Running SARS-CoV-2 Metagenomic Samples in Terra using Theiagen's Freyja FASTQ Workflow	
Document TG-FREY-01, Version 2	
Date:	Workflow Versions:
4/4/2024	PHB v2

1. PURPOSE/SCOPE

To standardize the process of running SARS-CoV-2 (SC2) metagenomic samples using Theiagen's Freyja FASTQ workflow in Terra to perform lineage deconvolution, abundance determination, and identify coverage metrics. Acceptable data types include Illumina paired end (PE) raw read files.

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

REQUIRED WORKFLOW INPUTS FILES

- Raw Illumina PE read files
- Primer bed file
- Reference genome
- Curated lineages file*
- Usher 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
TG-FREY-04	Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow

4. PROCEDURE

4.1 RUNNING THE FREYJA FASTQ WORKFLOW

1. Open Terra and navigate to the *workflows* tab within the workspace containing wastewater data
2. Select the *Freyja_FASTQ_PHB* workflow (Fig 1)
3. *Uncheck use call caching* (Fig 2)

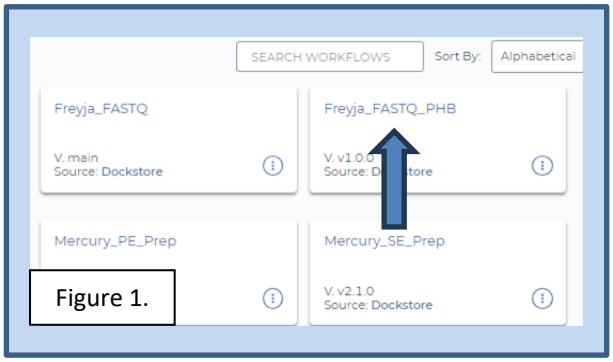

Figure 1.

*Freyja_FASTQ_PHB can be run without the curated_lineages and usher_barcodes input files by setting the update_db input value to true; this performs analysis using the most up-to-date reference files stored in the Freyja Github repository

4. Choose the latest version of *version 2* 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)

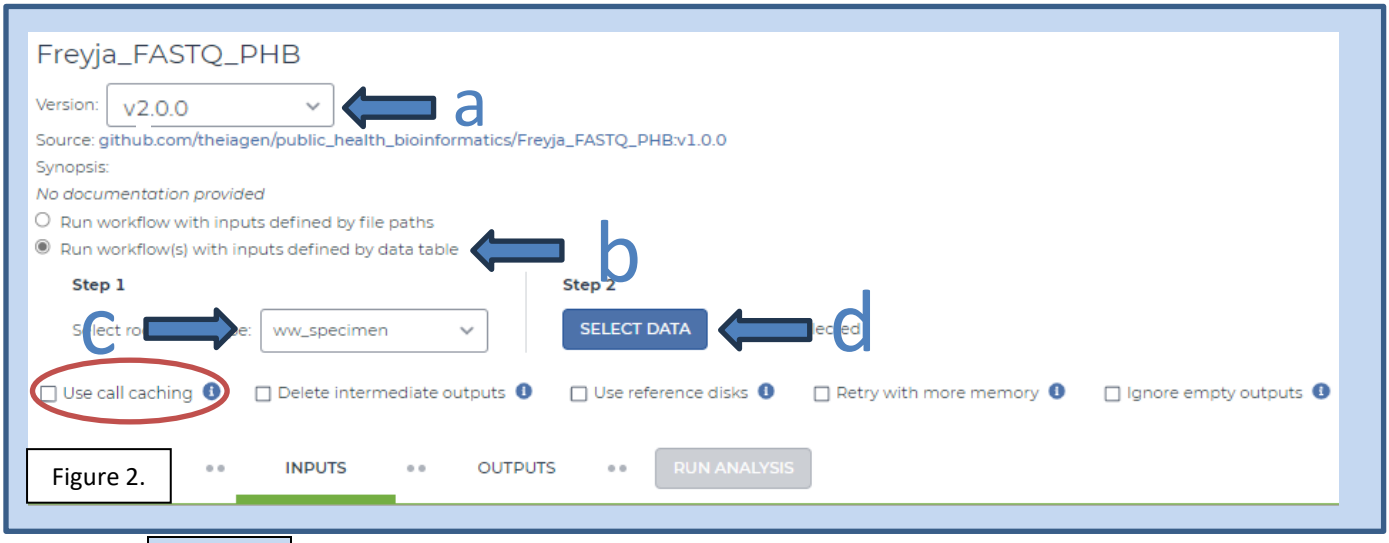

Figure 2.

