An Efficient Enzyme-Based Library Workflow for Multiplexed Long Read Whole Genome Sequencing

Maura Costello, Zac Zwirko, Christianto Putra, Ashley Silvia, Gavin Rush, Joseph Mellor

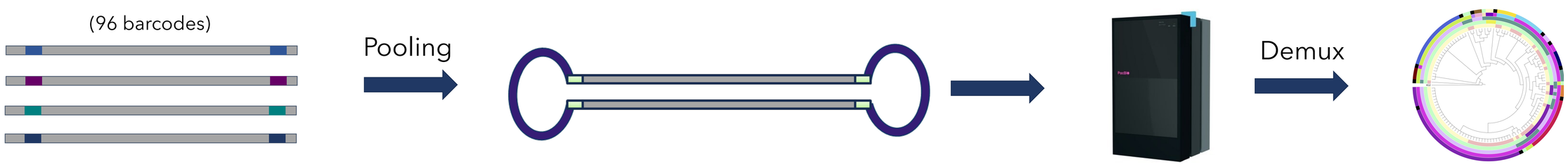
seqWell, Inc. Beverly, MA USA

Introduction

Long read sequencing technologies are unlocking new frontiers in genomic analysis. The recently released Revio™ instrument from Pacific Biosciences provides 15X higher throughput than the Sequel II™ enabling long read applications at unprecedented scale. Despite this increase in sequencing throughput, challenges in Pacific Biosciences' library construction protocols remain unresolved. Sample preparation costs are prohibitively high, fragmentation is performed through low-throughput mechanical means, and multiple micrograms of genomic DNA are typically needed to make a high-quality library.

To address these issues, we've developed a novel method that uses our LongPlex™ transposition kit to simultaneously fragment and index genomic DNA for subsequent pooling and multiplexed library construction using the SMRTbell® prep kit 3.0 (Pacific Biosciences). As proof of concept, we applied our approach to reference DNA (ATCC) from 8 different microbes spanning a broad range of GC contents using 250ng as input into tagmentation.

Workflow: LongPlex™ Multiplexed Library Construction for Microbial Sequencing and Assembly



STEP 1:
Fragmentation and indexing
Optional PCR

STEP 2:
Library construction and barbell ligation on multiplexed sample pool

STEP 3:
Sequencing on PacBio

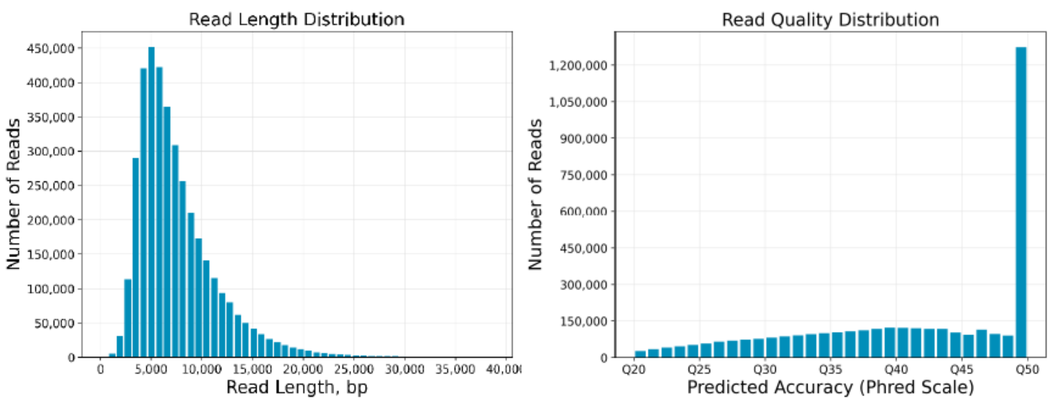
STEP 4:
Assembly of microbe genomes

Proof of Principle: 8-plex Library Construction on Microbes with Varied GC-Content

Experimental design:

- Tagmentation was performed on two sets of replicates using 250ng of genomic DNA from 8 different microbial isolate strains (ATCC) with GC-content ranging from 29-69% (Table 1)
- One set of replicates was amplified via PCR before pooling (PCR-plus) and one set was pooled during purification of tagmentation reactions (PCR-free)
- Each pool was carried into multiplexed library construction using SMRTbell® prep kit 3.0 (Pacific Biosciences)
- Resultant multiplexed libraries were pooled and sequenced on the Revio™ instrument using a 25M flowcell
- Reads were demultiplexed using Lima, alignment data was collected via Picard, and assembly was performed using FLYE

Results:



Pool ID	HiFi Reads	HiFi Yield	HiFi Read Length	HiFi Read Quality
PCR+	2,699,332	20,932,012,278	7754	Q32
PCR-	1,098,305	7,491,698,994	6821	Q33

Figure 1. Sequencing Quality Metrics HiFi read counts, insert lengths, and quality scores from multiplexed libraries generated with or without PCR before library construction.

