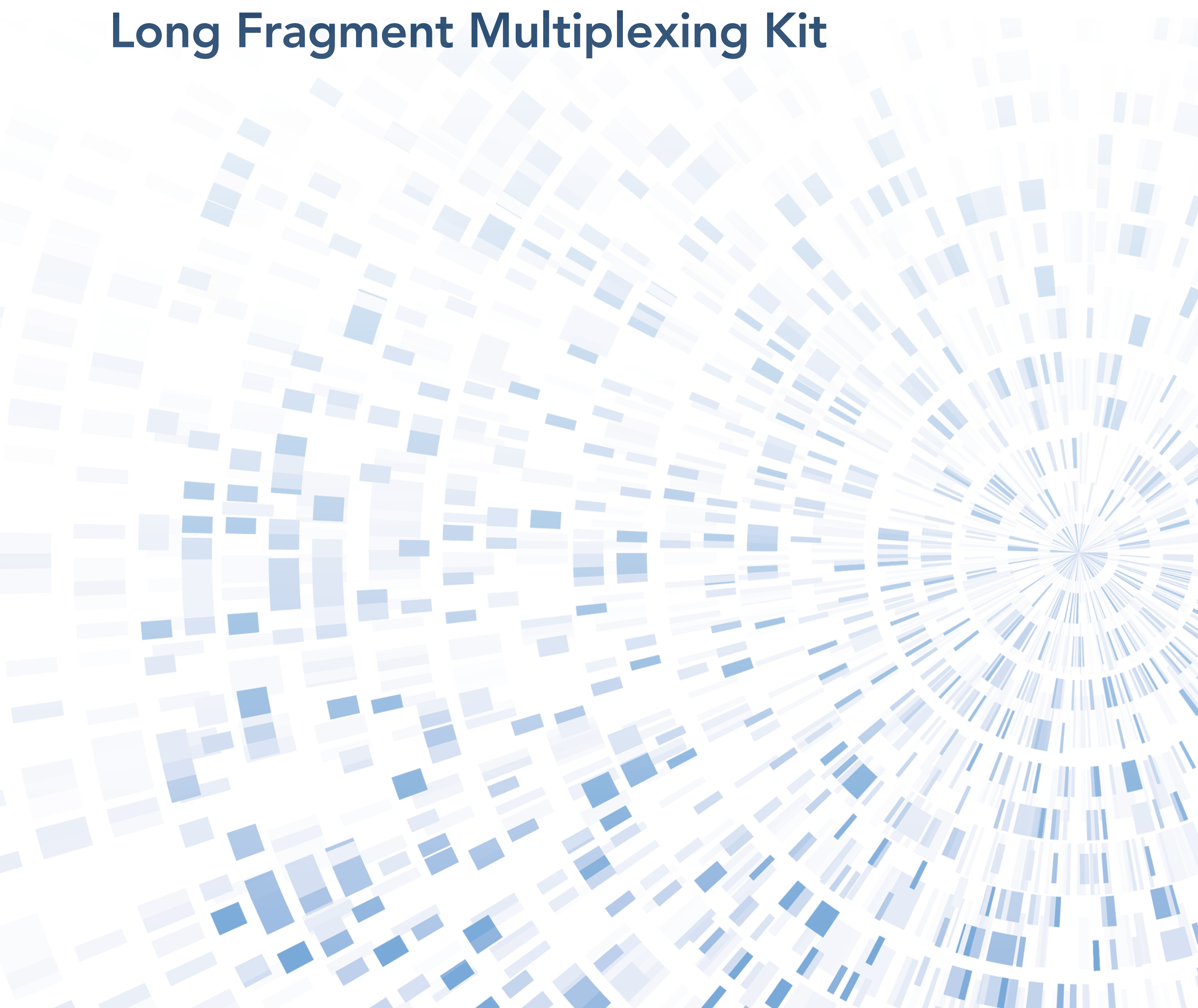
Application Note

seqWell | PacBio

Enabling Highly Scalable Microbial Whole Genome Sequencing on PacBio® HiFi Long-Read Systems using LongPlex™ Long Fragment Multiplexing Kit



LongPlex Long Fragment Multiplexing Kit Key Features:

- **SPEED:** Simple, automation-friendly workflow reduces time and steps
- **PERFORMANCE:** Simultaneous fragmentation and indexing without compromising library quality
- **SCALABILITY:** 96 Unique Dual Indexes (UDI) can be combinatorically expanded with PacBio indexing
- **COST-EFFECTIVE:** Pool up to 96 samples with UDI indexing upstream of HiFi library prep

Introduction

Long-read sequencing is a powerful genomic tool that provides deeper insights beyond the reach of traditional short-read technologies. Advances in long-read technology accuracy and throughput, in addition to decreasing sequencing costs, have made large-scale applications such as microbial whole genome sequencing more feasible.

