

Download datasets from the HMP portal

Alise Ponsero

Abstract

This protocol explores how to download datasets from the HMP portal, either from a sample id number.

Citation: Alise Ponsero Download datasets from the HMP portal. **protocols.io**

dx.doi.org/10.17504/protocols.io.srzed76

Published: 21 Aug 2018

Protocol

Download a 16s RNA dataset from the HMP portal using a sample number

Step 1.

Note : this protocol uses the sample [3674d95cd0d27e1de94ddf4d2e0c0780](https://portal.hmpdacc.org/cases/3674d95cd0d27e1de94ddf4d2e0c0780) from the 16s-PP1 project from the HMP

In order to access the page containing information about the sample, search in your browser https://portal.hmpdacc.org/cases/YOUR_SAMPLE_NUMBER (eg. <https://portal.hmpdacc.org/cases/3674d95cd0d27e1de94ddf4d2e0c0780>). This page displays basic information about the sample, as the body site concerned, the subject gender and identifier for the study.

In the top right corner of the screen click on 'Add all files to the cart'. And go to your cart. Two types of files will be added in your cart :

- raw dataset (raw_16s) in SFF format
- a trimmed dataset (trimmed_16s) in Fasta format

Clicking on the file name will give you additional information about the file. This page also displays a 'download button', but as of today (08-21-18), this doesn't work. In order to download the file, copy and paste the complete file name in your browser (eg <http://downloads.hmpdacc.org/data/HM16STR/SRP002860/SRS147223.fsa>).

Notes : Several sequencing run may have been realized for a same given sample. For your project, you will need to download only one sequencing run, as a rule you can choose to download the bigger run in file size.

Be also mindful that for some samples, a WGS sequencing has been realized on this sample. Look carefully at the data type before downloading.

Download a WGS dataset from the HMP portal using a sample number

Step 2.

Note : this protocol uses the sample [3674d95cd0d27e1de94ddf4d2ec63794](https://portal.hmpdacc.org/cases/3674d95cd0d27e1de94ddf4d2ec63794) from the WGS-PP1 project from the HMP

In order to access the page containing information about the sample, search in your browser https://portal.hmpdacc.org/cases/YOUR_SAMPLE_NUMBER (eg. <https://portal.hmpdacc.org/cases/3674d95cd0d27e1de94ddf4d2ec63794>). This page displays basic information about the sample, as the body site concerned, the subject gender and identifier for the study.

In the top right corner of the screen click on 'Add all files to the cart'. And go to your cart. Various types of files will be added in your cart :

- raw dataset (wgs_raw) in FASTQ format
(if several raw datasets are available, select the file with a name looking like <http://downloads.hmpdacc.org/data2/hhs/genome/microbiome/wgs/analysis/hmwgsqc/v2/SRS013951.tar.bz2>)
- assembled datasets (wgs_assembled) in FASTA format
(if several assembled datasets are available, select the file with a name looking like <http://downloads.hmpdacc.org/data/HMASM/IDBA/SRS013951.fna.bz2>)

Clicking on the file name will give you additional information about the file. This page also displays a 'download button', but as of today (08-21-18), this doesn't work. In order to download the file, copy and paste the complete file name in your browser (eg <http://downloads.hmpdacc.org/data/HMASM/IDBA/SRS013951.fna.bz2>).

Notes : Several sequencing run may have been realized for a same given sample. For your project, you will need to download only one sequencing run.

Be also mindful that for some samples, a 16s RNA sequencing has been realized on this sample. Look carefully at the data type before downloading.