Low-Input Single Molecule HiFi Sequencing for Metagenomic Samples

Meredith Ashby[1], Shreyasee Chakraborty[1], Daniel Portik[1], Primo Baybayan[1], Jonas Korlach[1], Sergey Nurk[2], Sergey Koren[2], Karen Lolans[3], Dawn Gratalo[4], Mark Driscoll[4], Christopher Quince[5], A. Murat Eren[3], Adam M. Phillippy[2]

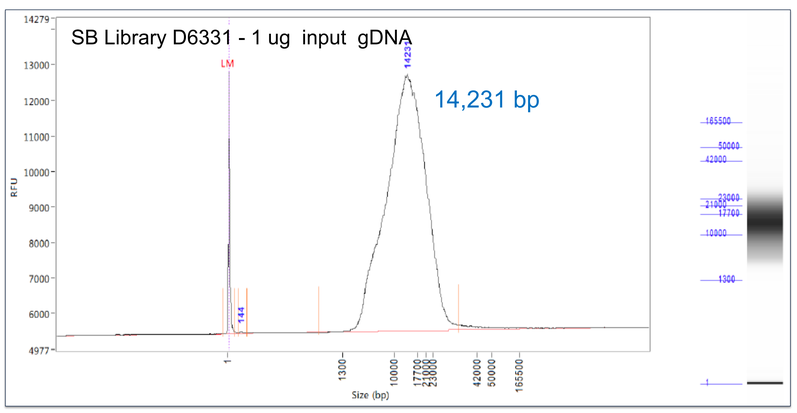
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025; [2] National Human Genome Research Institute, Bethesda, MD, USA ; [3] University of Chicago, Microbiology, Chicago, IL, USA; [4] Shoreline Biome, Framington, CT; [5] University of Warwick, Coventry, UK

Abstract

HiFi sequencing on the PacBio Sequel II System enables complete microbial community profiling of complex metagenomic samples using whole genome shotgun sequences. With HiFi sequencing, highly accurate long reads overcome the challenges posed by the presence of intergenic and extragenic repeat elements in microbial genomes, thus greatly improving phylogenetic profiling and sequence assembly. Recent improvements in library construction protocols enable HiFi sequencing starting from as low as 5 ng of input DNA. Here, we demonstrate comparative analyses of a control sample of known composition and a human fecal sample from varying amounts of input genomic DNA (1 ug, 200 ng, 5 ng), and present the corresponding library preparation workflows for standard, low input, and Ultra-Low methods. We demonstrate that the metagenome assembly, taxonomic assignment, and gene finding analyses are comparable across all methods for both samples, providing access to HiFi sequencing even for DNA-limited sample types.

DNA Extraction with Shoreline Breaker

Figure 1. Shoreline Breaker was used to extract consistently high molecular weight DNA from the ZymoBIOMICS Gut Microbiome Standard D6331, with peaks >10 kb.



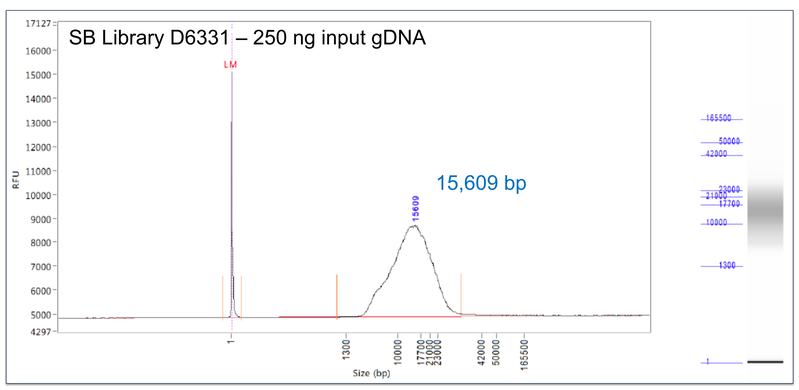
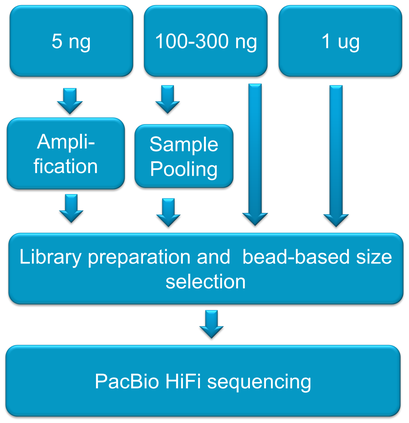


