

# Local Run Manager 16S Metagenomics Analysis Module

## Workflow Guide

Overview	3
Set Parameters	3
Analysis Methods	5
View Analysis Results	6
Analysis Report	7
Analysis Output Files	8
Custom Analysis Settings	8
Technical Assistance	10



This document and its contents are proprietary to Illumina, Inc. and its affiliates ("Illumina"), and are intended solely for the contractual use of its customer in connection with the use of the product(s) described herein and for no other purpose. This document and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina. Illumina does not convey any license under its patent, trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

The instructions in this document must be strictly and explicitly followed by qualified and properly trained personnel in order to ensure the proper and safe use of the product(s) described herein. All of the contents of this document must be fully read and understood prior to using such product(s).

FAILURE TO COMPLETELY READ AND EXPLICITLY FOLLOW ALL OF THE INSTRUCTIONS CONTAINED HEREIN MAY RESULT IN DAMAGE TO THE PRODUCT(S), INJURY TO PERSONS, INCLUDING TO USERS OR OTHERS, AND DAMAGE TO OTHER PROPERTY, AND WILL VOID ANY WARRANTY APPLICABLE TO THE PRODUCT(S).

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2018 Illumina, Inc. All rights reserved.

All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

## Overview

The Local Run Manager 16S Metagenomics analysis module classifies organisms from a metagenomic sample by amplifying specific regions in the 16S ribosomal RNA. This workflow is exclusive to Prokaryotes, which includes Bacteria and Archaea. The 16S Metagenomics workflow generates a classification of reads at several taxonomic levels: kingdom, phylum, class, order, family, and genus or species.

The 16S Metagenomics workflow demultiplexes indexed reads, generates FASTQ files, and then classifies reads.

## Compatible Library Types

The 16S Metagenomics analysis module is compatible with specific library types represented by library kit categories on the Create Run screen. For a current list of compatible library kits, see the [Local Run Manager support page](#) on the Illumina website.

## About This Guide

This guide provides instructions for setting up run parameters for sequencing and analysis parameters for the 16S Metagenomics analysis module. For information about the Local Run Manager dashboard and system settings, see the *Local Run Manager Software Guide (document # 1000000002702)*.

## Set Parameters

- 1 If needed, log in to Local Run Manager.
- 2 Select **Create Run**, and select **16S Metagenomics**.
- 3 Enter a run name that identifies the run from sequencing through analysis.  
The run name can contain alphanumeric characters, spaces, and the following special characters:  
`~!@#\$%\_-\_{}`.
- 4 **[Optional]** Enter a run description to identify the run.  
The run description can contain alphanumeric characters, spaces, and the following special characters:  
`~!@#\$%\_-\_{}`.

## Specify Run Settings

- 1 Select the library prep kit from the Library Prep Kit drop-down list.
- 2 Specify the number of index reads.
  - ▶ 0 for a run with no indexing
  - ▶ 1 for a single-indexed run
  - ▶ 2 for a dual-indexed run
 Unsupported index reads for your library prep kit are automatically disabled.
- 3 Specify a read type: **Single Read** or **Paired End**.  
If your library prep kit supports only one option, the read type is automatically selected.
- 4 Enter the number of cycles for the run.
- 5 **[Optional]** If using custom primers, specify their information.  
Custom primer options may vary based on your instrument or Local Run Manager implementation.

## Specify Module-Specific Settings

- 1 If applicable, set the Reverse Complement option.  
Enabling this setting makes all reads reverse-complemented as they are written to FASTQ files. By default, the reverse complement setting is turned off.
- 2 Select the taxonomy database to use for the run.
  - ▶ **Species-level Classification**—Use an Illumina-developed species-level classification taxonomy database.
  - ▶ **Custom Taxonomy Database**—Use a custom taxonomy database for taxonomic classification.
- 3 If you selected **Custom Taxonomy Database** as the taxonomy database, enter the full file path for the custom taxonomy database file.

