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 several DNA samples run on a Sequel™ System.

Control Data:

2 kb Control Amplicon:
The Sequel SMRT Cell control template has an insert size of 1,966 bp. It was sequenced on a Sequel System using Sequel Sequencing chemistry v1.0 and Polymerase v1.0. It was run on the initial Sequel Instrument Control Software version 3.0.1 using one SMRT Cell 1M and data was collected for 4 hours. The reference for the dataset is provided as file "2kPacBioInternalControl".

Number of Reads	615,347
Number of Bases	6,158,392,776 bp
Average Read Length	10,008 bp
Read N50	17,708 bp
Read Length Max	66,055 bp
Single-pass Accuracy (Mode)	0.89
Single-pass Accuracy (Mean)	0.84

4.3 kb PlasmidBell Control:
The 4.3 PlasmidBell control has an approximate insert size of 4.3 kb. The sequence was constructed by linearizing PBR 322 plasmid followed by SMRTbell™ adapter ligation. It was sequenced on a Sequel System using Sequel Sequencing chemistry v1.1 and Polymerase v1.0. It was run on Sequel Instrument Control Software version 3.1 using one SMRT Cell 1M and data was collected for 6 hours. The reference for the dataset is provided as file "pBR322_plasmidbell_4361bp".

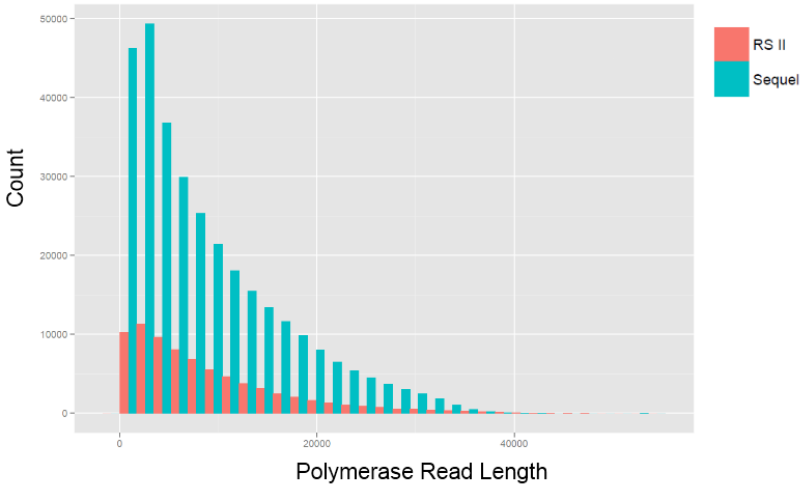
Number of Reads	403,715
Number of Bases	4,995,569,410 bp
Average Read Length	12,374 bp
Read N50	27,129 bp
Read Length Max	80,682 bp
Single-pass Accuracy (Mode)	0.86
Single-pass Accuracy (Mean)	0.83

Bacterial Data:

***E. coli* K12:**
This *E. coli* K12 dataset includes one SMRT Cell 1M of data generated with a Sequel System using Sequel Sequencing Kit v1.2 and Polymerase v1.0 run with Instrument Control Software v3.1.1. Data was collected for 6 hours. The library was sheared to a mode of 25 kb, converted to SMRTbell library, and size selected to 15 kb library using a BluePippin™ by Sage Sciences.

Number of Reads	314,591
Number of Bases	2,796,713,990 bp
Average Read Length	8,890 bp
Read N50	12,934 bp
Read Length Max	51,956 bp
Single-pass Accuracy (Mode)	0.87
Single-pass Accuracy (Mean)	0.83

Comparison of Sequel System and PacBio RS II[1] read length distribution



[1]PacBio RS II *E. coli* dataset: https://github.com/PacificBiosciences/DevNet/wiki/E.-coli-Bacterial-Assembly

***R. palustris*:**
This *R. palustris* dataset includes one SMRT Cell 1M of data generated with a Sequel System using Sequel Sequencing Kit v1.2 and Polymerase v1.0 run with Instrument Control Software v3.1.1. Data was collected for 6 hours. The library was sheared to a mode of 10 kb and converted to SMRTbell library.

Number of Reads	386,641
Number of Bases	3,681,982,243 bp
Average Read Length	9,523 bp
Read N50	17,335 bp
Read Length Max	42,884 bp
Single-pass Accuracy (Mode)	0.89
Single-pass Accuracy (Mean)	0.84

Reference Files:

Reference files provided for mapping results:

1. 2kb PacBio Internal Control, for mapping 2kb control data.
2. *Rhodobacter palustris* CGA0009
3. *Escherichia coli* K12
4. pBR322 plasmidbell, 4361bp, for mapping 4.3kb control data.