Figure 2. The Shoreline Breaker is a small, lightweight instrument that can perform mechanical lysis in one minute on up to 48 samples. The device improves lysis for difficult Firmicutes and spores while preserving HMW DNA, and is ideal for downstream applications including low-input library preparation.

Workflow

Figure 3. Sample preparation options for metagenomics samples.



Choose sample prep protocol based on DNA availability

For the Ultra-Low protocol, use proprietary (PCR-based) SMRTbell gDNA Sample Amplification Kit

HiFi SMRTbell library prep with SMRTbell Express Template Prep Kit 2.0 and size selection with diluted AMPure PB beads.

Sequel II System

HiFi Reads on the Sequel II System

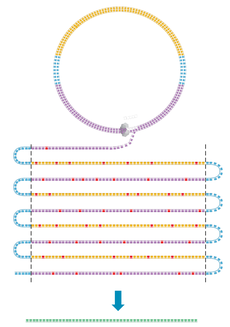


Figure 4. Sequencing advances on the Sequel II System have allowed the CCS method to be applied to far longer insert libraries than before. With average read lengths up to 100 kb and only 3 passes required to achieve Q20,10 kb shotgun metagenomics libraries can be sequenced at very high single molecule accuracy.

Sequencing and Assembly Performance

Table 1. For each of the libraries, greater than 1.9 million highly accurate HiFi reads (>Q20) with average subread length of 8–10 kb was achieved in a single SMRT Cell 8M. *De novo a*ssemblies of all three libraries had similar total sizes but somewhat different contig N50s.

Method	>Q20 BC reads	>Q20 Read Quality	Avg Read Length	Assembly Size	Contig N50
Ultra-Low input (ULI)	2,480,208	Q38	8,601 bp	139 Mbp	33,430
Low Input (LOW)	2,770,833	Q39	9,305 bp	120 Mbp	57,434
Standard Input (STD)	1,978,852	Q39	9,093 bp	106 Mbp	146,849

Figure 5. The GC content of HiFi reads was highly similar across the three methods, including the PCR-dependent Ultra-Low method. Kmer counts (K=22) were obtained from the HiFi reads and subsequently binned based on Kmer GC content. Comparisons of frequency counts for AT-rich and GC-rich bins across samples showed no significant differences occurring in the Ultra-Low sample, indicating GC biases are not created during the PCR step.

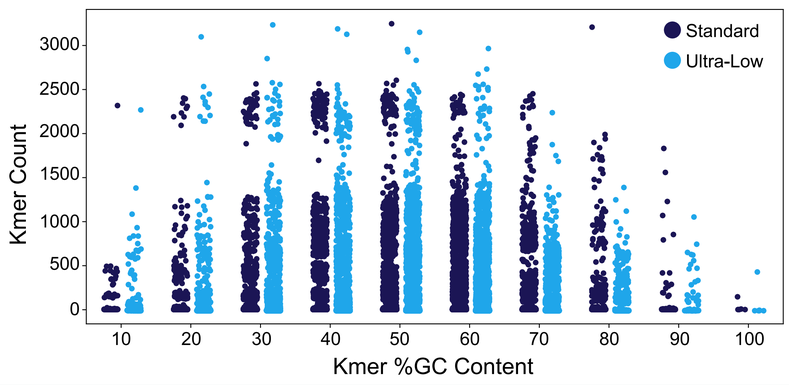


Figure 6. The relative abundances of 14 species detected differed from the theoretical values for the mock community. However, the inferred abundances were largely consistent across the three preparation methods, including the PCR amplification-dependent ULI method. Abundances are based on total aligned bases and are shown as relative proportions.

