

Obtaining FASTQ files off BaseSpace

FASTQ files store sequence and quality information for every read in a sample. These files can be very large but contain only plain text and can be opened in notepad, word, and many other programs.

Our water microbiome FASTQ files are stored as a project on the Illumina cloud service BaseSpace. You will need to create a free BaseSpace account to download these samples, and the process of acquiring them is described below.

Access the project

The first step to access these FASTQ files is to follow the link to the project and create a BaseSpace account. The links should be available at our water microbiome module page, but are:

Example Data Set (used in walkthrough):

<https://basespace.illumina.com/s/w2ZqhnVBEioR>

Additional Sample Data (contains additional samples as well):

<https://basespace.illumina.com/s/NaarQP0XZUjZ>

When you follow the link, you will be brought to the BaseSpace sign in page (Figure 1).

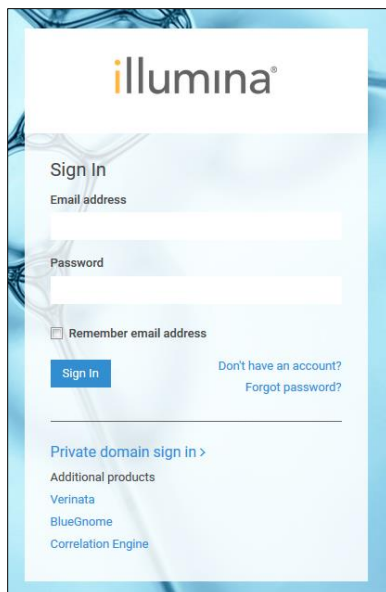


Figure 1 BaseSpace sign in page

Sign in, if you have an account, or use the “Don’t have an account?” link to sign up. Creating an account requires a valid email, but is free.

Once signed in, you will be brought to BaseSpace’s “Dashboard” page and will be prompted to accept a shared project. Accept the share and use the top menu to navigate to the “Projects” page (Figure 2).

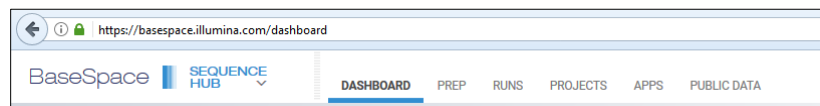


Figure 2 BaseSpace's top navigation menu, as seen from the Dashboard

Click on the project name to enter that project.

Accessing FASTQ files in a Project

Once in the project, use the side navigation bar to go to the sample list (Figure 3)

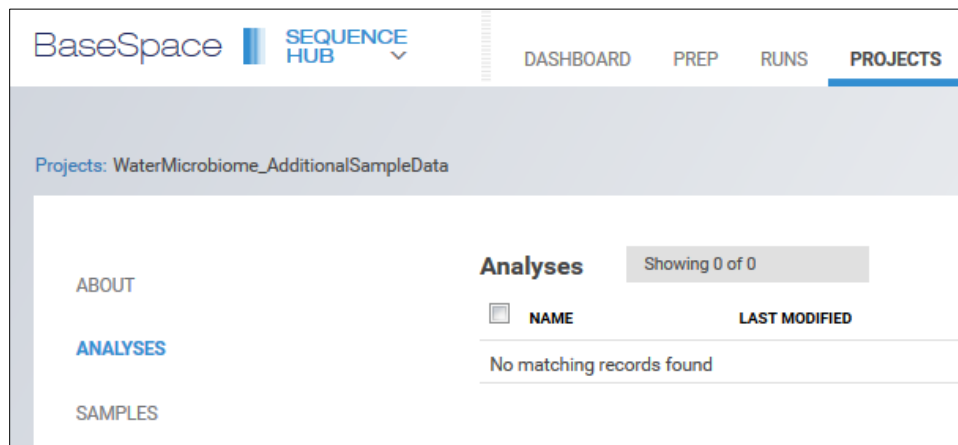


Figure 3 Side navigation bar, as seen from the Analysis tab.

