Long-Read Low-Pass Sequencing: A Scalable Strategy for High-Resolution Genotyping

Maura Costello[1], Michelle Rahardja[1], Zac Zwirko[1], Christianto Putra[1], David Bays[1], Yanyan Liu[1], Sabina Gude[1], Joe Mellor[1], Kendall Lee[2], Josh Clevenger[2]

[1]seqWell, Inc. Beverly, MA USA [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL USA

Introduction

Long-read low-pass (LRLP) sequencing represents a new frontier in cost-effective, high-resolution genotyping for plant and animal breeding. Compared to short-read low-pass (SRLP) sequencing, LRLP delivers substantially improved genome coverage, mapping accuracy, and detection of structural variants (SVs) enabling a far more comprehensive representation of complex, repetitive, and polyploid genomes. Despite these clear benefits, broad adoption of long-read sequencing in breeding pipelines has long been constrained by throughput and cost limitations. seqWell's LongPlex™ technology provides a scalable, high-throughput fragmentation and multiplexing workflow that addresses these limitations by dramatically simplifying sample preparation to enable cost effective LRLP for large scale breeding projects. Here, we present performance data for LRLP sequencing compared to short-read data using twelve peanut (Arachis hypogaea) DNA samples. These results illustrate that LRLP sequencing, enabled by scalable technologies like LongPlex, is poised to redefine cost-effective genotyping in modern agriculture. By delivering high-resolution variant discovery at population scale, LRLP empowers more accurate genomic prediction, accelerates selection cycles, and expands the practical applicability of long-read sequencing across diverse breeding programs.

LongPlex™ Multiplexing Kit Workflow

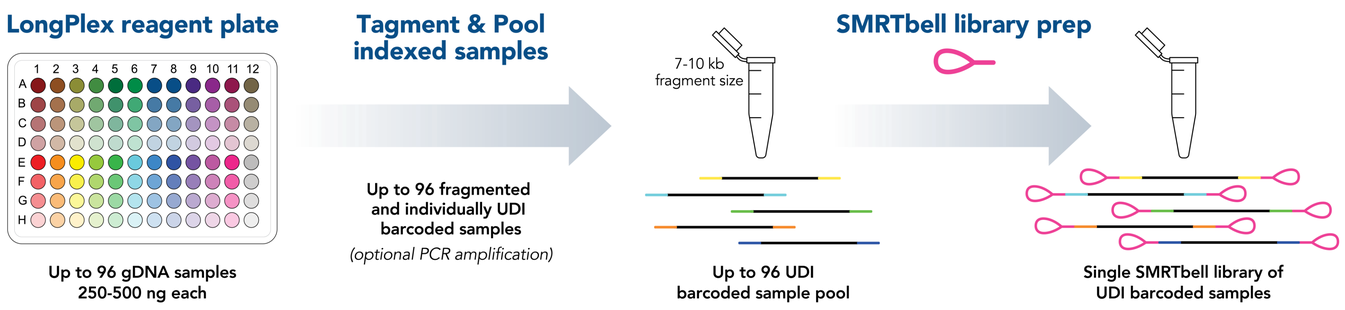


Figure 1. LongPlex uses Tn5 transposase to simultaneously fragment and barcode DNA samples in a rapid enzymatic workflow. Barcoded samples are then pooled together for further processing through PacBio SMRTbell prep kit 3.0. Post Revio sequencing, samples are demultiplexed using customized LIMA scripts followed by alignment and analysis.

Methods

- The LongPlex Multiplexing kit was used to fragment and barcode twelve peanut DNA samples. All 12 barcoded samples were then pooled for library construction using SMRTbell library preparation following the manufacturer's protocol (Figure 1) and sequenced on PacBio's Revio instrument on a 25M flowcell to approximately 1x coverage per sample.
- Sequencing data were demultiplexed using customized LIMA scripts. Large structural variations were identified using PacBio's pbmm2 and pbsv (v. 2.9.0) pipelines, with TRv2 as the reference genome.
- For comparison, short-read libraries were prepared using the Twist 96-plex Library Prep Kit and sequenced using Illumina NovaSeq 6000 system at Discovery Life Sciences (Huntsville, AL).

