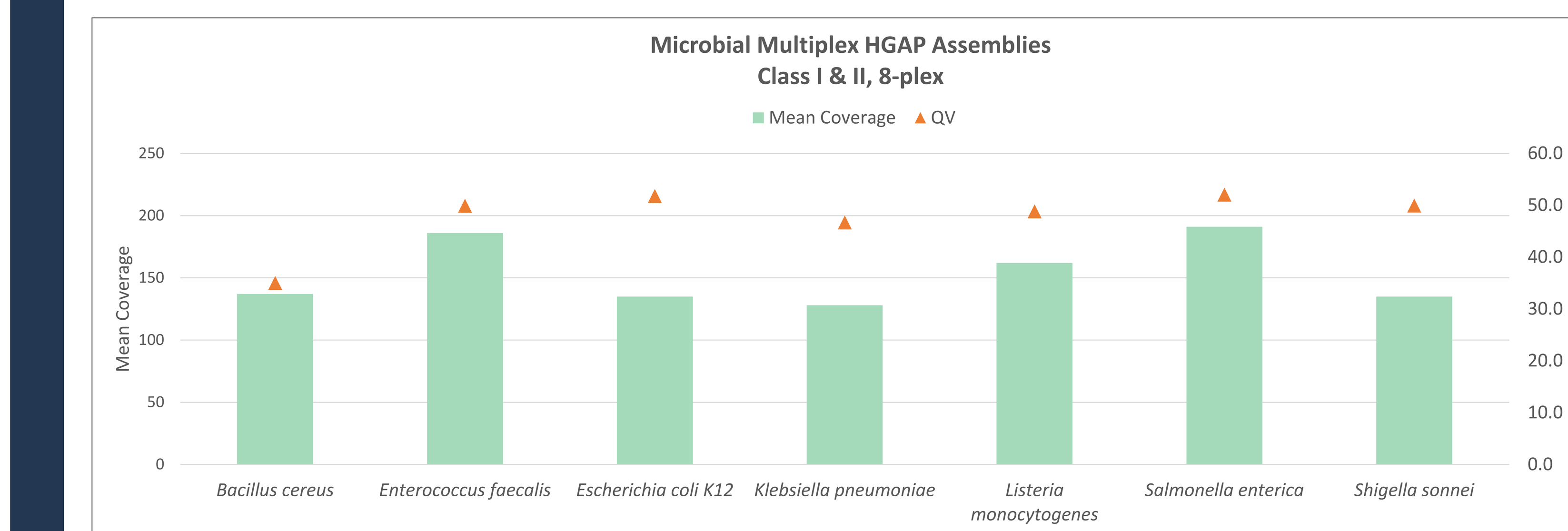

Figure 8. Mean coverage and assembly quality.
The pooling calculator results in relatively even coverage of all samples, enabling high consensus accuracy. Consensus quality values were typically around 50 for chromosomal sequences.

Multiplexing Difficult (Class III) Microbes

A second pool of four class III microbes (defined as containing repeats >7 kb) was tested for performance with the protocol described earlier, with several important changes:

- Pool size limited to 24 Mb to allow more coverage per genome
- Shear size increased to 20 kb to generate subreads spanning long repeat sequences
- Data collection time increased to 20 hours to sequence through long inserts
- Minor modifications made to HGAP parameters for *Pseudomonas aeruginosa*

Microbe	Genome Size, Mb	Mean Coverage	# Plasmids + Chromosomes	# Contigs	Contig N50
Burkholderia cepacia ATCC 25416	8.6	508	4	5	3,397,800
E coli strain W	4.7	806	1	1	4,746,024
Staphylococcus aureus HPV107	3.0	698	2	2	2,962,740
Pseudomonas aeruginosa	6.9	935	2	2	6,814,161
Sum of genomes	**23.2**				

Table 6. Comparison of reference and HGAP assembly results for class III microbes.
Note that *Burkholderia* has 3 chromosomes and 1 plasmid.