The PacBio Revio™ long-read system delivers up to 360 Gb[1, 2] of HiFi bases per day, with over 90% of bases at Q30 or higher. To realize the full throughput capacity of the Revio, it requires highly scalable production of sequence-ready SMRTbell® libraries.

Traditional library preparation methods that rely on mechanical DNA shearing to generate long molecule templates can be lengthy and resource-intense processes that become a bottleneck for high throughput projects without the right automation equipment. The LongPlex Long Fragment Multiplexing Kit uses a highly scalable plate-based tagmentation strategy to simultaneously fragment and tag gDNA to generate multiplexed pools of 8 – 10 kb (mean) DNA fragments that eliminates the need for mechanical shearing. LongPlex pools are then brought through standard SMRTbell library prep for sequencing on PacBio HiFi long-read systems (Figure 1). Additionally, PacBio SMRTbell indexed adapters can be added to LongPlex pools to enable combinatorial indexing to further increase throughput and reduce overall per-sample sequencing costs. On the PacBio Revio system, up to 384 microbial samples can be multiplexed on a single SMRT® Cell, and up to 1,536 (384 x 4) per sequencing run.

In this application note, we highlight the seamless combination of seqWell's LongPlex Long Fragment Multiplexing Kit and library prep using the PacBio SMRTbell prep kit 3.0 to produce sequence-ready multiplexed libraries for microbial whole genome sequencing on any of PacBio's HiFi long-read systems. Specifically, we demonstrate the ability to multiplex 24 samples, while consistently producing mean HiFi read lengths >6 kb with a high level of coverage uniformity for eight microbial genomes with highly variable GC content on the Revio system.

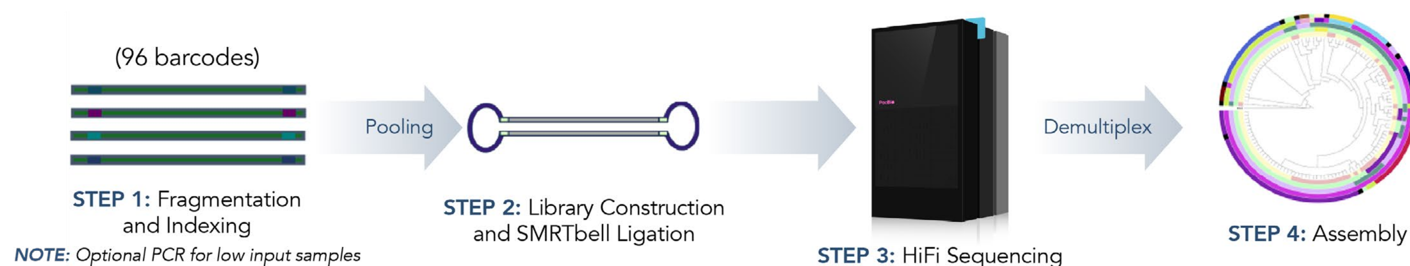


Figure 1: Workflow Scheme. Microbial genomic DNA samples are fragmented, tagged, and pooled using the seqWell LongPlex Long Fragment Multiplexing Kit. The pooled samples undergo library preparation using the PacBio SMRTbell prep kit 3.0, followed by sequencing on the PacBio Revio system. Sequence data are demultiplexed and assembled.

Materials and Methods

Genomic DNA (gDNA) and DNA quantification

Eight microbial gDNA preparations were obtained from the American Type Culture Collection (ATCC®). The set included microbes with GC-content ranging from 29-69% (Table 1). These samples were quantified using an Infinite® F200 PRO microplate reader (Tecan) and Quant-IT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher Scientific) prior to library preparation.

