Running Influenza A, H5Nx Metagenomic Samples in Terra using Theiagen's Freyja FASTQ Workflow	
Document TG-FREY-H5NX, Version 1	
Date:	Workflow Versions:
07/08/2025	PHB v3

1. PURPOSE/SCOPE

To standardize the process of running Influenza A, H5Nx (H5Nx) metagenomic samples using Theiagen's Freyja FASTQ workflow in Terra to perform lineage deconvolution, abundance determination, and identify coverage metrics. This SOP is specific to Illumina paired end (PE) raw read files. While this SOP can be used for any subtype of H5 influenza, please note that this SOP should NOT be used to run Influenza A, H3N2 samples.

2. REQUIRED RESOURCES

- Computer
- Internet connection: at least 10 and 5Mbps for download and upload speeds, respectively
- Internet browser
  - Google Chrome, Firefox, or Edge
- Google account
- Terra account, linked to Google account
- Metagenomic Illumina PE raw read files uploaded to Terra workspace
- Theiagen's Freyja_FASTQ_PHB Workflow in Terra
  - See the link "Running Freyja on other pathogens" for more details and resources.
  - See Appendix 10.1

REQUIRED WORKFLOW INPUTS FILES

- Raw Illumina PE read files
- [Primer bed file]
- Reference genome
- [barcodes metadata file]*

3. RELATED DOCUMENTS

Document Number	Document Name
TG-TER-03	Uploading Local or SRA NGS Data & Creating a Results Metadata Table in Terra

4. PROCEDURE

4.1 RUNNING THE FREYJA FASTQ WORKFLOW

1. Open Terra and navigate to the *workflows* tab of the workspace containing wastewater data
2. Select the *Freyja_FASTQ_PHB* workflow (Fig 1)

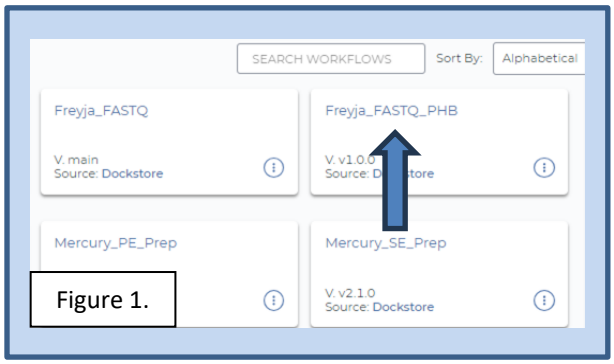


Figure 1.

3. *Uncheck* *use call caching* (Fig 2)
4. Choose the latest version of *version 3* in the version dropdown field, or the workflow version that was used during internal assay validation (Fig 2, a)
5. Select the second bullet to *run workflow(s) with inputs defined by data table* (Fig 2, b)
6. Select the relevant data table name under the select *root entity type* dropdown (Fig 2, c)
7. Click *select data* (Fig 2, d) and in the pop-up window *select the checkbox* for each sample to be included in the analysis (Fig 3)

