

# Analysis Procedure – Multiplexed Microbial Assembly with SMRT® Link

## Before You Begin

### Multiplexed Microbial Assembly with SMRT Link

Running multiplexed samples for assembly directly in the SMRT Link GUI is not currently supported as an integrated end-to-end workflow. The tools to do so are readily available both in the SMRT Link GUI and on the command line. These instructions will work with SMRT Analysis v4.0 and later.

There are two methods for running SMRT Link analysis applications on a per-barcode basis.

- GUI or manual analysis
- Command Line or automated analysis

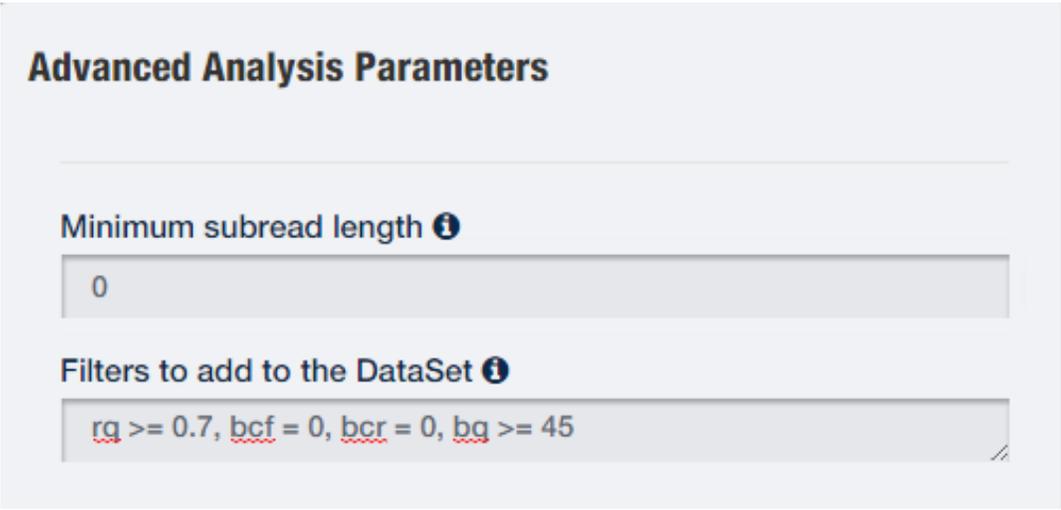
### Barcoded SubreadSet

Both methods listed above require running the standard Barcoding analysis prior to the steps listed here. The output barcoded subreadset from the Barcoding analysis application is the input here.

### Collating Results

A final option to collate reports from multiple SMRT Link jobs (using the same analysis application) is provided at the end for command-line users.

STEP	✓	Manual HGAP per Barcode Using the GUI	Notes
1		<p>Select barcoded data by going to the SMRT Link home page:</p> <p>SMRT Analysis &gt; Create New Analysis &gt; Select your barcoded dataset &gt; Next</p> <p>Note that subreadsets with barcodes have the label "(barcoded)" added to the "Name" field.</p>	
2		<p>Select <b>Assembly (HGAP 4)</b> from the Analysis Application drop-down menu and set the approximate Genome Length.</p>	

3	<p>In Advanced Analysis Parameters &gt; Filters to add to the DataSet, add comma-separated filters to the input dataset (see examples below):</p> <p>Example A (symmetric):  <code>rq &gt;= 0.7, bcf = 0, bcr = 0, bq &gt;= 45</code></p> <p>Example B (asymmetric,universal):  <code>rq &gt;= 0.7, bcf = 0, bcr = 1, bq &gt;= 45</code></p> <ul style="list-style-type: none"> <li>● <b>rq</b>: Read Quality filter. Default setting is usually sufficient.</li> <li>● <b>bcf</b>: Barcode Forward. The 0-indexed position of the forward barcode in the BarcodeSet used for barcoding*</li> <li>● <b>bcr</b>: Barcode Reverse. The 0-indexed position of the reverse barcode in the BarcodeSet used for barcoding</li> <li>● <b>bq</b>: Barcode Quality Score. SW alignment score of the barcode on integer scale of 0-100. Recommended minimum 45.</li> </ul> 	
4	Change any other parameters as desired. Click <b>Start</b> .	
5	Manually Repeat steps 1-4 for each barcode in the dataset.	

## Indexing Barcodes

For the two example filter parameters above and the following snippet of the first four records in the barcode fasta, take a look at the fasta file underlying the BarcodeSet that was used for barcoding in order to identify indices with the barcode names.

```
>lbc1
TCAGACGATGCGTCAT
>lbc2
CTATACATGACTCTGC
>lbc3
TACTAGAGTAGCACTC
>lbc4
TGTGTATCAGTACATG
```

The examples correspond to the following barcodes:

Example A (symmetric): **lbc1**(index 0 on both sides of the insert)

Example B (asymmetric,universal):**lbc1--lbc2**(index 0 with the forward primer, index 1 with the reverse primer)

## Automated HGAP Job Submission per Barcode using Python

Automating barcoded analysis submission can be done using the tools and Python APIs in the SMRT Link installation. The scripts needed for this tutorial can be found below. The scripts use the [pbcommand](#) library and the [pbservice](#) tool to interact with SMRT Link services. You will need to know the location of your SMRT Link install to modify the scripts for your system.