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)
   - *a* *Click the* checkbox dropdown and 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*

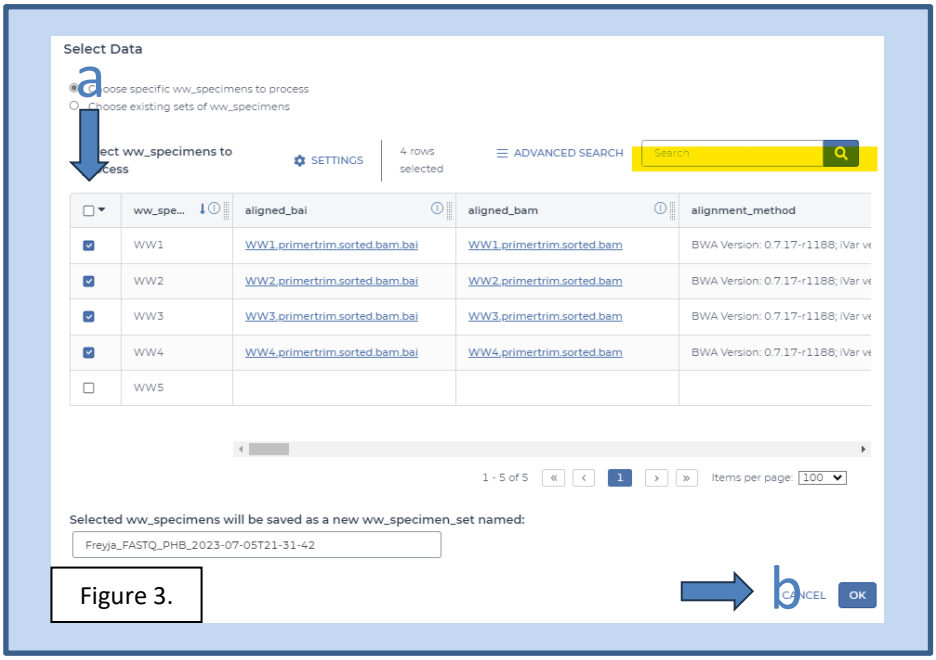

Figure 3.

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

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

i. Primer bed file: *workspace.SWIFT_primer_bed*

1. For other primer sets, ensure primer bed files are uploaded to the workspace; they will then be available in the dropdown as *workspace.[FILENAME]*
   a. *See appendix 10.2 for adding workspace elements and files to Terra*
2. *Find other SC2 primer bed files available on the Theiagen Docker Image and Reference Materials for SARS-CoV-2 Genomic Characterization Notion page under Terra Resources for TheiaCoV > SARS-CoV-2 Primer Scheme BED Files*

ii. Raw read1 file: *this.read1*

iii. Raw read2 file: *this.read2*

iv. Reference genome: *"gs://theiagen-public-files/terra/freyja-files/nCoV-2019.reference.fasta"*