Organism	%GC	Genome Size (Mb)	PCR-free Max Contig (Mb)	PCR-plus Max Contig (Mb)	PCR-free Coverage	PCR-plus Coverage
Clostridioides difficile	29	4.3	4.1	4.1	160	565
Staphylococcus epidermidis	32	2.6	2.6	2.6	396	1292
Bacillus cereus	35	5.4	5.4	5.4	74	90
Bacillus subtilis	44	4.2	1.2	4.2	23	391
Escherichia coli	51	4.6	4.6	2.4	191	1156
Enterobacter cloacae	55	5.3	5.3	0.0	136	0
Bordetella pertussis	67	4.0	4.0	1.1	259	741
Rhodobacter sphaeroides	69	4.5	3.2	1.9	209	688

Table 1. Coverage and Maximum Contig Length by GC Content and Genome Size Average deduplicated coverage and maximum contig sizes generated with and without PCR bef ore library construction.

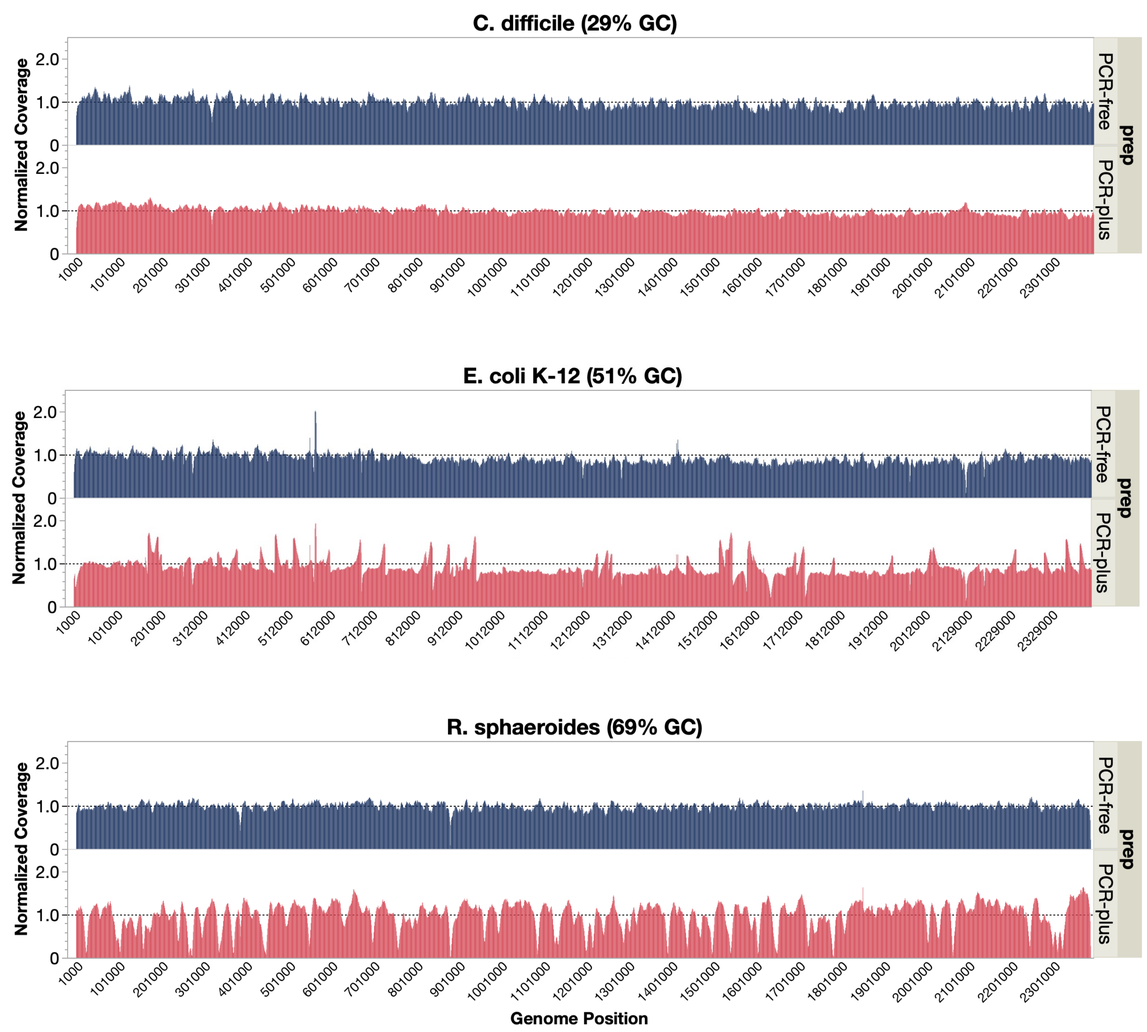


Figure 2. Uniformity of Coverage with Varied GC-content Normalized deduplicated coverage across 3 bacterial genomes with high, low, and medium GC-contents calculated using 1000 base windows. Multiplexed libraries created with PCR shown in red, PCR-free libraries shown in blue.

Summary and Conclusions

- LongPlex reagents enable simultaneous indexing and fragmentation up to ~8 kb allowing 8 or more samples to be successfully multiplexed upstream of Pacific Biosciences SMRTbell® prep kit 3.0 library preparation
- Performing PCR after tagmentation increases the mass of fragmented and barcoded material available for SMRTbell prep and SMRT cell loading, but amplification can lead to decreased uniformity of coverage for samples of higher GC contents (Figure 2)
- Higher plex pooling of up to 96 LongPlex barcoded samples into a single SMRTbell prep can lead to significant cost saving in SMRTbell library reagents
- LongPlex is currently in Early Access. For more information, please contact earlyaccess@seqwell.com