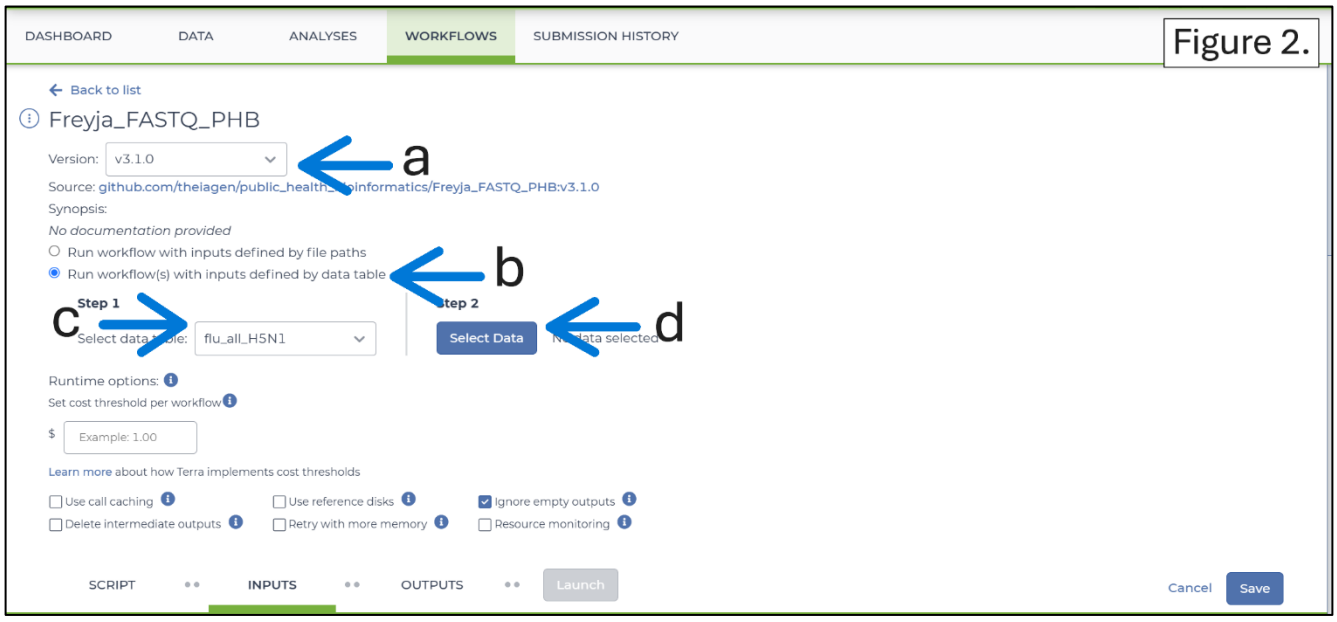
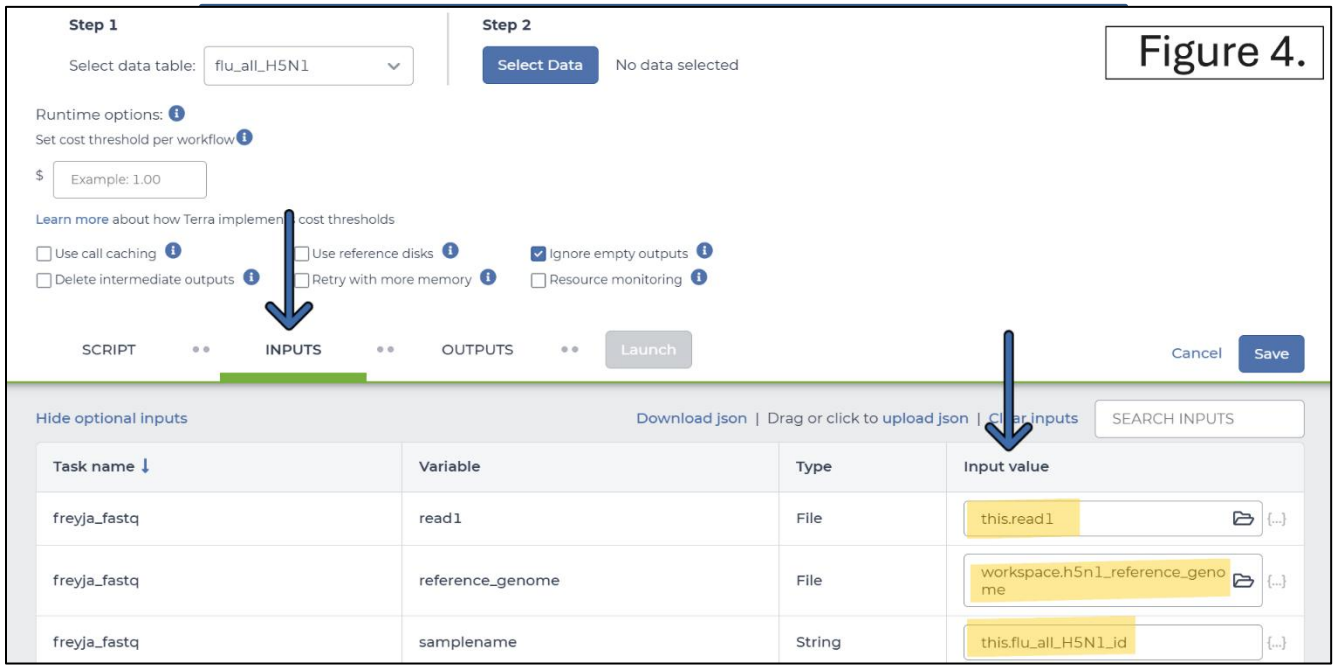


Figure 2.



a *Click th*e checkbox dropdown and select "all" *to select all samples in the data table; if the checkbox at the top is checked, only the first 100 samples in the data table will be selected*

b *Additionally, a subset of samples may be chosen using the search bar to filter before selecting the checkbox at the top to only select samples matching the search criteria (Fig 3, highlight)*

c Scroll to the bottom and click *ok*

8. Click on the inputs tab to specify settings (Fig 4)

a Manually set the first three attributes to the following, respectively

i. Reference genome can be found here: https://github.com/andersen-lab/Freyja-barcodes/tree/main/H5Nx/latest as "reference.fasta". This file will need to be downloaded to your computer and uploaded to the workspace data in the Data tab of your Terra workspace (see appendix 10.2 for adding workspace elements and files to Terra).

ii. Unique Terra data table name: *this.sample_id*

iii. Raw read1 file: *this.read1*

1. read2 is further down the page of inputs. If you are supplying Illumina PE reads, you will need to supply *this.read2* to the *read2* input.

