

Bioinformatics Lab Guide 1:

Navigating and Using GalaxyTrakr (and Galaxy)

James B. Herrick, Ph.D.
James Madison University

Navigation

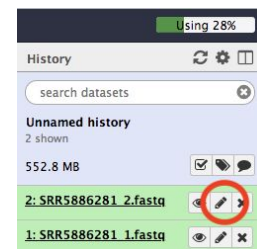
1. If you haven't already, go to <https://account.galaxytrakr.org/Account/Register> and sign up for a GalaxyTrakr account. (Note: this registration is separate from Galaxy Registration)
2. Go to galaxytrakr.org and log on to your account.
3. If it's not already open, click on 'Analyze Data' at the top of the window.

Uploading Data

1. In the left pane ("Tools"), click on "Get Data" and then "FASTQ Dump paired downloader".
2. In the window that opens, copy and paste the following SRR number: **SRR5886281** and hit 'Execute'. This downloads two ('forward' and 'reverse') paired end files from the [NCBI Short Read Archive \(SRA\)](https://www.ncbi.nlm.nih.gov/sra/), which stores raw high-throughput sequencing data. SRR5886281 is the SRA number for HJ1, the first *Salmonella* isolated in our lab (from Muddy Creek on 2016-10-22).
3. In the History pane on the right side, you will see the names of the reads. The color will change from gray (preparing to upload) to yellow (uploading) to green (uploading complete).
4. Re-name each file by adding the strain name HJ1_ at the beginning. This helps us keep the file straight by proving a complete lab name.
 - a. Click on the pencil icon to the right of one name:
 - b. Type "HJ1_" before the SRR number in the "name" field and click 'Save'.

Note: "HJ1" is our lab strain name, using the initials of those who isolated the strain (in this case Charles Holmes and Sophie Jurgensen) plus a numeric designation. Always insert these in the file name, but please be careful that they match up properly.

- c. Repeat for the other file.



Fastq files

1. Click on the eyeball icon to the left of the pencil icon to see what's in each file. These are “fastq” files, which include both the sequence itself and a quality score for each base called, using an ascii code (see below).

“Fastq” files (from Wikipedia)

A FASTQ file normally uses four lines per sequence.

- Line 1 begins with a '@' character and is followed by a sequence identifier and an *optional* description (like a [FASTA](#) title line).
- Line 2 is the raw sequence letters.
- Line 3 begins with a '+' character and is *optionally* followed by the same sequence identifier (and any description) again.
- Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

A FASTQ file containing a single sequence might look like this:

```
@SEQ_ID
GATTGGGGTTCAAAGCAGTATCGATCAATAGTAAATCCATTGTTCAACTCACAGTTT
+
!"*((( (**+))%%%++))(%%%%).1***-+*"))**55CCF>>>>>CCCCCCC65
```

The character '!' represents the lowest quality while '~' is the highest. Here are the quality value characters in left-to-right increasing order of quality ([ASCII](#)):

```
!"#$%&'()*+,-./0123456789;:<=>?@ABCDEFGHIJKLMNPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
```

Take a quick tour of the user interface

In the middle of the page, click on the button that says “Take an interactive tour: [Galaxy UI](#).” Run through this quick orientation to GalaxyTrakr. **Note:** when you come to “Select and load a tool for your analysis by clicking the underlined link,” click on one of the tools (like “Trimmomatic”) to see examples of the different sections they’re talking about. You can then run the analysis on strain HJ1, as an example.

Take a quick tour of the History pane

Click on 'Analyze Data' again (at the top of the page) and then click on the "Take an interactive tour: [History](#) button to run through a quick orientation to the History pane. Explore all the buttons on the history pane just to get used to using it, as this is where your work will be stored, viewed, downloaded, etc.

Sharing Histories

1. To share your history, click on the gear icon at the top of the history pane and choose "share and publish".
2. You won't need to share your HJ1 history, but once you begin to analyze and assemble one of our new (DG) genomes, you'll need to share your history with me. To do so, click on "share with a user" under "Share History with Individual Users" and type my email address "herricjb@jmu.edu" in the field.

Other Implementations of Galaxy

Galaxy main: <https://usegalaxy.org/>

Galaxy @ Pasteur Institute: <https://galaxy.pasteur.fr/>

Galaxy Tutorials and Help

- GalaxyTrakr User Guide: <https://goo.gl/PXUct5>
- [Galaxy 101](#) - the basic introduction to Galaxy's interface, its functionality, and workflows. Start here if you never used Galaxy before.
- [Uploading data](#) - how to get data into Galaxy.
- [Histories](#) - in Galaxy uploaded data and analysis results reside within the history pane. This tutorial explains how history works.
- [Where to go to get help](#):
 - GalaxyTrakrSupport@fda.hhs.gov
 - [Galaxy Support Page](#) -- "Chat" is useful for immediate answers.
 - [BioStars](#) is a very useful bioinformatics help site. The Biostars forum *specifically dedicated to Galaxy* is at: <https://biostar.usegalaxy.org/>