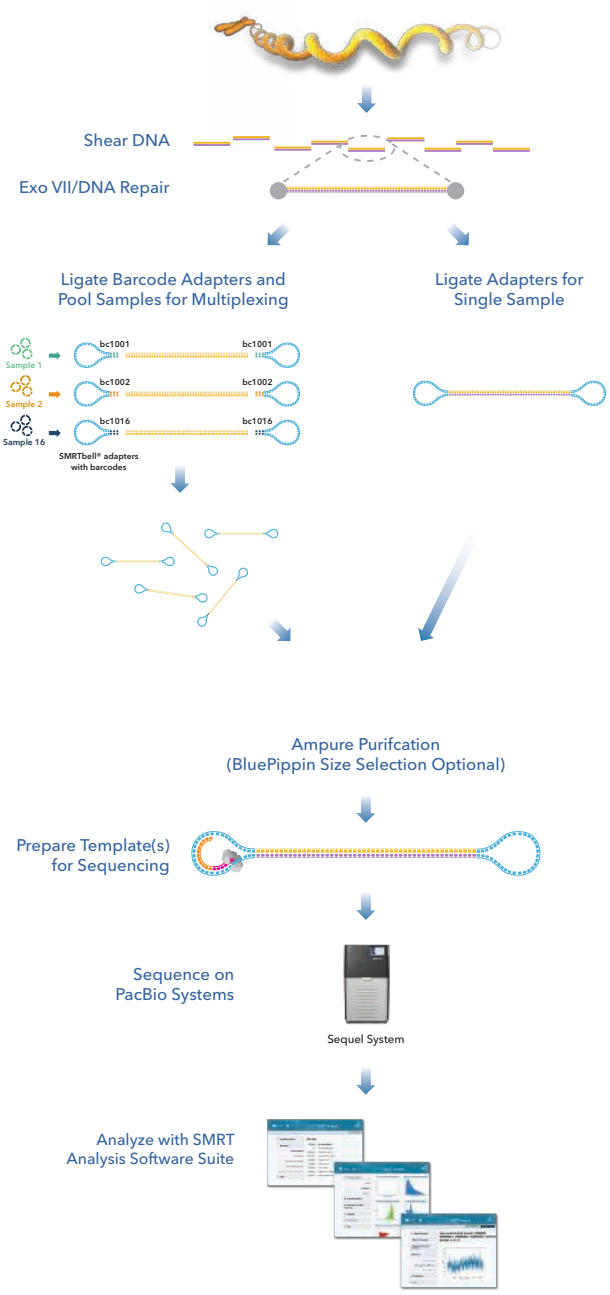
MICROBIAL WHOLE GENOME SEQUENCING
BEST PRACTICES

MICROBIOLOGY AND INFECTIOUS DISEASE

With Single Molecule, Real-Time (SMRT®) Sequencing, you can affordably characterize complete microbial genomes. For most microbes, closed genomes and accessory plasmids can be assembled using PacBio® data from single libraries in a single run.

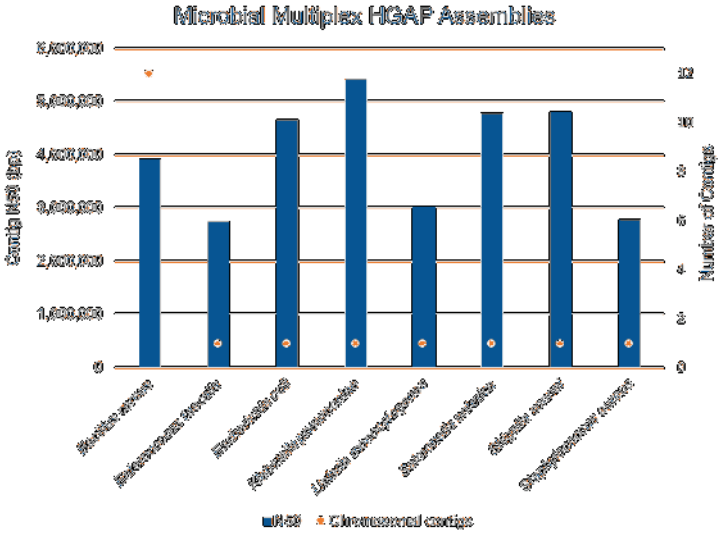
FROM GENOMIC DNA TO COMPLETE GENOME IN A SINGLE EXPERIMENT



SAMPLE PREPARATION RECOMMENDATIONS

- Use our barcoded adaptor kits to reduce the cost of microbial multiplexing[1]
- Multiplex up to 30 Mb of microbial genomes per SMRT Cell 1M on the Sequel® System to assemble most bacterial genomes into 5 contigs or fewer, exclusive of plasmids
  - Use our Microbial Multiplexing Calculator to simplify equimolar pooling[2]
  - Adjust planned multiplexing depth to balance cost constraints with your requirements for genome completeness[2]
- Closure of class III complexity genomes with large repeat regions may require 20–30 kb library preparations and may not be suitable for multiplexing[3,4]
- Use the recommended starting input of high-quality DNA (1.5 µg)

GENERATE NEAR-COMPLETE MICROBIAL ASSEMBLIES WITH MULTIPLEXING AND SEQUEL SYSTEM

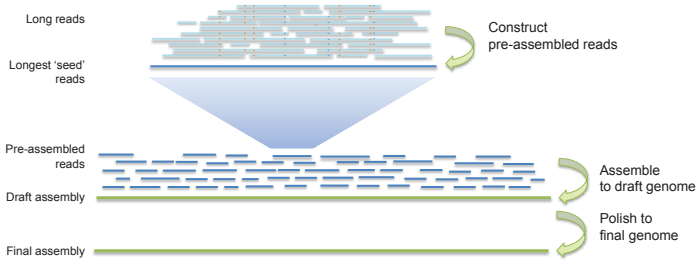


Shown above are Hierarchical Genome Assembly Process (HGAP)4 de novo assembly statistics from an 8-plex pool of bacteria relevant to food safety and nosocomial infections, sequenced in one attempt using the standard protocol and assembled using the recommended HGAP4 parameters without adjustment.