Long-read Sequencing Recovers Structural Variation Missed by Short-reads

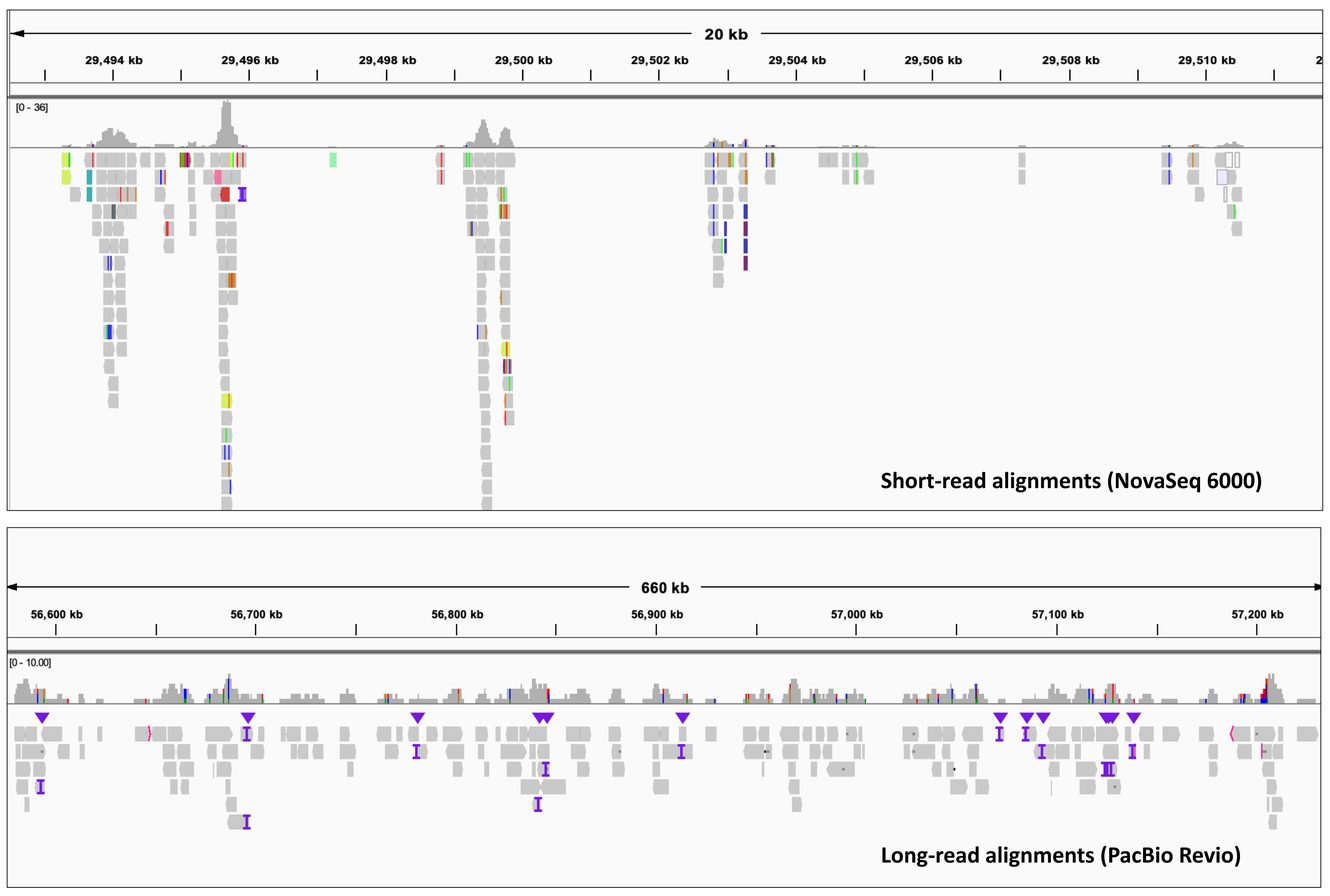


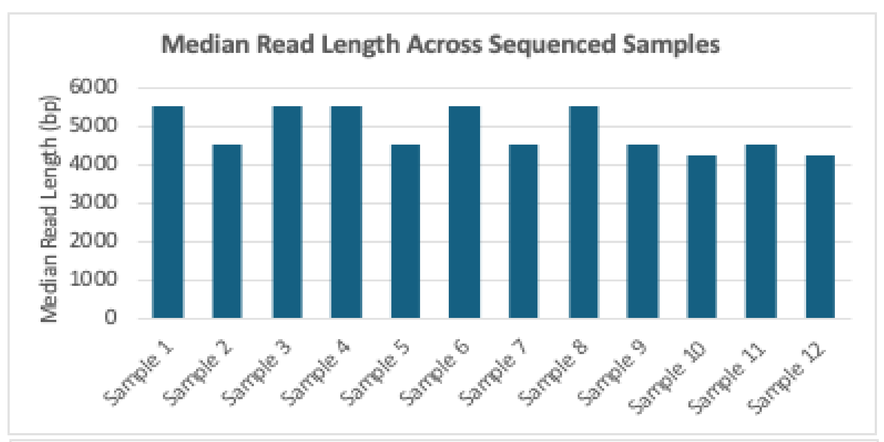
Figure 2. Long-reads deliver more contiguous and informative genome coverage compared to short-read low-pass sequencing. An Integrated Genomics Viewer (IGV) screenshot of short-read and long-read alignments to the peanut reference illustrate key differences in coverage patterns. Top: Short-read sequencing produces highly stacked, hence limiting the coverage across the genome. Bottom: Long-read low-pass sequencing using LongPlex provides broader, more even coverage, enabling detection of structural variants and improved representation of repetitive elements.

Table 1. Long-read low-pass sequencing reveals mid- and large-sized structural variants too large to be reliably detected by short-read methods.

Sample #	Average Insertion Length (bp)	Average Deletion Length (bp)	Sample #	Average Insertion Length (bp)	Average Deletion Length (bp)
Sample 1	336	3416	Sample 7	143	1154
Sample 2	213	1532	Sample 8	368	2232
Sample 3	192	1332	Sample 9	155	802
Sample 4	197	1001	Sample 10	257	2631
Sample 5	122	736	Sample 11	171	1070