b Scroll down the inputs to *update_db*. You will *either* have to provide input to *update_db* and *freyja_pathogen* OR provide an input for *freyja_barcodes* (Fig 5.1 and Fig 5.2).

i. **Option 1**: set *update_db* to *true* **and** set *freyja_pathogen* to "H5NX" (Fig 5.1)

ii. **Option 2**: Specify the *freyja_barcodes* file used to assign H5Nx lineages (Fig 5.2)

1. The H5Nx barcodes file in use by the Andersen lab can be found here as the *barcodes.csv* file: https://github.com/andersen-lab/Freyja-O/tree/main/H5Nx/latest. This may be saved in the workflow from prior analysis or copied from the respective Terra workspace files (see appendix 10.2 for details on copying workspace filepaths). It is not necessary to provide a lineage_metadata file to run Freyja_FASTQ for H5Nx.

iii. **(Optional)**: Provide *primer_bed* file for amplicon sequencing samples (Fig 5.3). Primer bed file: *workspace.[FILENAME]*

1. For appropriate H5Nx primer sets, ensure primer bed file (.bed file containing the primers used during sequencing) is uploaded to the workspace; it will then be available in the dropdown as *workspace.[FILENAME]*. *If amplicon sequencing was not done, there is no primer bed file; this field can be left blank. Freyja can be run without a primer bed file even for amplicon sequencing, as this is an optional field, but this is not recommended because primers will not be trimmed.*
    a. *See appendix 10.2 for adding workspace elements and files to Terra*

