A streamlined workflow for high-throughput, multiplexed HiFi sequencing of microbial genomes

Meredith Ashby[1], Harsharan Dhillon[1], Aaron Wenger[1], Joan Wilson[1], Shreyasee Chakraborty[1], Davy Lee[1], Christopher Dunn[1], Ivan Sovic[1], Zev Kronenberg[1], Heather Ferrao[1], Primo Baybayan[1]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

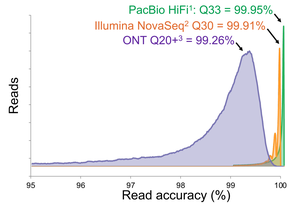
Introduction

Highly accurate, long HiFi reads produced by the PacBio Sequel IIe system have brought new levels of contiguity, completeness, and accuracy to large genome assembly. HiFi reads are similarly beneficial for microbial genome assembly, as higher quality assemblies enhance our ability to investigate foodborne illnesses and monitor antimicrobial resistance. However, obstacles in library preparation workflow and cost have limited use in public health. Here, we introduce a highly parallelizable shearing method, a streamlined library prep workflow, and an assembly algorithm based on HiFi reads that enables a high throughput, end-to-end solution for microbial genome assembly.

High raw accuracy enables workflow improvements

- Long reads improve assembly contiguity. However, once median read length exceeds the median repeat length of a genome, accuracy improvements yield greater gains in contiguity than longer reads.
- Because even very similar repeats can be distinguished with HiFi reads, shorter libraries can produce the same or better assembly results.
- Tolerance for shorter input DNA simplifies sample handling and allows closed genomes even from lower quality samples.

Figure 1. Raw accuracy by technology

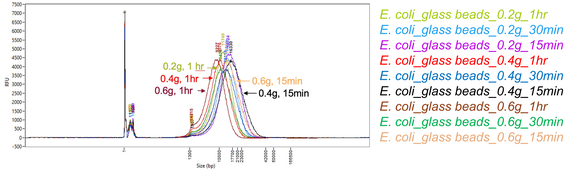


- Higher raw accuracy reduces the coverage requirements for assembly, increasing sample throughput and reducing cost per sample

Scalable, low-cost DNA shearing

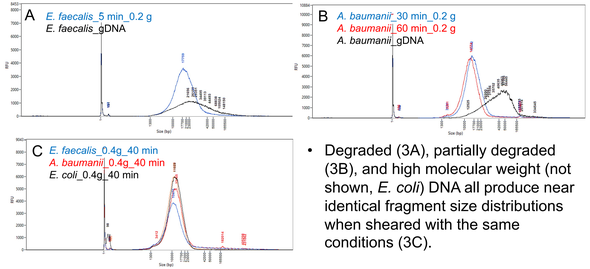
- The need for high molecular weight DNA can add complexity, cost and time to long read library preparation. In addition, Megaruptor 3 system and g-TUBE shearing do not scale efficiently beyond 8 or 24 samples, respectively.
- We present a novel DNA shearing method compatible with HiFi sequencing using acid washed glass beads (Sigma Aldrich, G8772) and agitation on a Vortex Genie2.

Figure 2. Glass bead shearing of high molecular weight E. coli gDNA produces a tight peak



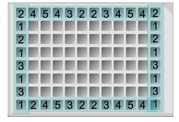
- 300 ng of E. coli gDNA in 50 µL of water was added to 0.2–0.6 g of glass beads in 300 µL of water. Samples were vortexed for 15 min – 1 hr at 3,000 rpm.
- The distribution of fragment sizes can be adjusted by changing the mass of beads or the vortex time.

Figure 3. Glass bead shearing produces consistent shears even with variable quality input DNA

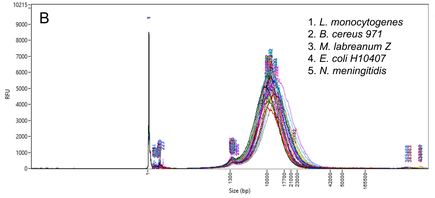


- Degraded (3A), partially degraded (3B), and high molecular weight (not shown, *E. coli*) DNA all produce near identical fragment size distributions when sheared with the same conditions (3C).

Figure 4. Up to 36 samples can be sheared in parallel per vortex device



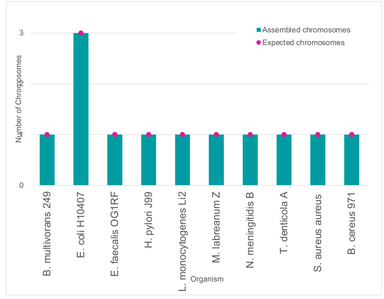
- A volume-matched scoop was used to aliquot ~0.2 g of beads into 36 outer wells of a V-shaped bottom, 96 deep-well plate.
- Aliquots of 5 microbial samples were added to wells as marked. The plate was sealed with optical film and secured to the Vortex Genie2 using a fit-for-purpose foam head attachment.



- Samples were sheared for 1 hr at 3,000 rpm, producing overlaid distributions with peaks between 9 kb–13 kb.

