

This README file provides descriptive information for the contents found in this directory.

The dataset released in this directory contains the results of PacBio® SMRT® Sequencing for an *Arabidopsis thaliana* Ler-0 genome run on a Sequel™ System.

Data Generation:

Purified *Arabidopsis* genomic DNA was sheared to an average size of 32 kb and converted to SMRTbell™ templates, followed by a 20 kb size selection performed on a BluePippin™ system (Sage Science). Each SMRT Cell was loaded at an on plate concentration of 144 pmol of library and run for 6 hours on the Sequel System using the modified chemistry. Collectively, the two SMRT Cells produced 10.8 Gb of data, contained in 1.1 million reads, with half of the data in reads greater than 16,400 bp in length.

Raw Data combined 2 SMRT Cells:

Number of Reads	1,135,065
Number of Bases	10.8 Gb
Average Read Length	9,474 bp
Read N50	15,377 bp
Mapped Read Length N50	16,411 bp
Mapped Subread Length N50	14,852 bp
Mapped Read Length Max	53,610 bp
Mapped Concordance (Mode)	0.88
Mapped Concordance (Mean)	0.84

Assembly:

Data were processed using HGAP4 with a development version of SMRT Link v3.2, using a seed read cutoff of 6,000 bp for preassemble reads. Assembly was polished using the Arrow consensus caller. Post assembly processing included the following filtering: Contigs were filtered out if greater than 10% of contig bases are not polished due to low coverage.

Assembly size	122.9 Mb
Polished Contigs	238
Contig N50	10.4 Mb
Max Contig Length	15.0 Mb
Busco Complete Single copy genes	97.7%
Busco Fragmented genes	0.6%
Busco Missing genes	1.7%

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2016 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. Fragment Analyzer is a trademark of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.