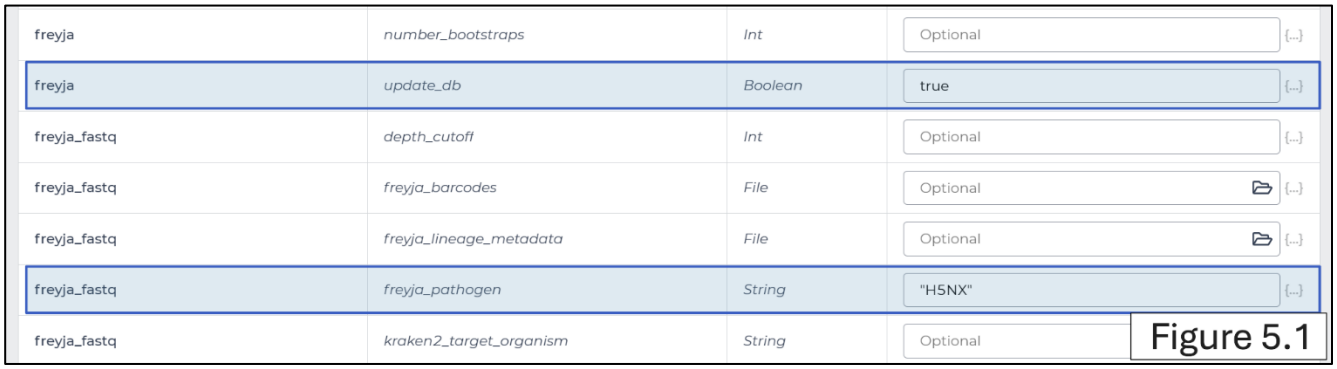


Figure 5.1

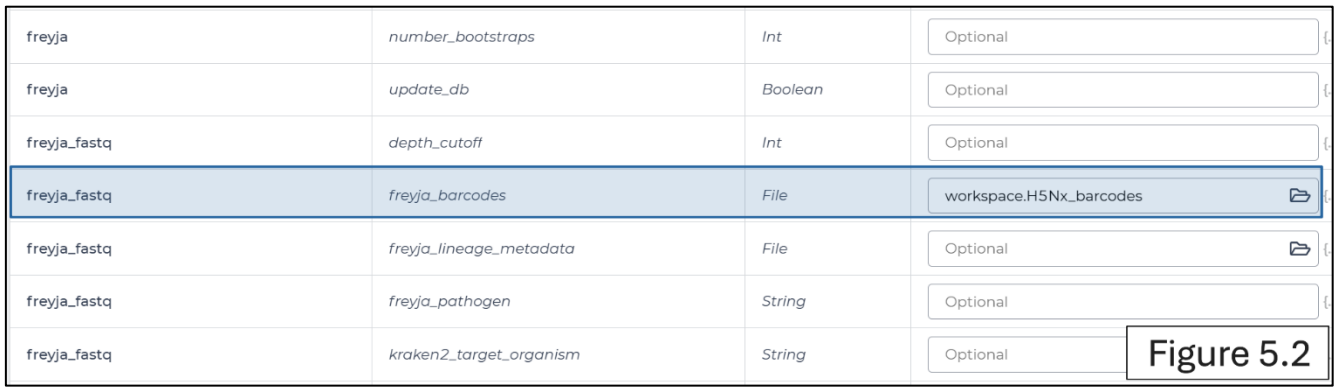
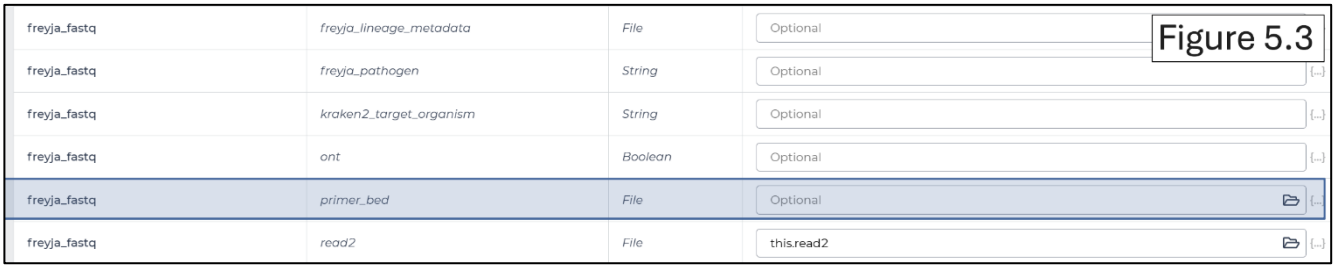
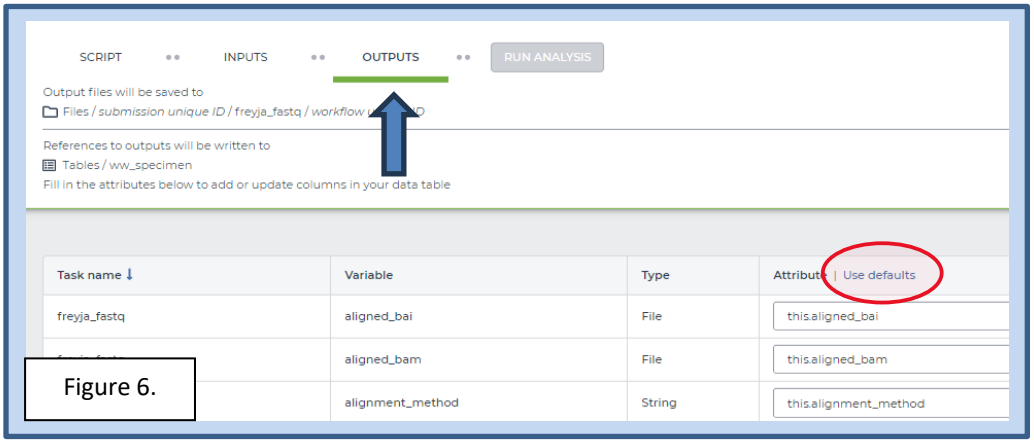


Figure 5.2



9. Specify outputs by clicking on the *outputs* tab and selecting *Use defaults* (Fig 6)



10. Click *save*
11. Launch the workflow by clicking *run analysis*; enter desired comments and click *launch*

4.2 DETERMINING LINEAGES, ABUNDANCES, AND COVERAGE METRICS

1. In the *data* tab, navigate to the Terra data table containing H5 metagenomic data
2. Click *settings* and select *none* to deselect all output columns (Fig 7)
3. To simplify the table, select the following outputs:
   a. *freyja_ barcode_version*
   b. *freyja_demixed*
   c. *freyja_depths*
   d. *freyja_metadata_version*
   e. *freyja_variants*