Organism	%GC	Genome Size (Mb)
Clostridioides difficile	29	4.3
Staphylococcus epidermidis	32	2.6
Bacilius cereus	35	5.4
Bacilius subtilis	44	4.2
Escherichia coli	51	4.6
Enterobacter clocae	55	5.3
Pseudomonas aeruginosa	66	6.8
Rhodobacter sphaeroides	69	4.5

Table 1. Summary of microbial genomic DNA

Library construction with LongPlex Long Fragment Multiplexing Kit and conversion using SMRTbell prep kit 3.0

Microbial gDNA samples from the eight microbes were processed in quadruplicate using the LongPlex Long Fragment Multiplexing Kit beginning with three different DNA input levels (250, 375, and 500 ng) for a total of 96 samples. Samples were processed according to seqWell's PCR-free protocol[3] (Figure 2). The 96 samples were pooled into four (4) 24-plex samples prior to LongPlex bead-based size selection. The concentration and fragment profiles of the four LongPlex pools were determined by Qubit (1X HS kit) and Agilent Femto Pulse with the Genomic DNA 165kb Analysis Kit. Each of the four 24-plex LongPlex pools was independently processed using a single SMRTbell library preparation following the manufacturer's protocol[4] resulting in four SMRTbell libraries. The four libraries were pooled prior to loading on the Revio SMRT Cell for a 96-plex sequencing run.

LongPlex offers both PCR-free and PCR-plus workflows for microbial whole genome sequencing. The PCR-free workflow is best suited for samples with high-quality DNA or if methylation data is needed. An optional PCR step can be performed following sample tagmentation and bead clean up, prior to sample pooling, to accommodate lower DNA input amount, difficult or degraded samples.

- The LongPlex workflow is a rapid, automation-friendly method that eliminates mechanical DNA shearing.
- LongPlex enables flexible batch sizes that accommodate different levels of multiplexing.
- The PacBio SMRTbell prep kit 3.0 contains all reagents for library preparation and libraries are compatible with any PacBio long-read system.
- Optional use of SMRTbell adapter indexes allows for combinatorial multiplexing for scaling to >96 samples per SMRT Cell.

Figure 2: Combined workflow. Go from gDNA to sample-ready, highly multiplexed libraries in less than 6 hours. The LongPlex workflow can be completed in less than 2 hours, requiring only 20 minutes of hands-on time to fragment, tag and pool samples. Pooled samples then undergo library preparation using SMRTbell prep kit 3.0 library prep which can be performed in under 4 hours.

Sequencing and data analysis

Sequencing of PacBio SMRTbell libraries was performed on a PacBio Revio system. The sequencing data were demultiplexed using a modified *Lima* pipeline[5]. Library QC metrics and sequencing statistics, such as library insert size, library complexity, and genome coverage, were calculated using standard tools from the Picard tools, and assembly was done using FLYE version 2.95.

Results & Discussion

LongPlex Long Fragment Multiplexing Kit: a simple, scalable workflow for long-read sequencing

The LongPlex Long Fragment Multiplexing Kit was designed with simplicity in mind. This transposase-based workflow allows for streamlined generation of multiplexed long fragment libraries averaging 8 – 10 kb DNA fragments containing full length unique dual indexes (UDI) adapters in a single enzymatic step. Following LongPlex, pooled samples undergo SMRTbell library preparation to attach SMRTbell adapters (Figure 3).

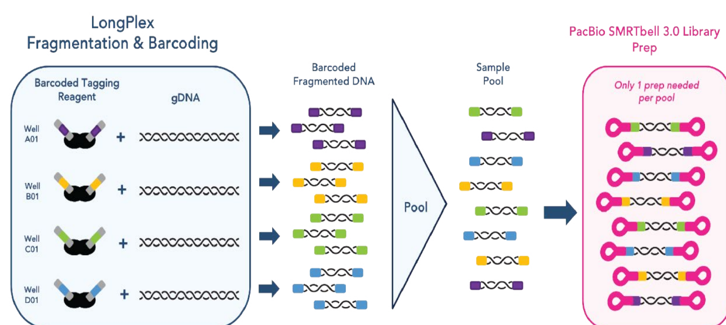


Figure 3: Molecular workflow. LongPlex uses loaded transposases to simultaneously fragment and index individual DNA samples. Indexed samples are pooled and SMRTbell adapters added via SMRTbell prep kit 3.0.

