

Analysis Procedure – Multiplexed Microbial Assembly with SMRT® Link v5.1.0

Upload the Barcode Adapter Kit’s Barcode FASTA file to SMRT Link (One-Time Operation)

For microbial multiplexing, use the following 16 barcoded adapters contained in the Barcoded Adapter Kits (8A and 8B):

- Barcoded Adapter Kit 8A: bc1001, bc1002, bc1004, bc1008, bc1009, bc1010, bc1012, bc1014
- Barcoded Adapter Kit 8B: bc1015, bc1016, bc1017, bc1018, bc1019, bc1020, bc1021, bc1022

The following steps describe how to import the file containing the barcode sequences for both kits into SMRT Link for use in demultiplexing.

1. Download the FASTA file containing the barcode sequences using the following link: https://www.pacb.com/wp-content/uploads/Sequel_RSII_16_Barcodes_v2.zip
2. Unzip to produce the file `Sequel_RSII_16_Barcodes_v2.fasta`.
3. Upload this file to a location on the SMRT Link file system. (Example: By using scp on an internal server. Note that support from your Systems Administrator for SMRT Link may be needed to transfer this file to the file system.)
4. Open SMRT Link and select **Data Management**.
5. Click **View or Import Sequence Data**.
6. Click **Import** and choose **Barcodes FASTA**.
7. Navigate to the file `Sequel_RSII_16_Barcodes_v2.fasta` at the location chosen in Step 3.
8. Click **Import**. The barcode file is imported and becomes available for use.

Set Up a Run on the Sequel® Instrument

Open the Run Design module in SMRT Link. Click **New Run Design**. Fill in the Sample Information section, then click the small arrow to open **Barcoded Sample Options**. Specify the following options:

1. Sample is Barcoded: **Yes**
2. Barcode Set: **Sequel_RSII_16_Barcodes_v2**
3. Same Barcodes on Both Ends of Sequence: **Yes**
4. Autofilled Barcode Name File: Click **Download File**
5. Barcoded Sample Name File: In the downloaded Autofilled Barcode Name File, fill in the desired values for Bio Sample Name, as shown below for the first 3 barcodes. Save the file and upload it using **Select CSV File**.

	A	B
1	Barcode Name	Bio Sample Name
2	bc1001_BAK8A--bc1001_BAK8A	E. coli
3	bc1002_BAK8A--bc1002_BAK8A	S. aureus
4	bc1004_BAK8A--bc1004_BAK8A	B. subtilis

6. Save the Run Design and launch the sequencing run.

Review the Resulting Data Sets

Once the run is complete, open the SMRT Link Data Management module. The barcoded samples are automatically demultiplexed. Click the small arrow next to the parent sample (**My microbial samples** in this example) to view the barcoded Data Sets in a hierarchical display:

Data Set Name	Bio Sample Name	Barcode Name
▼ My microbial samples	[multiple]	[multiple]
My microbial samples (E. coli)	E. coli	bc1001_BAK8A--bc1001_BAK8A
My microbial samples (S. aureus)	S. aureus	bc1002_BAK8A--bc1002_BAK8A
My microbial samples (B. subtilis)	B. subtilis	bc1004_BAK8A--bc1004_BAK8A

Click the Data Set name of the parent sample (**My microbial samples** in the above example) to view the Data Set details. The Barcodes section displays the number of reads per barcode and several graphs to assess the quality of the demultiplexing results. Refer to the SMRT Link User Guide v5.1.0, Demultiplex Barcodes Application section for further details on interpreting this report.

If review of the demultiplexing results reveals that any demultiplexing parameters need to be corrected (such as barcode set, barcode layout, or minimum barcode score), launch the **Demultiplex Barcodes** analysis application manually from the SMRT Analysis module, using the same Data Set as input.

Launch Assembly Analyses on the Microbial Data Sets

To launch one assembly analysis for each barcoded sample, open the SMRT Analysis module. Click **Create New Analysis**. Specify the following options:

1. Analysis Application: **Assembly (HGAP 4)**
2. Analysis Name: Enter an analysis name to persist across each analysis.
3. Data Sets: Locate the parent Data Set of the demultiplexed set (**My microbial samples**, in the above example). Click the small arrow in front of its name to expand all child data sets below it. Click **All** next to the parent Data Set name to select all children.
4. Analysis Type: **One Analysis per Data Set – Custom Parameters**