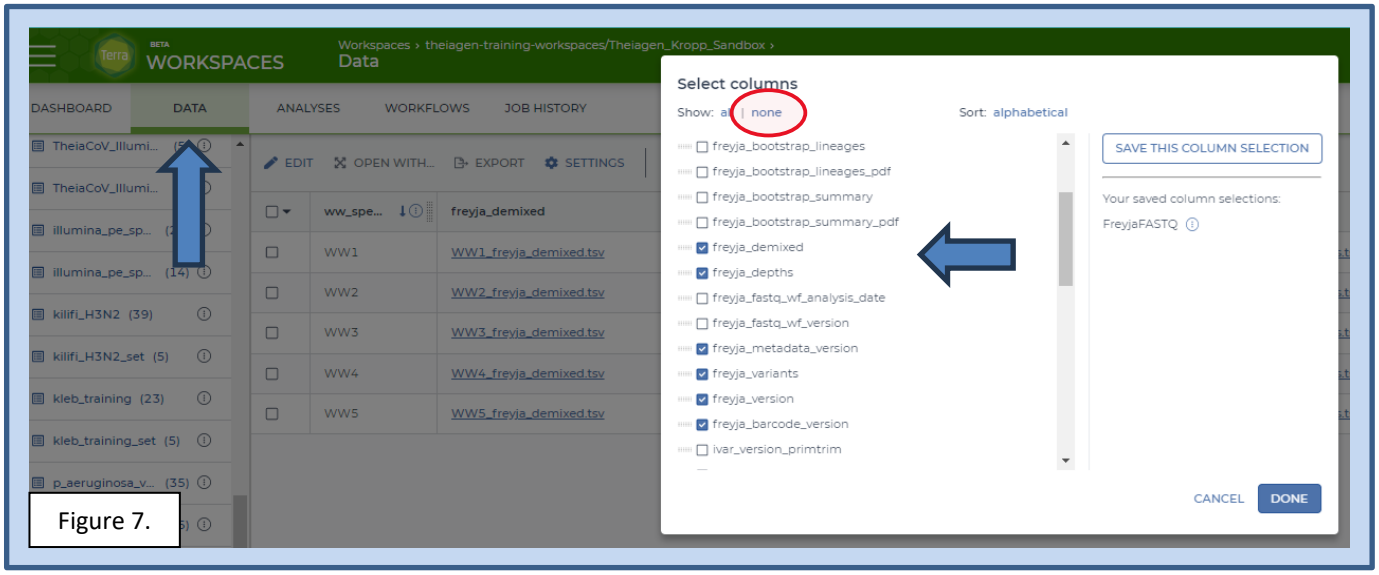


Figure 7.

4. Click on the *freyja_demixed column file* to determine the following sample information:
   a. Lineages identified
   b. Lineages and relative abundances of lineages
5. Click on the *freyja_variants column file* to see all variants identified within the sample
6. Click on the *freyja_depths column file* to determine the relative depth of coverage for every variant identified

5. QUALITY RECORDS

- Xu, X., et al. (1999). Influenza A virus (A/goose/Guangdong/1/1996 (H5N1)) hemagglutinin (HA) gene, complete cds. NC_007362.1. NCBI. https://www.ncbi.nlm.nih.gov/nuccore/NC_007362.1
- Workflow version and configuration settings (default and custom inputs)
- Curated lineages and usher barcodes files
- Raw read files
- freyja_demixed, freyja_variants, and freyja_depths tsv output files
- aligned_bam file for further visualizations

6. TROUBLESHOOTING

- Consult with internal staff familiar with this procedure or contact support@theiagen.com for troubleshooting inquiries
- For document edit requests, contact support@theiagen.com

7. LIMITATIONS

1. When creating visualizations from aggregated sample data over time, ensure all samples have been run with Freyja FASTQ using the same barcodes file

8. REFERENCES

1. Andersen Lab Github. https://github.com/andersen-lab/Freyja. Accessed on 07/09/2025..

9. REVISION HISTORY

Revision	Version	Release Date
Document creation	1	07/2025

10. APPENDICES

10. 1 IMPORTING FREYJA WORKFLOWS FROM DOCKSTORE

1. In the *Terra workspace* of interest, open the *workflows* tab and click *find a workflow* (Fig 8)
2. In the pop-up window, click *dockstore* (Fig 9)

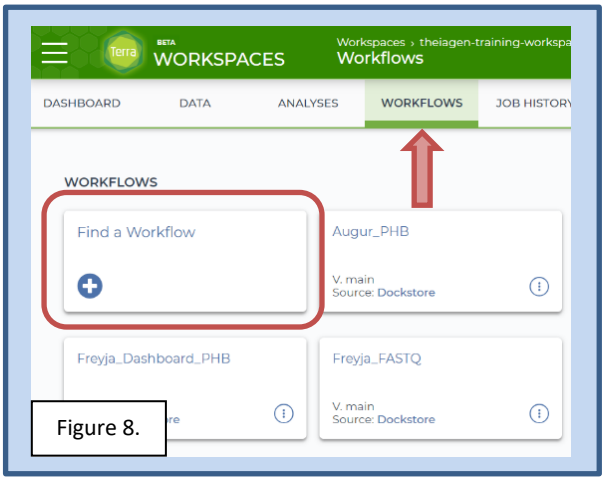


Figure 8.

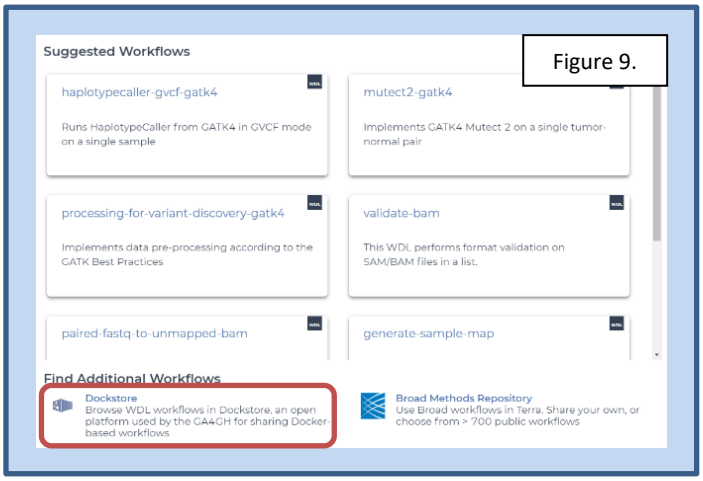


Figure 9.