Combinatorial multiplexing

Each LongPlex kit contains 96 unique i7 and i5 indexes enabling processing of up to 96 samples. The i7-index and i5-index tagging reactions are carried out simultaneously. In these studies, three DNA input levels were tested, 250, 375, and 500 ng, in quadruplicate on 8 microbial species. Following the tagging reaction, 24 samples were pooled volumetrically, underwent purification, and size selection according to the user guide[3], creating four 24-plex LongPlex long fragment pools labeled with seqWell UDIs.

Each of the four (4) 24-plex LongPlex pools independently underwent SMRTbell library preparation. The resulting 4 SMRTbell libraries were subsequently pooled prior to sequencing using a single Revio SMRT Cell (Figure 4). This sequential barcoding approach allows for combinatorial multiplexing of LongPlex 96 UDI barcodes with 384 PacBio SMRTbell barcodes, making this workflow highly scalable, theoretically resulting in 36,864 unique combinations.

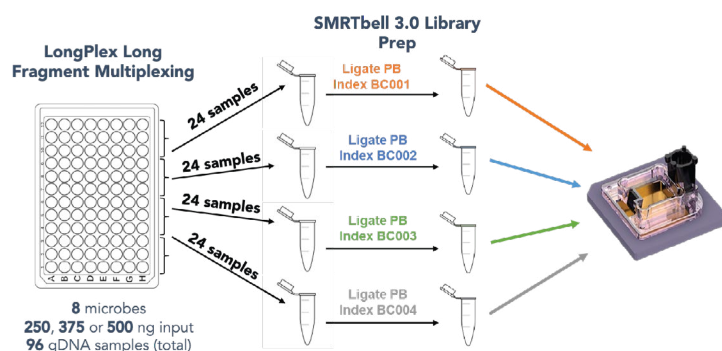


Figure 4: Multiplexing strategy. The early pooling of LongPlex prior to PacBio SMRTbell library preparation enables high production of massively multiplexed, sequence-ready libraries.

Reproducible generation of highly multiplexed sequence-ready, long-read libraries

Pooling of samples during LongPlex processing alleviates the need to individually purify samples. This greatly reduces the QC burden prior to SMRTbell library preparation and increases throughput capability and cost-effectiveness by lowering the number of SMRTbell library preps required without comprising on data quality.

Sample	Post LongPlex Femto size (bp)	Qubit Conc. Post LongPlex (ng/μL)	Total Material (ng)	Post SMRTbell prep conc (ng/μL)	Post SMRTbell Femto size (mean, bp)	Post SMRTbell Femto size (mode, bp)
24-Plex A	11,503	64.3	1800.4	26.1	12,472	8,443
24-Plex B	10,953	75.3	2018.4	31.8	11,444	9,302
24-Plex C	10,710	79.2	2217.6	15.1	11,902	9,463
24-Plex D	10,156	70.2	1965.6	18.7	10,112	9,490

Table 2. Summary of QC metrics post library preparation for each 24-plex pool

The QC results of the four (4) 24-plex library pools post LongPlex workflow show performance reproducibility and consistency across 8 different microbes with GC content ranging from 29 – 69% each at 3 different DNA input amounts (Table 2).

Highly reproducible fragment profiles were observed for the four (4) 24-plex pools following LongPlex tagmentation using a PCR-free workflow when assessed using the Femto Pulse (Table 2). The fragment sizes of the 24-plex libraries in these studies ranged from 10.1 - 11.5 kb. Additional QC was conducted following SMRTbell library preparation with final pool fragment sizes ranging from a mean of 10.1 – 12.4 kb in the region analysis of 5 – 20 kb and a mode of 8.4 – 9.5 kb (Figure 5).

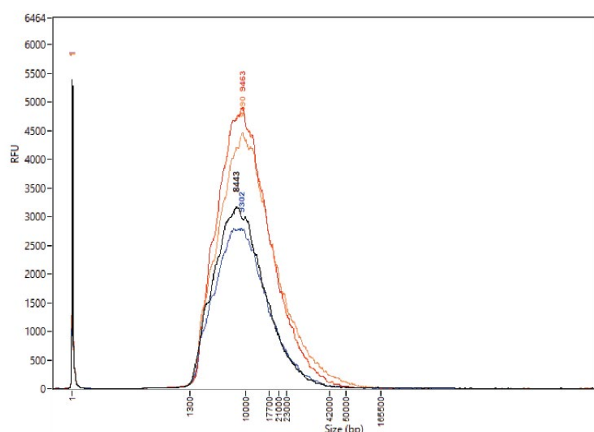


Figure 5: Fragment size distribution. Electropherograms show fragment sizes for the four (4) 24-plex pools following LongPlex and SMRTbell library prep processing as assessed using the Agilent Femto Pulse gDNA 165 kb Analysis.