## Specify Samples for the Run



Specify samples for the run using the following options:

- ▶ **Enter samples manually**—Use the blank table at the bottom of the Create Run screen.
- ▶ **Import sample sheet**—Navigate to an external file in a comma-separated values (\*.csv) format.

After you have populated the samples table, you can export the sample information to an external file. You can use this file as a reference when preparing libraries or import the file when configuring another run.

## Enter Samples Manually

To manually enter sample information, you must first select a Library Prep Kit in the Run Settings section.

- 1 Adjust the samples table to an appropriate number of rows.
  - ▶ In the Rows field, use the up/down arrows or enter a number to specify the number of rows to add to the table. Select  to add the rows to the table.
  - ▶ Select  to delete a row.
  - ▶ Right-click on a row in the table and use the commands in the contextual menu.
- 2 Enter a unique sample ID in the Sample ID field.  
Only use alphanumeric characters, dashes, or underscores. Do not include spaces.
- 3 Enter a sample name.  
This field is required when a sample sheet with a Sample\_Name column is imported.
- 4 **[Optional]** Enter a sample description in the Sample Description field.  
Only use alphanumeric characters, dashes, or underscores. Do not include spaces.
- 5 If you have a plated kit, select an index plate well from the Index well drop-down list and skip to step 8.
- 6 If applicable, specify an Index 1 sequence.  
Select **Show Index Sequence/Show Index Names** to toggle between showing the name of the index and the index sequence.
- 7 If applicable, specify an Index 2 sequence.



### NOTE

During analysis, the iSeq™ 100, MiniSeq™, and NextSeq™ Systems automatically reverse complement the i5 indexes in custom library prep kits. Make sure that the i5 indexes are in the forward orientation.

- 8 **[Optional]** Enter a sample project name in the Sample Project field.

Use alphanumeric characters, dashes, underscores, or spaces.

- 9 **[Optional]** Select **Export Sample Sheet** to export sample information in \*.csv format.
- 10 Select **Save Run**.

## Import Sample Sheet

- 1 If you do not have a sample sheet to import, see *Enter Samples Manually on page 4* for instructions on how to create and export a sample sheet. Edit the file as follows.
  - a Open the sample sheet in a text editor.
  - b Enter the sample information in the [Data] section of the file.
  - c Save the file. Make sure that the sample IDs are unique.
- 2 Select **Import Sample Sheet** at the top of the Create Run screen and browse to the location of the sample sheet.  
Make sure that the information in the sample sheet is correct. Incorrect information can impact the sequencing run.
- 3 When finished, select **Save Run**.

## Analysis Methods

The 16S Metagenomics analysis module performs the following analysis steps and then writes analysis output files to the folder.

- ▶ Demultiplexes index reads
- ▶ Generates FASTQ files

## Demultiplexing

Demultiplexing compares each Index Read sequence to the index sequences specified for the run. No quality values are considered in this step.

Index reads are identified using the following steps:

- ▶ Samples are numbered starting from 1 based on the order they are listed for the run.
- ▶ Sample number 0 is reserved for clusters that were not assigned to a sample.

## FASTQ File Generation

After demultiplexing, the software generates intermediate analysis files in the FASTQ format, which is a text format used to represent sequences. FASTQ files contain reads for each sample and the associated quality scores. Any controls used for the run and clusters that did not pass filter are excluded.

Each FASTQ file contains reads for only one sample, and the name of that sample is included in the FASTQ file name. FASTQ files are the primary input for alignment.

## Classification of Reads

The classification step uses ClassifyReads, a proprietary algorithm that provides species-level classification for paired-end reads. This process involves matching short subsequences of the reads (called words) to a set of 16S reference sequences. The accumulated word matches for each read are used to assign reads to a particular taxonomic classification. Analysis results list the total number of classified clusters for each sample at each taxonomic level. Statistics are written to the file Classification.txt.

## Current Taxonomy

The current taxonomy is stored in Taxonomy.dat. As of Local Run Manager v2.0, the 16S Metagenomics workflow generates classifications to the species level. For information about setting up analysis to genus level only, see [Sample Sheet Settings for Analysis on page 1](#).