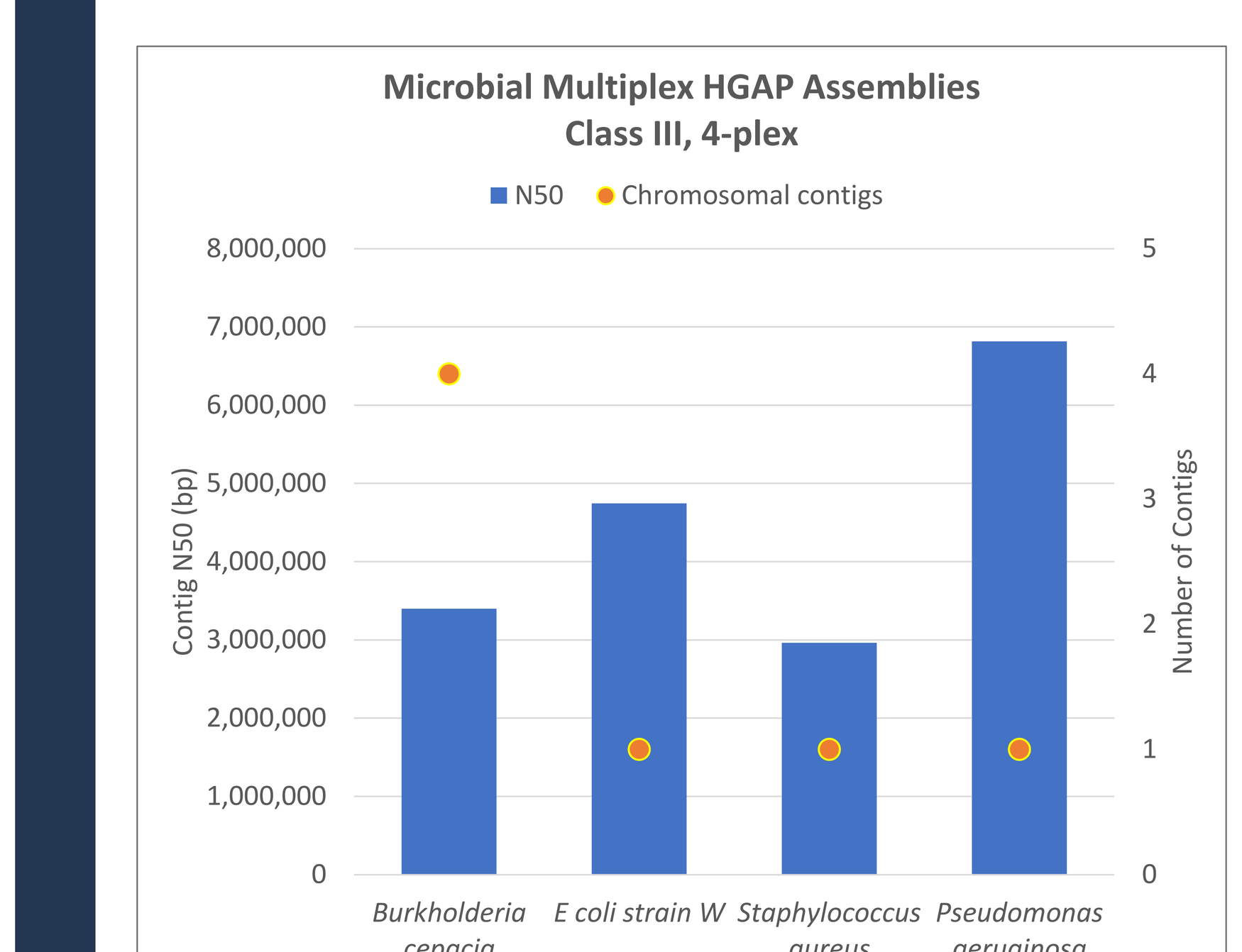

Figure 9. Number of chromosomal contigs and N50.
Chromosomes were assembled to single contigs for all four microbes.