High quality, long-read sequencing across a range of DNA input levels

The sequencing result shows a robust performance with a yield of 10.1 M HiFi reads, mean read length at 6.4 kb, and median HiFi read quality (Phred score) of Q46. All 24 samples were present in the data following demultiplexing for each 24-plex, indicating no drop-outs. The CV of read count between samples within each 24-plex ranged from 22% to 28%. Each 24-plex library pool yielded a demultiplexing rate >80%, with a range of 83.8% to 86.5% (Table 3).

Sequencing data show mean HiFi read length uniformity and total reads consistency (CV of 25.6% for all 96 samples), despite a 2-fold range of DNA input 250 – 500 ng and a wide variety of GC content bacteria from 29 – 69% (Figure 6 and Table 4). The ability to pool samples with varying input between 250 – 500 ng without compromising uniformity of read length increases efficiency by lowering the overall cost of reagents and labor, while maximizing the use of sequencing capacity. Uniformity of HiFi read lengths helps achieve even coverage across the genome, reducing gaps and ensuring all regions are adequately sequenced.

Microbial WGS with strong coverage uniformity across a range of GC content

The genome coverage is consistent for all four (4) 24-plex library pools. Figure 7 shows the uniformity of coverage across 8 bacterial genomes with various GC content. Uniformity of coverage ensures all regions of the genome are covered evenly, reducing the likelihood of missing important genetic information and minimizing sequencing bias, which can occur when some regions are overrepresented while others are underrepresented.

Six microbial genomes (GC content of 32 – 66%) assembled into a single full-length contig (Table 4), one of the key benefits of using long-read sequencing over short-read sequencing. The longer read length technologies

Sample	HiFi Reads	HiFi Read Length (mean, bp)	HiFi Read Quality (median)	% Demux	% Read CV	Mean Insert Size (bp)	% PF Reads Align	Mean Coverage	PCT_30X
24-Plex A	2,559,646	6,562	Q46	83.8	22.4	6,393	98.4	129.7	99.7
24-Plex B	2,099,116	6,423	Q46	84.6	26.0	6,258	98.4	105.6	99.2
24-Plex C	1,721,057	6,401	Q46	86.5	25.9	6,233	98.5	89.0	96.4
24-Plex D	1,785,947	6,374	Q46	85.5	28.2	6,208	97.5	87.7	95.2

Table 3. Summary of sequencing output for each 24-plex pool: demultiplex, alignment, and WGS metrics

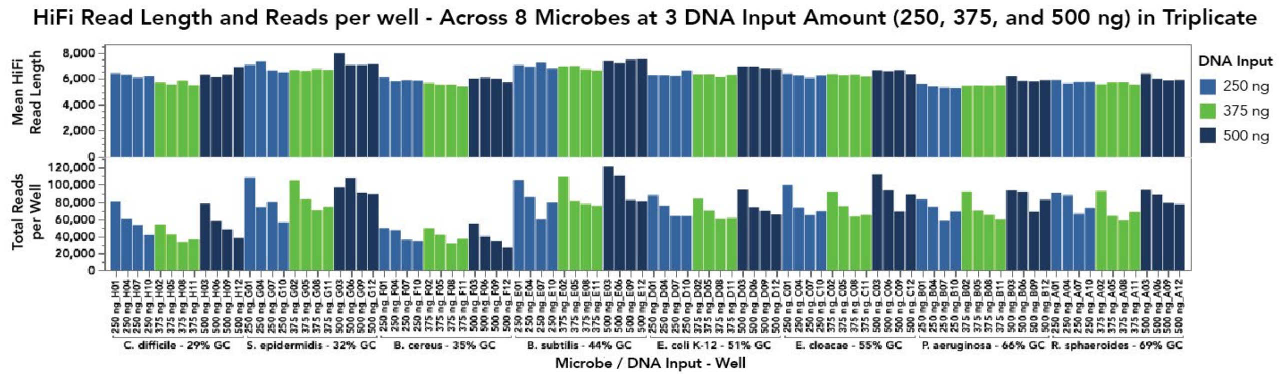


Figure 6: HiFi reads. Uniformity of mean HiFi read length and total reads of the four 24-plex across 8 different microbes with GC content ranging from 29 – 69%, each at 3 different DNA input amount: 250, 375, and 500 ng.