Create New Analysis - Settings
✕ CANCEL ▶ START AND CREATE NEXT

Analysis Application *

Analysis of Multiple Datasets

Analysis Type *

Assembly

Genome Length ⓘ

ADVANCED ANALYSIS PARAMETERS

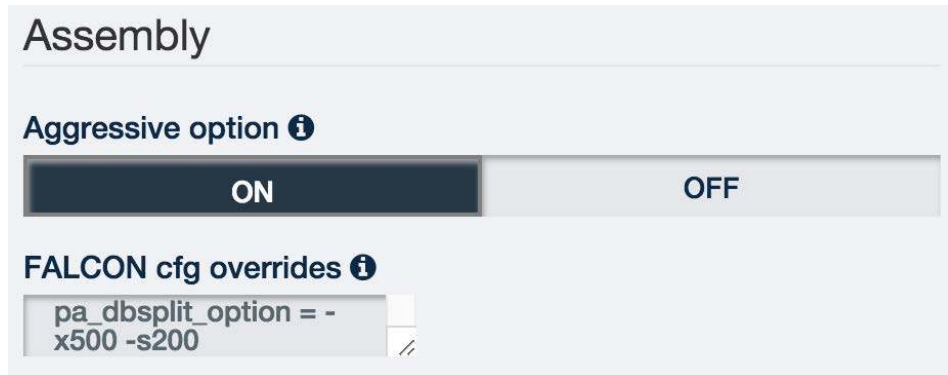
Analysis Name * ⓘ
 - My microbial samples (E. coli)

View:

Data Sets ☰

Data Set Name	Bio Sample Name
▼ My microbial samples All None	[multiple]
<input checked="" type="checkbox"/> My microbial samples (E. coli)	E. coli
<input checked="" type="checkbox"/> My microbial samples (S. aureus)	S. aureus
<input checked="" type="checkbox"/> My microbial samples (B. subtilis)	B. subtilis

- For the highlighted first Data Set, edit the **Genome Length** parameter if desired.
- Click **Advanced Analysis Parameters**. Make the following selections for best-practice microbial genome assembly: Change **Aggressive option** to **ON**, and under **FALCON cfg overrides**, paste the following text: `pa_dbsplit_option = -x500 -s200`



- Click **OK** to save the Advanced Analysis Parameters.
- Click **Start and Create Next**. This launches the first assembly analysis on the first Data Set and proceeds to the launch page of the next Data Set, preserving any edits to the parameters made in the previous page.
- Edit the **Genome Length** parameter as appropriate for the second Data Set. Do **not** edit the advanced parameters **Aggressive Option** and **FALCON cfg overrides**, which should still have the values `On` and `pa_dbsplit_option = -x500 -s200` set in the first analysis launch.
- Click **Start and Create Next**. Repeat the process for all Data Sets. After launching the last analysis, you are redirected to the SMRT Analysis module home page, which displays the status of each launched analysis.
- When an analysis completes, click its name to view the results.

Polished Assembly	
Value	Analysis Metric
1	Polished Contigs
4,642,492	Maximum Contig Length
4,642,492	N50 Contig Length
4,642,492	Sum of Contig Lengths
4,642,492	E-size (sum of squares / sum)

Additional Troubleshooting Guidance

If the **Assembly (HGAP 4)** analysis generates a microbial genome assembly with >5 contigs, here is an additional suggestion to help improve the assembly:

- In the **Preassembly** section of the analysis results page, check **Pre-Assembled Coverage (bases/genome_size)**. If the value is below 20, launch a new **Assembly (HGAP 4)** analysis on the same input Data Set. In the **Create New Analysis** page, click **Advanced Analysis Parameters** and change the

Seed coverage parameter from the default of 30 to a level that will approximately compensate. For example, if your preassembled coverage was 10 instead of the desired 20, try changing the seed coverage to 60 to get double the preassembled coverage, since the preassembled yield tends to be similar as seed coverage increases.

Circularizing Microbial Genomes

Circularizing the assembled genome is a step that can be performed after running **Assembly (HGAP 4)** to address issues with linear assemblies of circular genomes, such as removing common overlaps on each end of the contig. Circlator (<http://sanger-pathogens.github.io/circlator/>) is a tool to circularize genome assemblies that has been tested on PacBio data and that may be helpful to run downstream of **Assembly (HGAP 4)** analysis.

Revision History (Description)	Version	Date
Initial Release (Internal only)	01	April 2018
Initial Full Release	02	May 2018

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2018, Pacific Biosciences of California, Inc. All rights reserved. Information in this document is subject to change without notice. Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and to the applicable license terms at <https://www.pacb.com/legal-and-trademarks/terms-and-conditions-of-sale/>. 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, Inc. NGS-go and NGSengine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.