To download samples you can select multiple samples and use the “Download Samples” button (Figure 4, this requires downloading/installing the BaseSpace downloader).

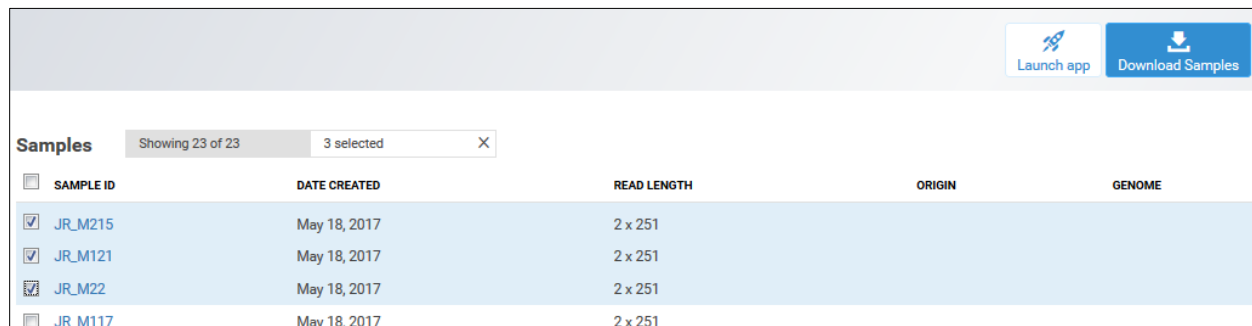


Figure 4 Selecting multiple samples to download at once.

You can also navigate to samples individually (clicking on their name) and using the sample entry to navigate to the individual FASTQ files (Figures 5 and 6).

Once you have the FASTQ files you need, follow the process outlined in the QIIME walkthrough to analyze them.

Details

Sample name	JR_M215
Date created	an hour ago
Genome	-
Paired end	Paired
Number of Clusters	182,629
Number of Reads	365,258
Read 1 length	251
Read 2 length	251
Size	61.08 MB

Origin

Run information is not available for this sample.

Files

NAME	SIZE	TYPE	STATUS	DATE
JR-M215_S11_L001_R2_001.fastq.gz	31.86 MB	.gz	UploadComplete	an hour ago
JR-M215_S11_L001_R1_001.fastq.gz	29.22 MB	.gz	UploadComplete	an hour ago

Figure 5 Sample entry. Click on the FASTQ file's name to download without using the BaseSpace downloader (Figure 6)

JR_M215 : Files

JR-M215_S11_L001_R2_001.fastq.gz

File is either not text-based or too large for viewing, please [download the raw file](#).

Opening JR-M215_S11_L001_R2_001.fastq.gz

You have chosen to open:

- JR-M215_S11_L001_R2_001.fastq.gz
which is: gzip (31.9 MB)
from: <https://basespace-data-east.s3-external-1.amazonaws.com>

What should Firefox do with this file?

Open with

Save File

Do this automatically for files like this from now on.

Figure 6 File page and the prompt to download after clicking "download the raw file".

File Names and sample Lists

File names are in the form: JR-M82_S7_L001_R1_001

JR: initials of the researcher to collect and process samples

M: water source (M for Mystic River or T for Tap water)

82: date of sample collection in MM/DD (leading zeros excluded)

S7: sample ID used by the MiSeq

L001: lane number

R1: read number (each sample should have R1 and R2 FASTQ files)

001: tile number

We generally refer to samples by their source and collection date, ex: M82 or T111.

The “Example Data Set” project contains ten samples from a single sequencing run and is a subset of the “Additional Sample Data” set. The “Additional Sample Data” project contains 23 samples from two sequencing runs.

Samples from the first run (*are part of the Example Data Set used in the QIIME walkthrough)

M82*

M91*

M101*

M111*

M317*

M815

M915

M1015

M1116

T82*

T111*

T117*

T815*

T1015*

T1116

Samples from the second run

M22

M117

M121

M215

T22

T31

T215

T317