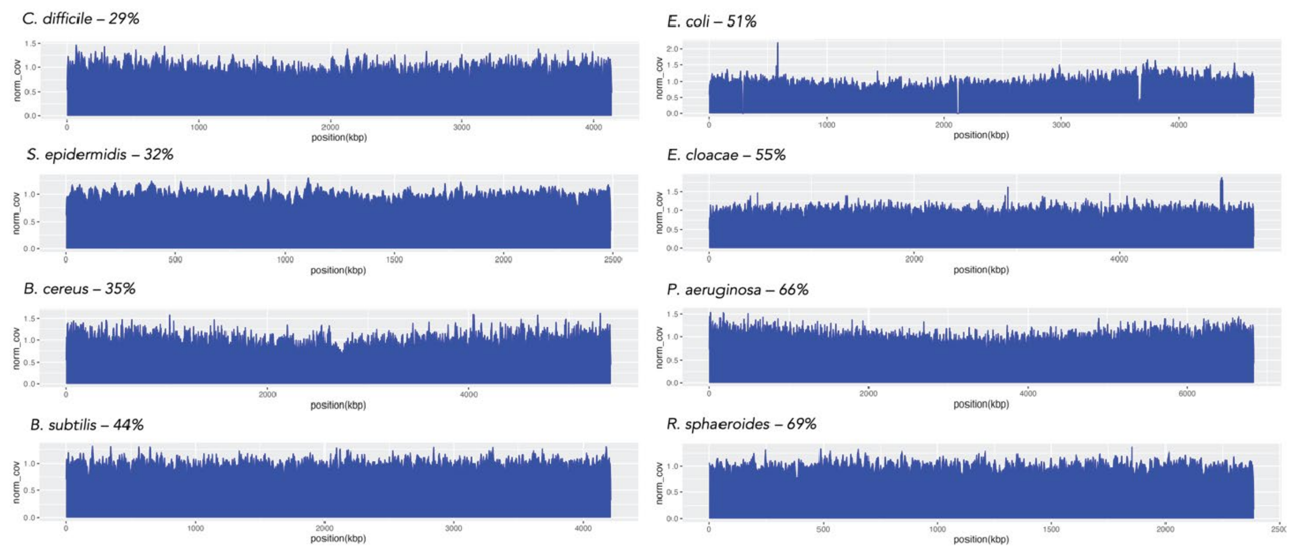


Figure 7: Uniformity of coverage across 8 different microbes with varied GC content. Normalized deduplicated coverage (y-axis) across 8 different microbes with GC content ranging from 29 – 69% calculated using 1000 base windows of genome position (x-axis).

like PacBio can cover larger portions of the genome in a single read, facilitating the assembly of genomes into fewer, longer contigs, and often achieving complete genomes in a single contig. This ensures that the entire genome is represented without gaps, providing a comprehensive view of the organism's genetic material. Moreover, single contig assemblies reduce errors and ambiguities that can arise from multiple contig alignments and scaffolding, leading to more accurate genomic data.

Two of the eight microbes missed several small contigs, two for *S. epidermidis* and one for *E. cloacae*. However, the largest contig size for those two microbes were successfully assembled. Additionally, ensuring high sequencing coverage helps increases the likelihood of spanning and accurately assembling high and low GC regions, which may be underrepresented. Adequate sequencing depth helps mitigate biases in sequencing that can arise from regions with extreme GC content (Figure 8).