DATA ANALYSIS SOLUTIONS WITH SMRT ANALYSIS AND OUR PARTNERS

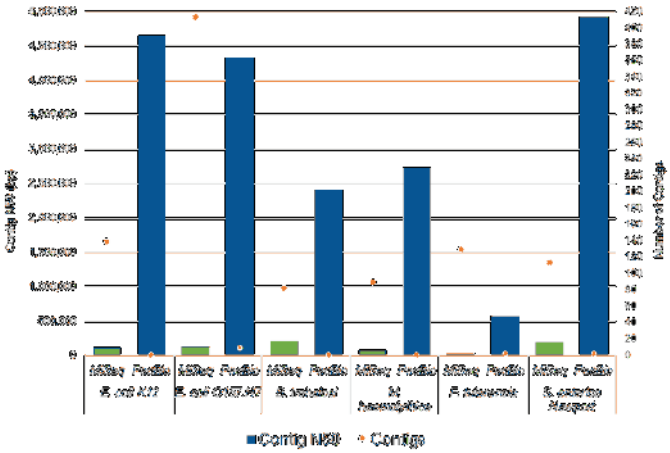
- Easily generate gold-standard genome and plasmid *de novo* assemblies with automated de-multiplexing of pooled samples and one-size-fits-most HGAP4 settings through the SMRT Analysis user interface[5,6]
- Alternatively, use cloud-based implementations of SMRT Analysis available from our data analysis partners DNAnexus or Bluebee
- Achieve high-quality consensus accuracies >99.999%
- Output data in standard file formats, BAM and FASTA/Q, for seamless integration with downstream analysis tools

HIERARCHICAL GENOME ASSEMBLY PROCESS



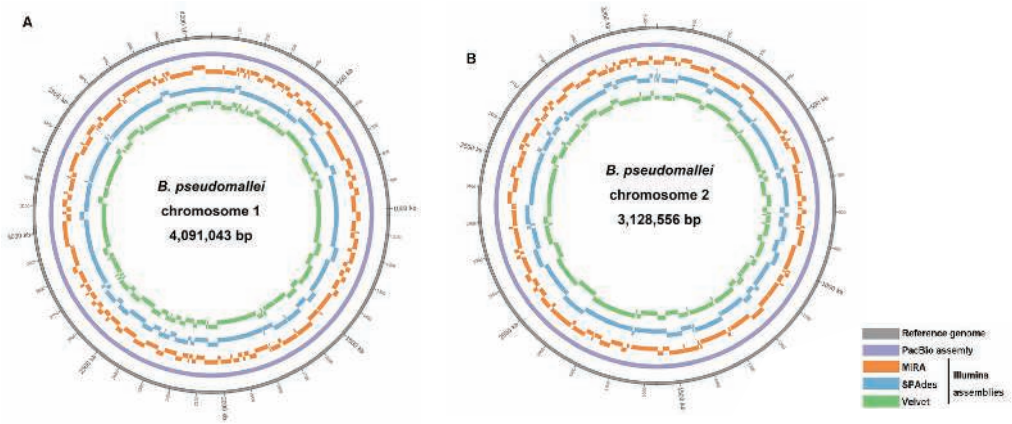
HGAP4, included in SMRT Analysis, utilizes all SMRT Sequencing data using the longest reads for contiguity and all reads to generate high-quality de novo assemblies with high consensus accuracy (>QV50)[7].

GENOME ASSEMBLY CONTINUITY AND CORRECTNESS



Published comparison between different sequencing platforms on the continuity and correctness of genome assemblies for different microbial strains. All PacBio assemblies had QV scores of >60, while MiSeq assemblies had an average QV of 51.[3]

CHARACTERIZE COMPLETE MICROBIAL GENOMES AND METHYLOMES



Scientists used SMRT Sequencing to resolve the genome of Burkholderia pseudomallei and compared their assembly results with hybrid and short-read assemblies. They concluded that PacBio long-read sequencing generated the highest quality, most cost-effective assembly and enabled closure of a complex microbial genome.[8]

KEY REFERENCES

1. Procedure & Checklist – Preparing Multiplexed Microbial SMRTbell® Libraries for the PacBio® Sequel® System.
2. Application Note: Microbial Multiplexing Workflow on the Sequel System.
3. Koren S. et. al. (2013) Reducing assembly complexity of microbial genomes with single molecule sequencing. *Genome Biology*. 14, R101.
4. Procedure & Checklist – Preparing >30 kb SMRTbell® Libraries Using Megaruptor® Shearing and BluePippin™ Size-Selection for PacBio RS II and Sequel® Systems.
5. Tutorial: Multiplexed Microbial Assembly.
6. Analysis Procedure – Multiplexed Microbial Assembly with SMRT® Link v5.1.0.
7. Chin, C.S. et al., (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT Sequencing data. *Nature Methods*. 10, 563–569.
8. Teng, J.L. et al., (2017) PacBio but not Illumina technology can achieve fast, accurate and complete closure of the high GC, complex *Burkholderia pseudomallei* two-chromosome genome. *Frontiers in Microbiology*, 8, 1448.

PN: BP101-053018