v. Unique Terra data table name: *this.ww_specimen_id*

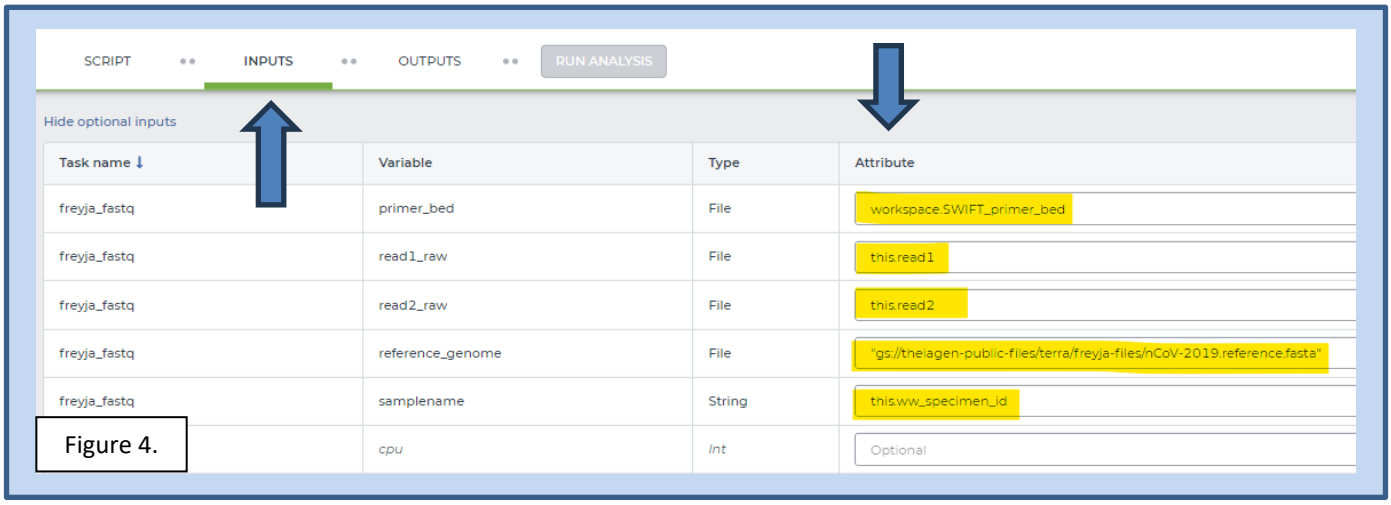

Figure 4.

b Specify the *curated lineages and usher barcodes files* used to assign SC2 lineages; follow one of the three options below (Fig 5):

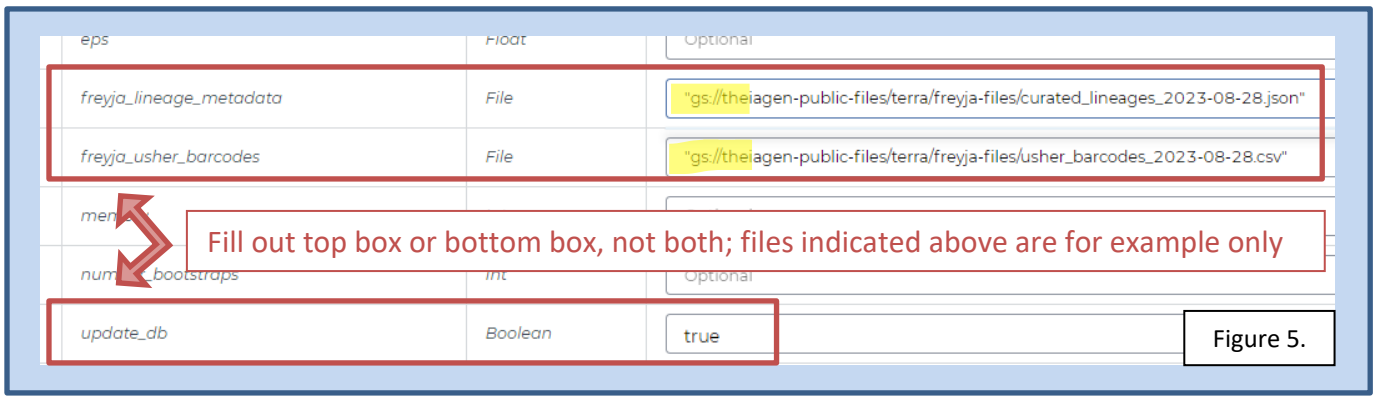

Figure 5.

i. Option 1: To run Freyja FASTQ for the first time or to use the most up-to-date SC2 curated_lineages and usher_barcodes files created by the Freyja developers[1], set the *update_db* input field to *true*

1. *To save these reference file versions for use in a future analysis, run the Freyja_Update_PHB workflow; see the TG-FREY-04: Freyja_Update SOP for details*

ii. Option 2: Run recently updated versions of the curated_lineages and usher_barcodes reference files saved as Terra workspace data elements using the Freyja_Update_PHB workflow and SOP