Sample 6	169	904	Sample 12	378	3852

High Level Sequencing Metrics

- Revio HiFi sequencing yield 43.4 Gb with read lengths averaging 5.3 kb.



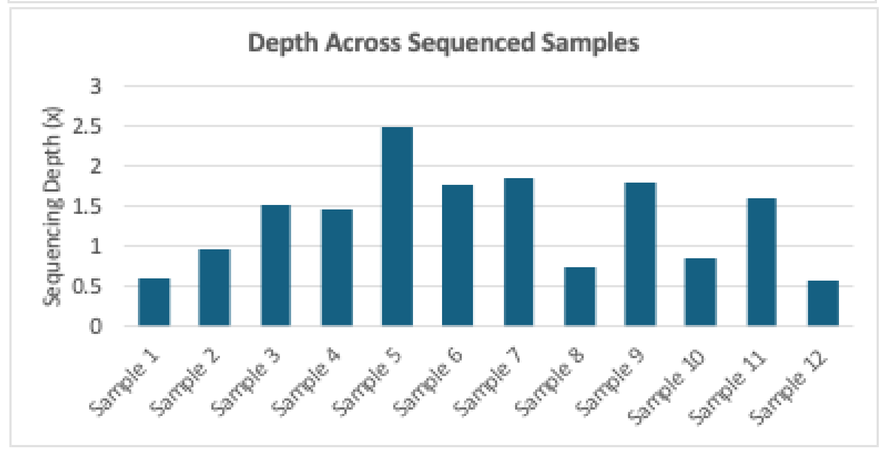


Figure 3. Top: Median read length measured for twelve sequencing samples, showing low variability ranging (CV of 11%) from 4,249 to over 5,499 bases. **Bottom:** Sequencing depth achieved across twelve long-read low-pass samples. Depths range from approximately 0.5x to 2.8x.

Summary and Conclusions

- Despite similar nominal depth (1.8x for short-read and 0.97x for long-read) long reads span gaps and low-mappability regions that short reads fail to cover, resulting in greater effective genome coverage (59% covered using long-read in comparison to 15% covered) and variant visibility at low pass depths (Fig. 2).
- Long-read low-pass sequencing revealed a broad spectrum of structural variation across samples, with average insertion sizes between 120 - 380 bp and deletion sizes ranging from 736 - 3,852 bp (Table 1). These SV sizes fall outside the detection limits of short-read technologies which have read-length constraints and mapping ambiguity in repetitive regions.