The taxonomy database for the 16S Metagenomics workflow is an Illumina-curated version of the Greengenes database ([greengenes.secondgenome.com/downloads/database/13\\_5](http://greengenes.secondgenome.com/downloads/database/13_5)). To generate species-level classifications, the following filters are applied:

- ▶ Filter all entries where the 16S sequence length was below 1250 bp.
- ▶ Filter all entries with more than 50 wobble bases (M, R, W, S, Y, K, V, H, D, B, and N).
- ▶ Filter all entries that are partially classified with no classification for genus or species.

The following taxonomic counts are available for the 16S Metagenomics workflow.

Taxonomy	Count
Kingdoms	3
Phyla	33
Classes	74
Orders	148
Families	321
Genera	1086
Species	6466

## Alternative Taxonomy Database

You can prepare an alternative taxonomy database using the tool CreateTaxonomyDatabase distributed with Local Run Manager. This tool is located in the Local Run Manager install folder, typically on the C: drive:

C:\Illumina\Local Run

Manager\Modules\MetagenomicsWorkflow\\Workflows\MetagenomicsWorker\CreateTaxonomyDatabase.exe.

CreateTaxonomyDatabase is a command-line tool; run it without arguments for a description of available options. For an example of a valid FASTA file, see:


[greengenes.lbl.gov/Download/Sequence\\_Data/Fasta\\_data\\_files/current\\_GREENGENES\\_gg16S\\_unaligned.fasta.gz](http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/current_GREENGENES_gg16S_unaligned.fasta.gz)

The 16S Metagenomics workflow provides species-level classification. To configure the workflow for genus-level classification, use the TaxonomyFile sample sheet setting and specify gg\_13\_5\_genus\_32bp.dat. For more information, see [Sample Sheet Settings for Analysis on page 1](#).



## Statistics Reporting

Statistics are summarized and reported, and written to the Alignment folder.

## View Analysis Results

- 1 From the Local Run Manager dashboard, select the run name.
- 2 From the Run Overview tab, review the sequencing run metrics.
- 3 To change the analysis data file location for future requeues of the selected run, select the **Edit**  icon, and edit the output run folder file path.

The file path leading up to the output run folder is editable. The output run folder name cannot be changed.

- 4 **[Optional]** Select the **Copy to Clipboard**  icon to copy the output run folder file path.
- 5 Select the Sequencing Information tab to review run parameters and consumables information.
- 6 Select the Samples & Results tab to view the analysis report.
  - ▶ If analysis was requested, select the appropriate analysis from the Select Analysis drop-down list.
- 7 **[Optional]** Select the **Copy to Clipboard**  icon to copy the Analysis Folder file path.

## Analysis Report

Analysis results are summarized on the Samples & Results tab. The report is also available in a PDF file format for each sample in the Analysis folder.

## Sample Configuration

Row	Description
Sample ID	The sample ID from the sample sheet. Sample ID must always be a unique value.
Sample Name	The sample name from the sample sheet.
Run Folder	The folder location of the run.
Taxonomy File	Name of the taxonomy file in *.dat format.

## Sample Information

Column	Description
Total Reads	The number of reads in a run.
Reads Passing Quality Filtering	The number of reads passing filter.
% Reads Passing Quality Filtering	The percentage of reads passing filter.

## Classification Results

Report	Description
Classification Statistics	A table with the number of reads and percentage of reads classified to each classification level: Kingdom, Phylum, Class, Order, Family, Genus, Species.
Classification Rate by Taxonomic Level	A bar graph showing the percentage of reads classified to each classification level.
Sunburst Classification Chart	A sunburst chart showing the relative abundance of the classification results within each taxonomic level. Click on any category to zoom in, click the center button to zoom out, and click <b>Reset Plot</b> to return to a Kingdom-level view.
Top 20 Classification Results by Taxonomic Level	A bar chart showing the relative abundance of the top 20 classification results within each taxonomic level. Mouse over any category to see its description and abundance.
Classification Results by Taxonomic Level	Tables for each taxonomic level showing the number and percentage of reads for the highest 8 taxonomic classifications. Pie charts for each taxonomic level show all classifications above 3.5% abundance. The "Other" category in the pie charts is the sum of all classifications with less than 3.50 % abundance.