- [Bash quickstart \(multiplexHGAP4.sh\)](#)
- [Splitting and Importing \(splitBarcodeUpload.py\)](#)
- [Submitting Jobs \(multiplexSubmit.py\)](#)
- [HGAP presets json \(presets\\_template.json\)](#)

STEP	✓	Quickstart to Automation	Notes
1		<p>Modify <b>multiplexSubmit.py</b> to correspond to where you saved the presets_template.json, your SMRT Analysis install directory, and the host and port (See 'About' link in the SMRT Link browser Menu).</p> <pre>#Set the following values at the top of the script PRESETS_TEMPLATE='/path/to/presets_template.json' PBSERVICE      ='/path/to/smrtlink/installdir/smrtcmds/bin/pbservice' DEFAULTHOST    ='smrtlink' DEFAULTPORT    =8081</pre>	
2		<p>Modify <b>multiplex.sh</b> to point to the SMRT Link python interpreter, and the two scripts you downloaded above.</p> <pre>#Set the following values in the script PYTHON=/path/to/smrtlink/installdir/private/otherbins/all/bin/python SPLITPROG=/path/to/splitBarcodeUpload.py JOBPROG=/path/to/multiplexSubmit.py</pre> <p>You might also consider setting the host and port values rather than passing them in as arguments to the script.</p>	

<p><b>3</b></p>	<p>Run the script using a barcoded subreadset xml as input:</p> <pre> NAME='My Multiplexed Jobs' BARCODESUBREADSET=/path/to/barcoded.subreadset.xml #Can use xml downloaded from SMRT Link BARCODEFASTA=/path/to/barcode.fasta #Barcode fasta used to score subreadset. Used for naming only. GENOMESIZE=4600000 #Estimated OUTDIR=/my/output/directory HOST=smrtlink #See SMRT Link 'About' page in browser PORT=8081  bash multiplexHGAP.sh \$NAME \$BARCODESUBREADSET \$BARCODEFASTA \$GENOMESIZE \$OUTDIR \$HOST \$PORT </pre> <p>The bash script runs the two python scripts. First run <b>splitBarcodeUpload</b> to generate subreadset.xml files, one for each barcode and import them to the SMRT Link server. Second run ``multiplexSubmit.py`` to start HGAP 4 jobs, one for each imported barcode subreadset using default parameters and specified estimated genomesize.</p> <p>After the script finishes, the output directory contains the following:</p> <pre> ls /my/output/directory  [NAME]_lbc*--lbc*.subreadset.xml #Filtering xmls, one for each barcode [NAME]_barcode lbc*--lbc*_HGAP_presets.json #HGAP presets, one for each barcode uploaded_subreadsets.csv #Table of imported datasets. (subsetName,subsetId,jobState) started_jobs.csv #Table of started jobs. (host,jobId,jobName,jobPath) </pre> <p>Jobs will usually still be running after the submit script finishes, and they will all run concurrently by default. If you are concerned for overloading your cluster environment with too many jobs, you can set the <b>SLEEP</b> parameter in multiplexHGAP.sh to wait N minutes between submitting jobs.</p>	
-----------------	--	--

## Splitting and Importing each barcode

The script **splitBarcodeUpload.py** can be used independently to upload barcoded subsets to you SMRT Link server for further analysis in any downstream SMRT Link analysis application. See `python splitBarcodeUpload.py --help` for more information.

Note that this script does *not* physically demultiplex the original subreads.bam. Each imported subset is an xml file pointing to the original barcoded subreadset with an associated set of filters for barcode value and quality.

See [documentation](#) for more information on the dataset model.

## Multiple Job Submission

The script `multiplexSubmit.py` can also be used independently to submit multiple SMRT Link jobs given a `presets.json` template and a csv file of input subreadsets IDs from your server. Just change the template in the script to use this with other SMRT Link analysis applications. The script will pick up the available options from the json file.

```
PRESETS_TEMPLATE='/path/to/presets_template.json'
```

See `multiplexSubmit.py --help` after setting the preset template.

The json file linked to on this wiki is for HGAP 4 and can be generated using `pbsmrtpipe` from the SMRT Analysis environment. Generating a json file for resequencing can be done as follows:

```
pbsmrtpipe show-template-details pbsmrtpipe.pipelines.sa3_ds_resequencing -j  
presets.json
```

## Comparing Results

The following script is provided as a way to collate results reports for multiple SMRT Link jobs. Note that this tool depends on the python module [pandas](#) (not included in the SMRT Analysis python) as well as [pbcommand](#).

- [collateReports.py](#)

This takes a csv file with the minimum columns `host`, `obId`, and `jobName` and attempts to collate reports from the listed jobs for export to csv and excel formats.

```
head started_jobs.csv
```

```
host,jobId,jobName,jobPath  
smrtlink,1934,12plex_bsub (barcode=lbc34--lbc34)_HGAP,/path/to/jobs/001934  
smrtlink,1935,12plex_bsub (barcode=lbc40--lbc40)_HGAP,/path/to/jobs/001935  
smrtlink,1936,12plex_bsub (barcode=lbc29--lbc29)_HGAP,/path/to/jobs/001936  
smrtlink,1937,12plex_bsub (barcode=lbc38--lbc38)_HGAP,/path/to/jobs/001937  
smrtlink,1938,12plex_bsub (barcode=lbc70--lbc70)_HGAP,/path/to/jobs/001938
```

See `collateReports.py --help` for more information.