Figure 5. Sequencing of glass bead sheared microbial samples

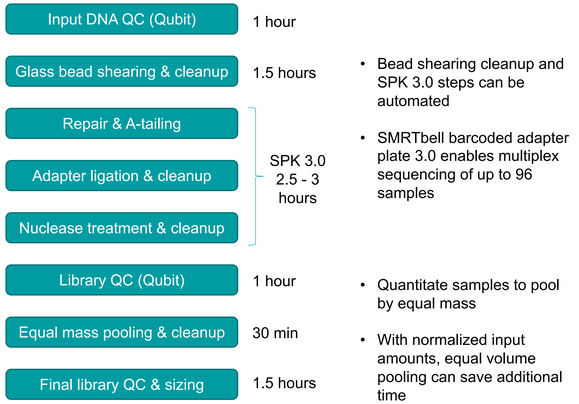
- 10 diverse microbial samples were sheared, prepped, and sequenced on the Sequel IIe system in 1 SMRT Cell 8M.
- The data was down-sampled to 30-fold coverage and assembled using the Microbial Assembly workflow in SMRT Link.
- All samples produced closed, circular genomes



Simpler, automatable SMRTbell library prep

- The new SMRTbell library prep kit (SPK3.0) has fewer steps, fewer reagents, and is automatable.
- Paired with glass bead shearing, SPK 3.0 creates an efficiently scalable high throughput workflow.

Figure 6. SMRTbell library prep workflow for up to 96 samples

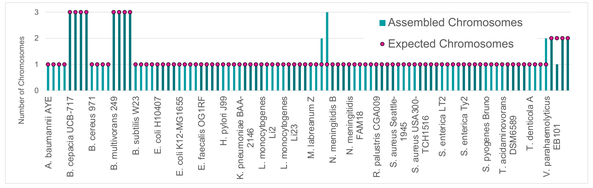


- Bead shearing cleanup and SPK 3.0 steps can be automated
- SMRTbell barcoded adapter plate 3.0 enables multiplex sequencing of up to 96 samples
- Quantitate samples to pool by equal mass
- With normalized input amounts, equal volume pooling can save additional time

HiFi sequencing and assembly results

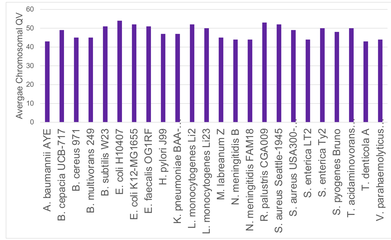
The PacBio Microbial Assembly tool, which launches automatically after sequencing, leverages HiFi accuracy and read length to address the unique challenges of bacterial genome assembly with distinct chromosomal and plasmid assembly stages, circular-aware polishing, oriC rotation, and NCBI-ready output.

Figure 8. HiFi data produces closed genomes with default settings



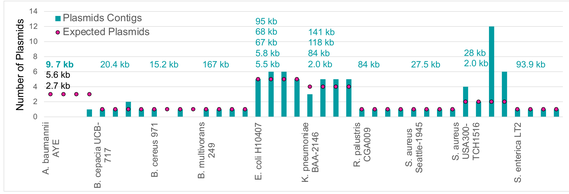
- (Fig. 8) 96 SMRTbell libraries were made from 24 distinct bacteria relevant to food safety, hospital outbreaks, and infectious disease, encompassing a wide range of GC content, genome size, and genome complexity.
- Without any manual curation, nearly all samples yielded closed chromosomes. Those that did not could be closed with hifiasm using custom settings.

Figure 9. Microbes assembled with HiFi data have gold standard accuracy



Accuracy of representative examples from the 96-plex sequencing run

Figure 10. HiFi sequencing consistently recovers closed plasmids



- Expected plasmid sizes are noted above samples grouped by species. Plasmids in green were recovered in at least one replicate.
- Plasmid recovery is shown using default parameters without curation. Manual review can improve contiguity (*S. aureus USA300-TCH1516; E. coli H10407*) and identify assembly artifacts.
- Recovery of smaller plasmids may require use of smaller insert sizes.

Conclusions

- HiFi data allows highly accurate, closed microbial genomes to be produced with less coverage from shorter libraries, providing flexibility in DNA handling and increasing the number of samples that can be multiplexed.
- Glass bead shearing is an inexpensive, robust method of producing DNA suitable for HiFi sequencing in a highly parallel manner, removing a significant barrier to high throughput SMRT sequencing of microbial genomes.
- The SPK 3.0 further simplifies the SMRTbell prep workflow, reducing handling times and enabling automation by eliminating size selection and time-consuming manual QC steps.
- The Microbial Assembly workflow in SMRT Link consistently produces highly accurate, closed genomes from HiFi data with default settings, completing an end-to-end high throughput workflow for microbial whole genome sequencing.

References

1. PacBio HiFi: HG003 18 kb library, Sequel II system chemistry 2.0, precisionFDA Truth Challenge V2
2. Illumina: HG002 2×150 bp NovaSeq library, precisionFDA *Truth Challenge* V2
3. Oct 2021 GM24385 Q20+ Simplex Dataset Release