## Analysis Output Files

The following analysis output files are generated for the 16S Metagenomics analysis module and provide analysis results for alignment. Analysis output files are located in the Alignment folder.

File Name	Description
Classification.txt	Contains the total number of classified clusters for each sample at each taxonomic level.
Demultiplexing (*.demux)	Intermediate files containing demultiplexing results.
DemuxSummary*	Intermediate files containing demultiplexing summary results.
FASTQ (*.fastq.gz)	Intermediate files containing quality scored base calls. FASTQ files are the primary input for the alignment step.
*.rarefaction.txt	Contains the number of unique genera discovered by the number of reads classified.

## Supplementary Output Files

The following output files provide supplementary information, or summarize run results and analysis errors. Although these files are not required for assessing analysis results, they can be used for troubleshooting purposes.

File Name	Description
AnalysisLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages.
AnalysisError.txt	Processing log that lists any errors that occurred during analysis. This file will be empty if no errors occurred.
Summary.txt	Contains summary statistics specific to the run.

## Custom Analysis Settings

Custom analysis settings are intended for technically advanced users. If settings are applied incorrectly, serious problems can occur.

### Add a Custom Analysis Setting

- 1 From the Module-Specific Settings section of the Create Run screen, select .
- 2 Select **+ Add custom setting**.
- 3 In the custom setting field, enter the setting name as listed in the Available Analysis Settings section.
- 4 In the setting value field, enter the setting value.
- 5 To remove a setting, select **✕**.



## Sample Sheet Settings for Analysis

Parameter	Description
Adapter	<p>Only specify for custom library prep kits.</p> <p>Specify the 5' portion of the adapter sequence to prevent reporting sequence beyond the sample DNA.</p> <p>Illumina recommends adapter trimming for Nextera libraries and Nextera Mate Pair libraries.</p> <p>To specify 2 or more adapter sequences, separate the sequences by a plus (+) sign. For example: CTGTCTCTTATACACATCT+AGATGTGTATAAGAGACAG</p>
AdapterRead2	<p>Only specify for custom library prep kits.</p> <p>Specify the 5' portion of the Read 2 adapter sequence to prevent reporting sequence beyond the sample DNA.</p> <p>Use this setting to specify a different adapter other than the one specified in the <b>Adapter</b> setting.</p>
TaxonomyFile	<p>This setting overrides the taxonomy database; default is taxonomy.dat.</p> <p>Species-level classification is enabled, by default. For faster, but less granular genus-level classification, specify gg_13_5_genus_32bp.dat.</p>

## Technical Assistance

For technical assistance, contact Illumina Technical Support.

Website: [www.illumina.com](http://www.illumina.com)  
 Email: [techsupport@illumina.com](mailto:techsupport@illumina.com)

### Illumina Customer Support Telephone Numbers

Region	Toll Free	Regional
North America	+1.800.809.4566	
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Belgium	+32 80077160	+32 34002973
China	400.066.5835	
Denmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Germany	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Ireland	+353 1800936608	+353 016950506
Italy	+39 800985513	+39 236003759
Japan	0800.111.5011	
Netherlands	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Norway	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan	00806651752	
United Kingdom	+44 8000126019	+44 2073057197
Other countries	+44.1799.534000	

Safety data sheets (SDSs)—Available on the Illumina website at [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

Product documentation—Available for download in PDF from the Illumina website. Go to [support.illumina.com](http://support.illumina.com), select a product, then select **Documentation & Literature**.



Illumina  
5200 Illumina Way  
San Diego, California 92122 U.S.A.  
+1.800.809.ILMN (4566)  
+1.858.202.4566 (outside North America)  
techsupport@illumina.com  
www.illumina.com

**For Research Use Only. Not for use in diagnostic procedures.**

© 2018 Illumina, Inc. All rights reserved.

**illumina**<sup>®</sup>