3. To find the Theiagen Freyja FASTQ workflow, type "*Freyja_FASTQ_PHB*" in the search bar (Fig 10)
4. In the left hand sidebar, scroll down to Organization and select "*theiagen*" (Fig 10)
5. Find the workflow by looking at the file path suffix; click the name to *open the workflow* (Fig 10)

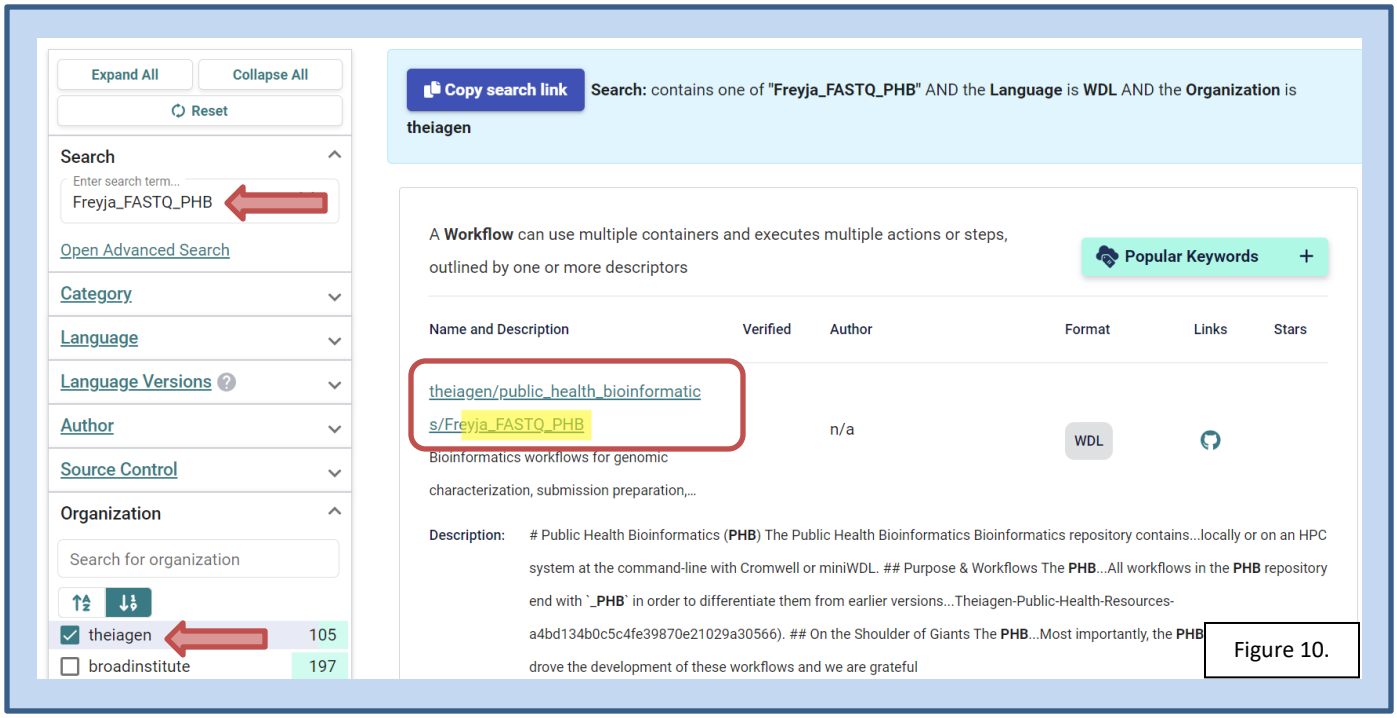


Figure 10.

6. Click *Terra* to launch the workflow in Terra (Fig 11)
7. Choose the *destination workspace* in the dropdown and click *import* or create a new workspace (Fig 12)