Organism	%GC	Genome Size (Mb)	Genome Contig Count	Assembled Contig Count	PCR-free Max Contig (Mb)	PCR-free Max Coverage
Clostridioides difficile	29	4.3	3 (All circularized)	3	4.1	113
Staphylococcus epidermidis	32	2.6	5 (1 circularized)	3	2.6	289
Bacilius cereus	35	5.4	2 (1 circularized)	2	5.4	55
Bacilius subtilis	44	4.2	1 (All circularized)	1	4.2	180
Escherichia coli	51	4.6	2 (All circularized)	2	4.6	108
Enterobacter clocae	55	5.3	4 (3 circularized)	3	5.3	107
Pseudomonas aeruginosa	66	6.8	1 (All circularized)	1	6.8	73
Rhodobacter sphaeroides	69	4.5	10 (6 circularized)	10	2.4	118

Table 4. Summary of coverage and maximum contig length for 8 different organisms varied in genome size and GC content

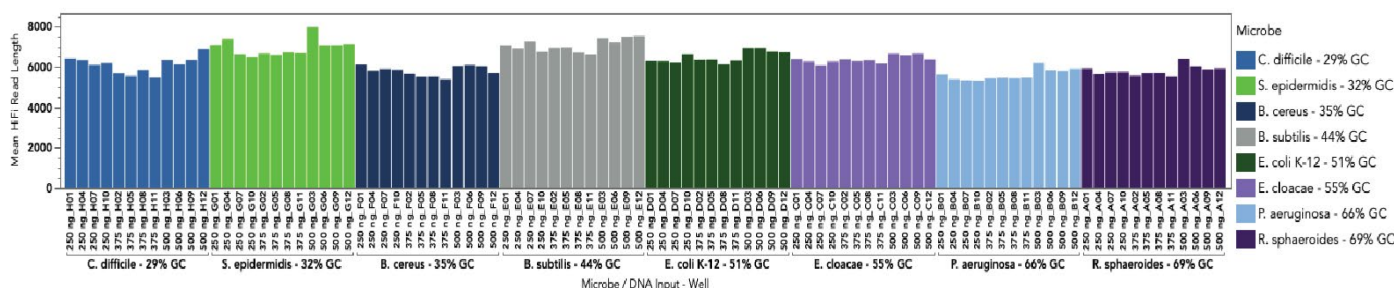


Figure 8: Uniformity of coverage upon sample multiplexing. Uniformity of mean HiFi read length of the four 24-plex across 8 different microbes with GC content ranging from 29 – 69%. A uniform coverage helps produce more accurate and contiguous genome assemblies by providing consistent data across the entire genome.

Conclusions

Realize the full capacity of your PacBio HiFi long-read sequencer

The LongPlex Long Fragment Multiplexing Kit enables scalable generation of long fragment sample pools upstream of SMRTbell library preparation supporting high throughput, long-read sequencing. Sample pooling and combinatorial multiplexing make the combination of LongPlex and PacBio HiFi long-read systems an **accurate, highly scalable, and cost-effective** solution for long-read microbial WGS.

These studies demonstrated the rapid generation of highly multiplexed, sequence-ready SMRTbell libraries with fragment profiles >8 kb. 24-plex libraries produced highly uniform sequence coverage for eight microbes with GC content ranging from 29 – 69% and for differing sample DNA inputs (250 – 500 ng) highlighting the ability to multiplex samples without sacrificing data quality.

Key LongPlex Long Fragment Multiplexing Kit benefits include:

- **SPEED:** Eliminate mechanical shearing using a fully scalable, enzymatic method to simultaneously fragment & tag genomic DNA
- **SIMPLICITY:** Automation-friendly, plate-based method with total workflow performed in <2 hours (20 minutes hands-on time)
- **SCALABILITY:** Massive multiplexing using 96 unique dual indexes (UDI) that can be combinatorially expanded with PacBio SMRTbell indexing
- **SAVINGS:** Early sample pooling greatly reduces all-in cost per sample for long-read sequencing without sacrificing data quality

References

[1]HiFi yield specification based on HG002/GM24385 human DNA extracted with Nanobind CBB kit and prepared with SMRTbell prep kit 3.0.

[2]HiFi yield is dependent on library fragment size. Yield is typically lower for shorter libraries.

[3]Protocol: LongPlex™ Long Fragment Multiplexing Kit. Download here.

[4]User Guide: Preparing whole genome and metagenome sequencing libraries using SMRTbell® prep kit 3.0. Download here. Samples should be eluted in 32 µl of low TE buffer and brought into the Repair and A-tailing step in the SMRTbell prep kit 3.0 protocol.

[5]LongPlex demultiplexing script – *lima* modified: Download here.

© 2024 seqWell Inc. All rights reserved.
Unless otherwise stated in references, presented data on file. For Research Use Only. Not for use in diagnostic procedures.
All product names and trademarks are the property of their respective owners.