1. ***NOTE:*** *Freyja_Update_PHB should be run prior to running Freyja_FASTQ_PHB for this option*
2. Enter additional Freyja_FASTQ_PHB workflow inputs as follows:
   a. freyja_lineage_metadata: *workspace.FreyjaLineageMetadata*
   b. freyja_usher_barcodes: *workspace.FreyjaUsherBarcodes*

iii. Option 3: To use specific versions of the curated_lineages and usher_barcodes files that have been used previously for analysis, enter the corresponding *"gs://[FILENAME]…"* filepaths for *freyja_lineage_metadata* and *freyja_usher_barcodes* input values; these 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)

iv. ***NOTE****: When aggregating and analyzing samples sequenced over time, it is important that the same curated lineages and usher barcodes files are used to run Freyja FASTQ workflow and identify lineages present within a sample. Updated files can contain new lineage assignments that samples run with Freyja FASTQ using previous file versions will be missing*

9. Specify outputs by clicking on the *outputs* tab and *use defaults* (Fig 6)
10. Click *save*
11. Launch the workflow by clicking *run analysis*; enter desired comments and click *launch*

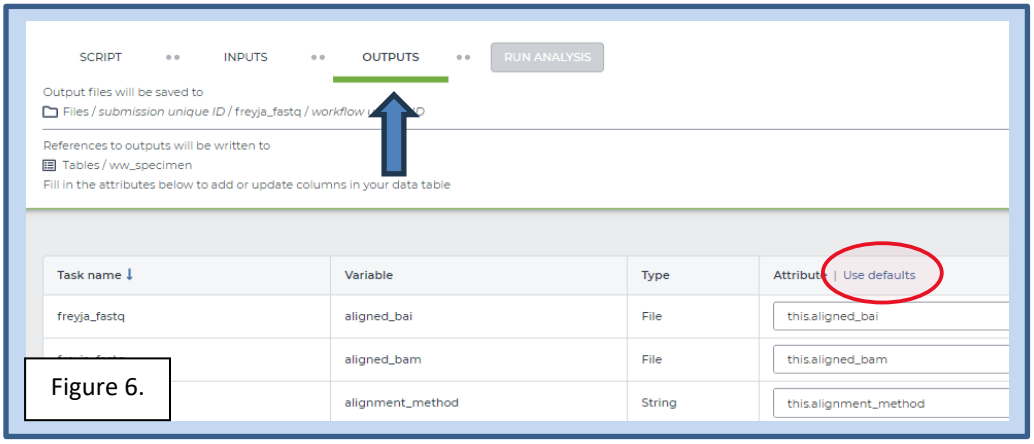

Figure 6.

4.2 DETERMINING LINEAGES, ABUNDANCES, AND COVERAGE METRICS

1. In the *data* tab, navigate to the Terra data table containing SC2 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*