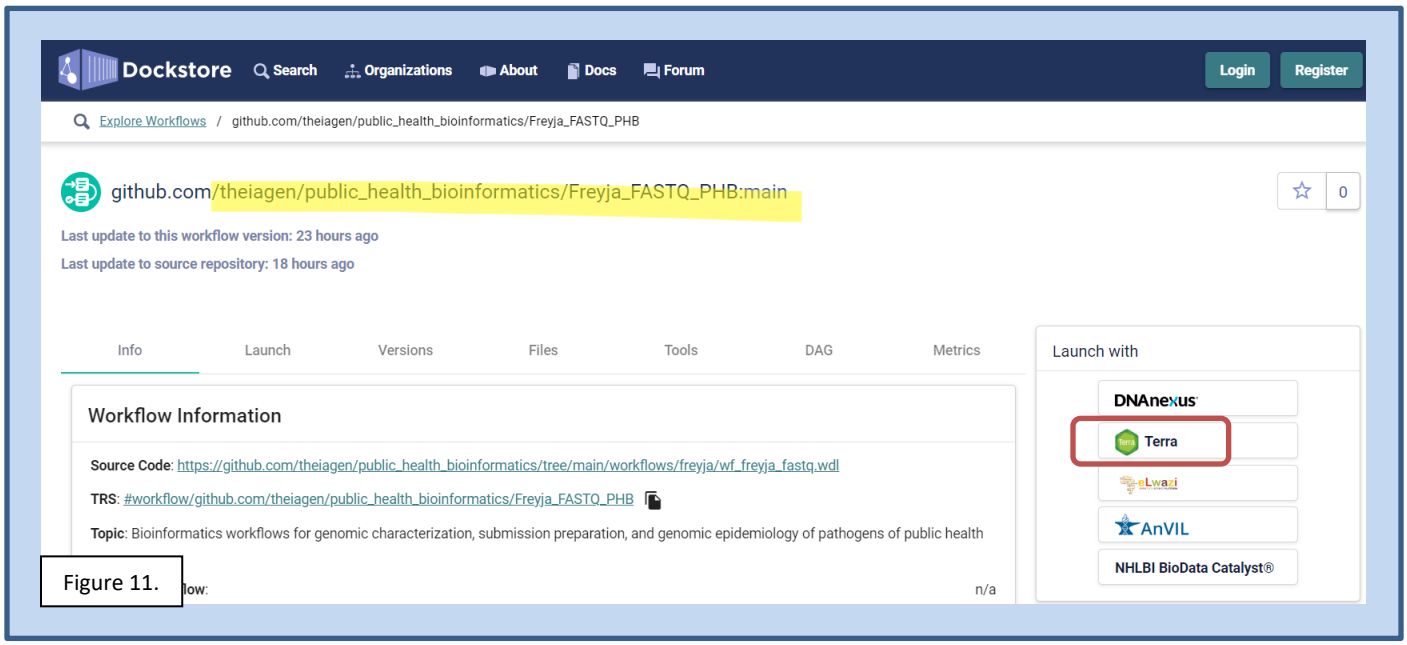


Figure 11.

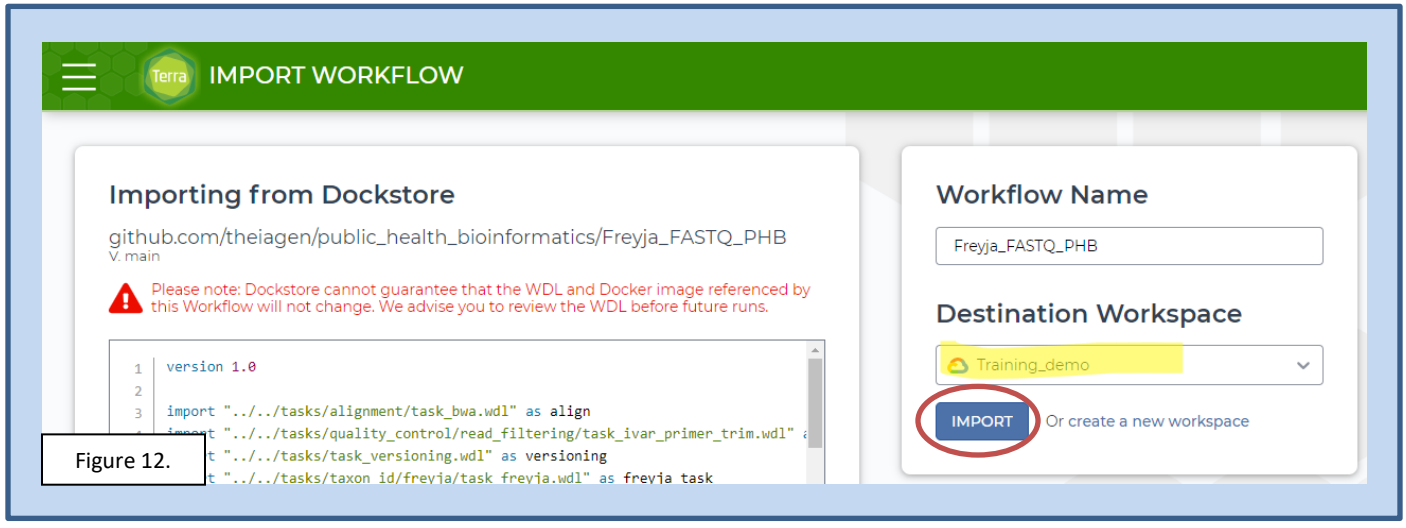


Figure 12.

