

Analysis Procedure – Multiplexed Microbial Assembly with SMRT® Link v6.0.0 and SMRTbell® Express Template Prep Kit 2.0

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

- Barcoded Overhang Adapter Kit 8A [PN 101-628-400]: bc1001, bc1002, bc1003, bc1008, bc1009, bc1010, bc1011
- Barcoded Overhang Adapter Kit 8B [PN 101-628-500]: 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_16_Barcodes_v3.zip
2. Import the Sequel_16_Barcodes_v2.zip file in SMRT Link. For information on how to do this refer to “Importing Data” in SMRT Link User Guide v6.0.0.

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_Barcodes_v3**
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**.

| Barcode Name | Bio Sample Name |
|----------------------------------|-----------------|
| bc1001_BAK8A_OA--bc1001_BAK8A_OA | E. coli |
| bc1002_BAK8A_OA--bc1002_BAK8A_OA | S. aureus |

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 number in the Demultiplexed Subsets column (**16** in this example) to view the barcoded Data Sets in a separate table:

| Data Set Name | Demultiplexed Subsets | Bio Sample Name | Barcode Name |
|--------------------------------------|-----------------------|-----------------|--------------|
| My microbial samples | 16 | [multiple] | [multiple] |

| Data Set Name | Bio Sample Name | Barcode Name |
|--|-----------------|----------------------------------|
| My microbial samples (E. coli) | E. coli | bc1001_BAK8A_OA--bc1001_BAK8A_OA |
| My microbial samples (S. aureus) | S. aureus | bc1002_BAK8A_OA--bc1002_BAK8A_OA |

From the Data Management module home page, 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 v6.0.0, 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 number in the Demultiplexed Subsets column (16) to view the child data sets. Click the checkbox in the column header 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: Assembly (HGAP 4) ▼

Analysis Name: My microbial assembly - My microbial samples (E. coli)

Analysis of Multiple Data Sets ← **Members of My microbial samples**

Analysis Type: One Analysis per Data Set - Custom Parameters ▼

Assembly

Genome Length: 5000000

ADVANCED ANALYSIS PARAMETERS

Data Sets Advanced Search Search... ☰

| <input checked="" type="checkbox"/> | Data Set Name | Bio Sample Name |
|-------------------------------------|------------------------------------|-----------------|
| <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 |

5. For the first Data Set, edit the **Genome Length** parameter if desired.
6. 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.
7. Edit the **Genome Length** parameter as appropriate for the second Data Set.
8. 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.
9. When an analysis completes, click its name to view the results.

▶ Analysis Overview

▼ Polished Assembly

Summary Metrics

Contig Coverage vs Confidence

▶ Realignment to Draft Assembly

▶ Coverage

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, below are some additional suggestions to help improve the assembly, which can be tried individually or together.

In the Preassembly section of the analysis results page, check **Pre-Assembled Coverage** bases/genome_size).

1. If the preassembled coverage is below 20:
 - a. Launch a new **Assembly (HGAP 4)** analysis using the same input Data Set.
 - b. 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. This is more frequently found in older sequencing chemistries prior to Sequel Sequencing Kit 3.0 and when library quality may be less than ideal.
 - c. Click **OK** to confirm the Advanced Analysis Parameters changes and run analysis.
2. If the preassembled coverage is above 20:
 - a. Launch a new **Assembly (HGAP 4)** analysis using the same input Data Set. In the **Create New Analysis** page, click **Advanced Analysis Parameters**.
 - b. In the **Assembly** section – **FALCON cfg overrides** type or copy/paste the following:
ovlp_DBsplit_option = -s50

Note that this setting does not always make an impact, but it can be helpful when there is a large difference in coverage across the genome, such as when the DNA was extracted from bacteria in an early log phase.
 - c. Click OK to confirm Advanced Parameters changes and run analysis.

Note: Be sure to use the troubleshooting guidance provided in this document for SMRT Link v6.0.0, as opposed to previous versions which are more relevant to SMRT Link v5.1.0. The guidance provided for earlier software versions are not forward compatible.

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. | 01 | February 2019 |

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2019, 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.