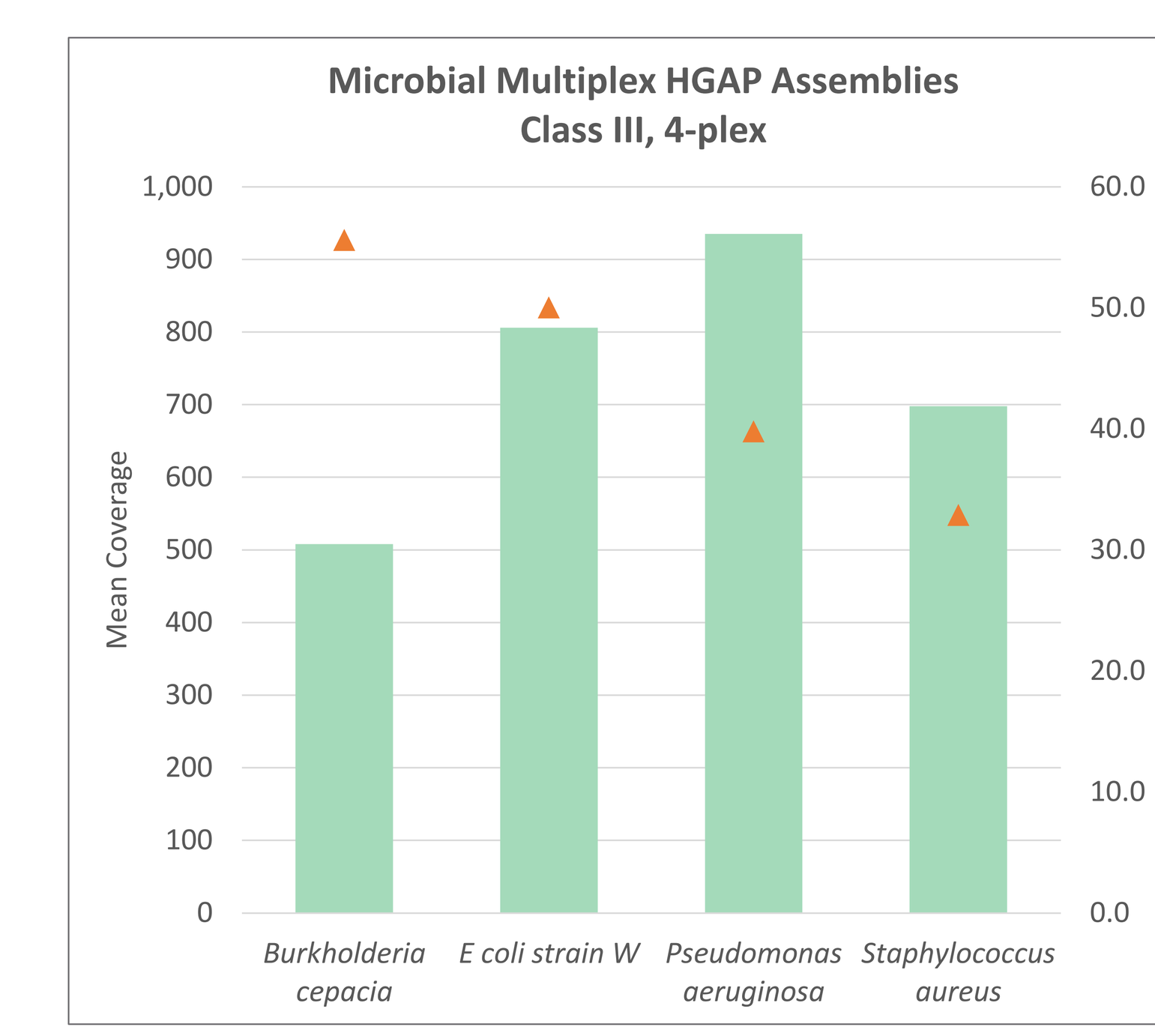

Figure 10. Mean coverage and assembly quality.
Chromosomal consensus quality values were based on closest available reference genomes in NCBI. Note that the *S. aureus* reference is a fragmented assembly from 2005.

Conclusions

- Extraction with Circulomics Nanobind produces high-purity, high molecular weight gDNA from a wide variety of microbe types, generating excellent input for PacBio SMRT Sequencing.
- With the Sequel System, microbes totaling ~30 Mb in genome size may be pooled and sequenced on one SMRT Cell 1M in a cost-efficient, rapid-turnaround process that produces complete closed genomes for most microbes. Complete plasmid sequences are often included as well.
- The data analysis workflow is highly streamlined, with automated demultiplexing and simple push-button assembly.
- In addition to a low number of contigs, assembly produces highly accurate consensus sequences, with quality values typically ranging from 40 to over 50.
- The combination of Circulomics Nanobind extraction and PacBio SMRT Sequencing can generate very good results with class III microbes, using a few modifications to the workflow.