10.2 ADDING TERRA WORKSPACE DATA ELEMENTS AND FILES

1. Navigate to the *Terra workspace* where analysis will be run
2. To upload local files, open the *Files* tab in the bottom left of the workspace (Fig 13)
   a. Click *upload* and select the file of interest; ensure the file name does not contain spaces
   b. Once the upload is complete, *right click* on the file name and click *copy link*
3. Open the *workspace data* tab (Fig 13) and click the *blue plus symbol* in the bottom right (Fig 13)
4. Click in the *key field* and *name the element* being added (Fig 14)
   a. E.g. to add a primer bed file, the key SWIFT_primer_bed *may be used*
5. In the value field, choose *string* as the value type
   a. *Paste the file path*; the string must start with *"gs://[FILENAME]…"*
      i.Add a *description* (e.g. updated date/initials), if desired and click the *blue checkmark* (Fig 14)

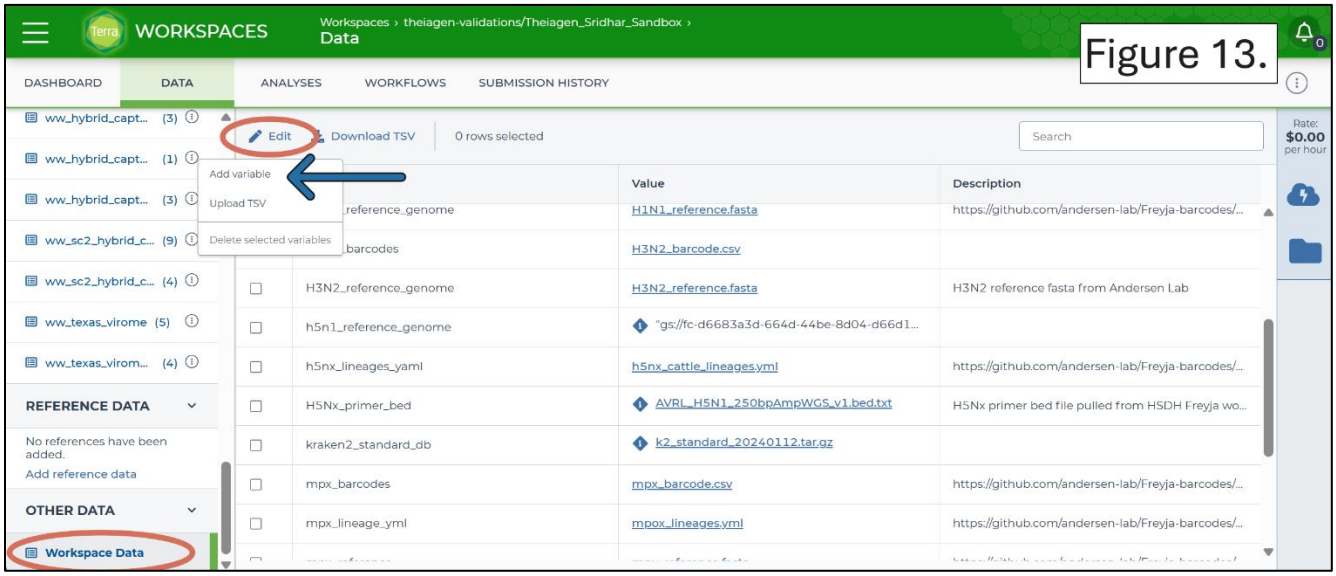


Figure 13.

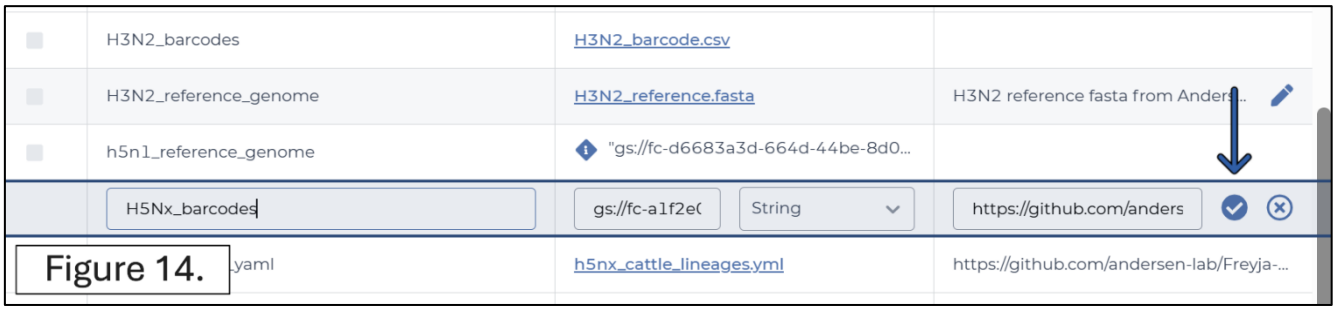


Figure 14.