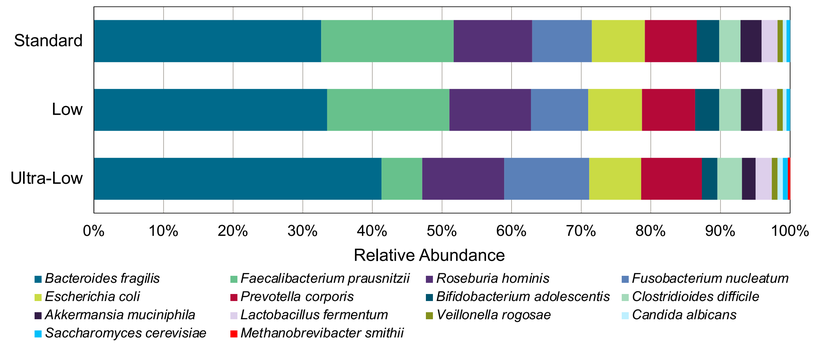
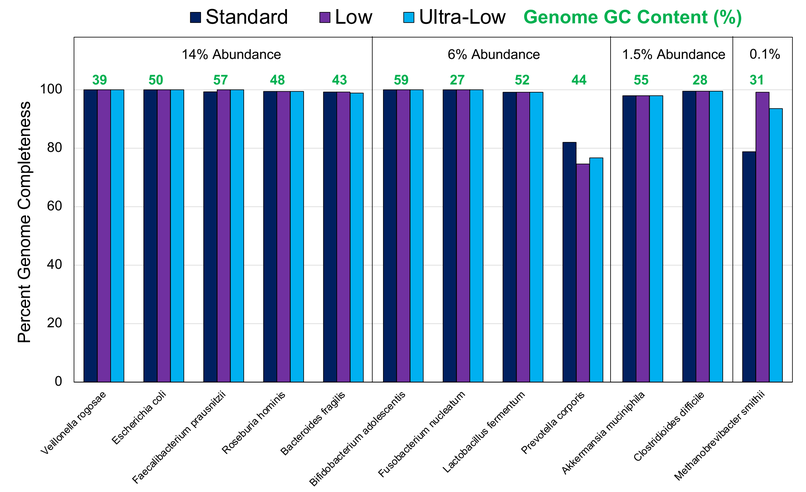
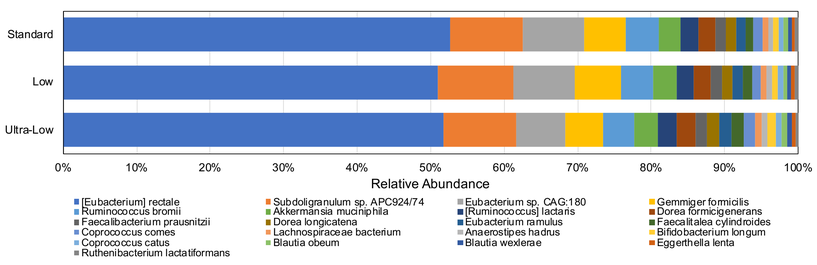


Figure 7. For all three preparation methods, the genomes of species with abundances as low as 1.5% were complete. As few as 2,500 reads or 10-fold coverage were sufficient to reconstruct a complete bacterial chromosome. Notably, a closed MAG was generated with 10-fold coverage for a 0.1% abundant species.



Fecal Sample Sequencing

Figure 8. A human fecal sample was also prepared with the standard, low, and Ultra-Low methods and the results were compared. The relative abundances of the top 21 species detected in the sample were consistent across all three methods.



Conclusions

These new methods will allow HiFi sequencing to be used even for sample types with limited starting material, allowing for reference quality MAGs, discovery of novel genes without the need for *de novo* assembly, identification of biochemical pathways, and resolution of taxa below the species level with confidence.