Best Practices for Whole Genome Sequencing Using the Sequel System

Nick Sisneros, Shreyasee Chakraborty, Sarah Kingan, Richard Hall, Joan Wilson, Christine Lambert, Kevin Eng, Emily Hatas and Primo Baybayan
PacBio, 1380 Willow Road, Menlo Park, CA 94025

Abstract

Plant and animal whole genome sequencing has proven to be challenging particularly due to genome size, high density of repetitive elements and heterozygosity. The Sequel System delivers long reads, high consensus accuracy and uniform coverage which enable more complete, accurate, and contiguous assemblies of these large, complex genomes. The latest Sequel chemistry can produce 5–8 Gb per SMRT Cell with reduced input SMRTbell libraries (as low as 5 pM). Read lengths averaging 10–15 kb can be routinely achieved, with the longest reads >60 kb. Furthermore, 50% of useable bases are in reads greater than 20 kb.

Here, we recommend the best practices for whole genome sequencing and de novo assembly of complex plant and animal genomes. Guidelines for constructing large-insert SMRTbell libraries (>30 kb) to generate optimal read lengths using the latest Sequel chemistry are presented. We also describe ways to maximize library yield per preparation from as little as 5 μg of sheared genomic DNA. The combination of these advances makes plant and animal whole genome sequencing a practical application of the Sequel System.

Library Construction Recommendations

Recommended Shearing Devices for Large-insert Fragments

<table>
<thead>
<tr>
<th>Recommended Shearing Devices</th>
<th>Large-insert Construction Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEMTO Pulse</td>
<td>High molecular weight DNA</td>
</tr>
<tr>
<td>High Molecular Weight DNA</td>
<td>- Pulsed Field Gel Electrophoresis</td>
</tr>
<tr>
<td>PFGE (power supply)</td>
<td>- PFGE (power supply)</td>
</tr>
<tr>
<td>- Optimization of shearing</td>
<td>- Proper size selection cutoff</td>
</tr>
<tr>
<td>- Post selection DNA Damage</td>
<td>- Following loading recommendations</td>
</tr>
<tr>
<td>Repair Improves Read Length</td>
<td></td>
</tr>
</tbody>
</table>

Large-insert Library Construction Workflow

1. Select the appropriate library size.
2. Run A FAGE run provides the library quality, using either the Biorad CHIEF Mapper (A) or Sugar Science Pippin Pulse 800 (B) PFGE instruments. The gel image (C) shows high molecular weight DNA (∼50 kb) which can support the desired size of DNA. Lane 4 shows a less than clear DNA with a smear up to 80 kb. Depending on the necessary degradation, the sample may be used directly for library construction. A size selection cutoff of 10 Mb usually generates good standard lengths.
3. While both CHIEF Mapper and Pippin Pulse are reliable systems for characterizing genomic DNA, electrophoresis run times are increased (>16 hrs) and require significant amounts of DNA as input. Alternatively, high-throughput capillary-based electrophoresis systems are available with a 10X reduction in run time and reduced consumption of large-insert libraries. More importantly, the system requires program mg quantities of DNA.
4. Human genomic DNA was also loaded on the CHIEF Mapper and Pippin Pulse. Separation observed in CHIEF Mapper is comparable performance on the Pippin Pulse.

Data Analysis

Hierarchical Genome Assembly Process (HGAP)

HGAP utilizes all PacBio data using the longest reads for contiguity and all reads to generate high-quality de novo assemblies with high consensus accuracy (>99%).

Diploid-aware Genome Assembly with FALCON

In diploid assembly, error-corrected reads are assembled using a string graph of the read overlap, generating primary and alternative contigs that represent the alternative alleles, or structural variants (SVs), between the haplotypes. FALCON is available through GitHub.

Summary and Resources

- The Sequel System achieves avg read lengths of 10–15 kb with throughput of 5–8 Gb per SMRT Cell 1M
- Follow best practices to improve performance and overall project results
- Pulsed Field Gel Electrophoresis is important for assessing incoming genomic DNA, sheared DNA, SMRTbell library and final size-selected SMRTbell Library
- The Megaruptor system is recommended for sequencing DNA >30 kb
- Optimize shearing conditions by performing tests prior to large-scale shearing
- Treat size-selected libraries with DNA Damage repair enzymes
- De novo assembly using either HGAP or FALCON algorithms

Resources:
For all PacBio library prep and sequencing protocols, visit https://www.pacb.com/support/documentation;
FALCON available on GitHub: https://github.com/PacificBiosciences/FALCON

References:

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2017 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Sequel, and Sequel are trademarks of Pacific Biosciences. Biorad and SageScience are trademarks of Biorad Laboratories and GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.