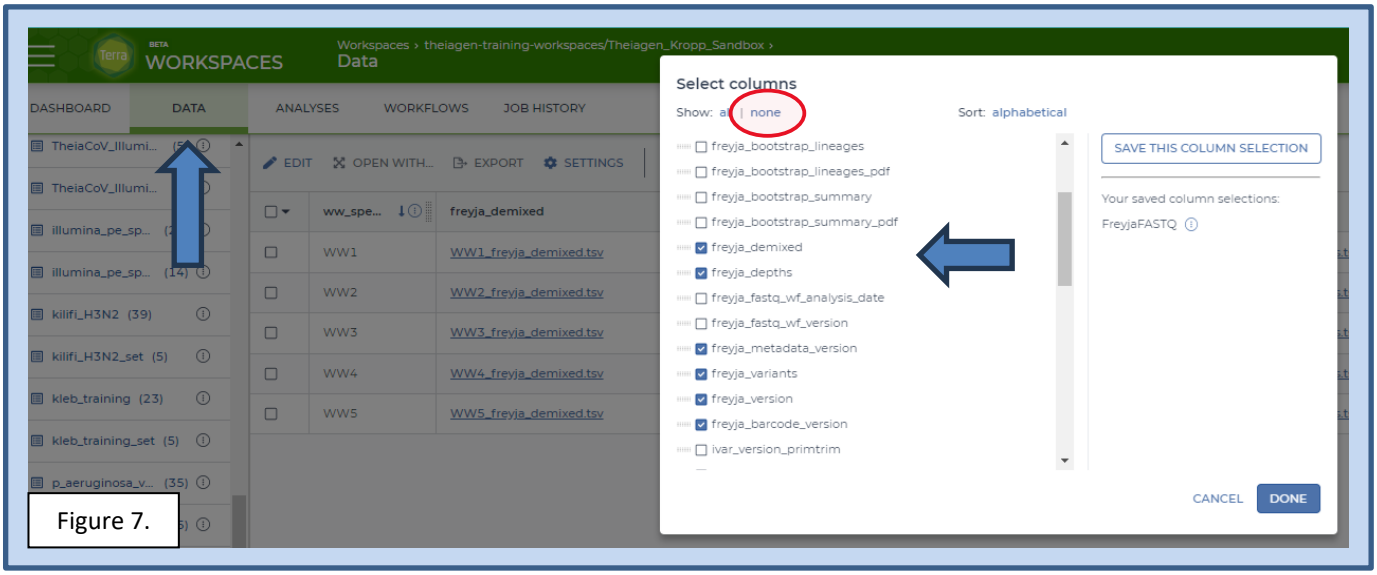

Figure 7.

4. Click on the *freyja_demixed column file* to determine the following sample information:
   a. Lineages identified
   b. Distribution of variants of concern (VOCs)
   c. 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

- Wu, F., et al. (2020). Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. NC_045512.2. [FASTA Genome Assembly]. NCBI. https://www.ncbi.nlm.nih.gov/nuccore/1798174254.
- 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 curated_lineages and usher_barcodes files
2. Freyja FASTQ can only be used to analyze SC2 data from Illumina PE sequencing files

8. REFERENCES

1. Andersen Lab Github. https://github.com/andersen-lab/Freyja. Accessed on 4/5/2024.
2. Primer BED files available at Theiagen *Docker Image and Reference Materials for SARS-CoV-2 Genomic Characterization Notion page*

9. REVISION HISTORY

Revision	Version	Release Date
Document creation	1	8/2023
Added workflow input information and usage of curated_lineages and usher_barcodes reference files, updated quality records and references; added appendices	2	4/2024

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)

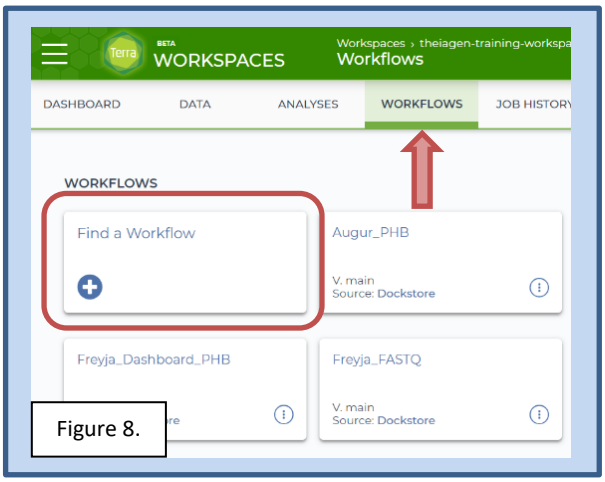

Figure 8.

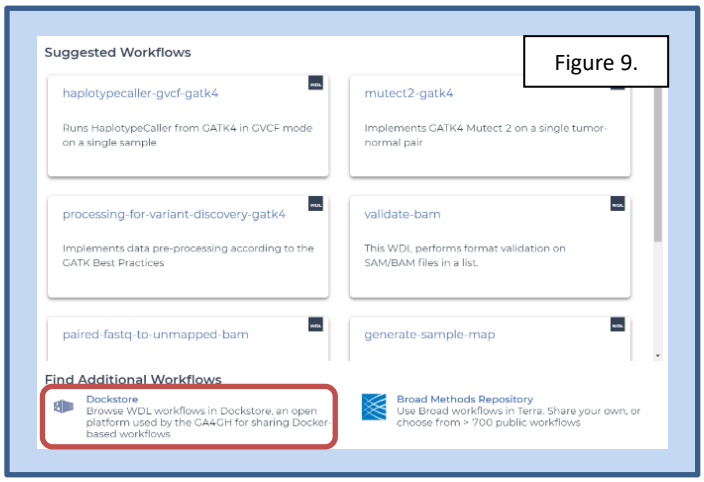

Figure 9.

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)

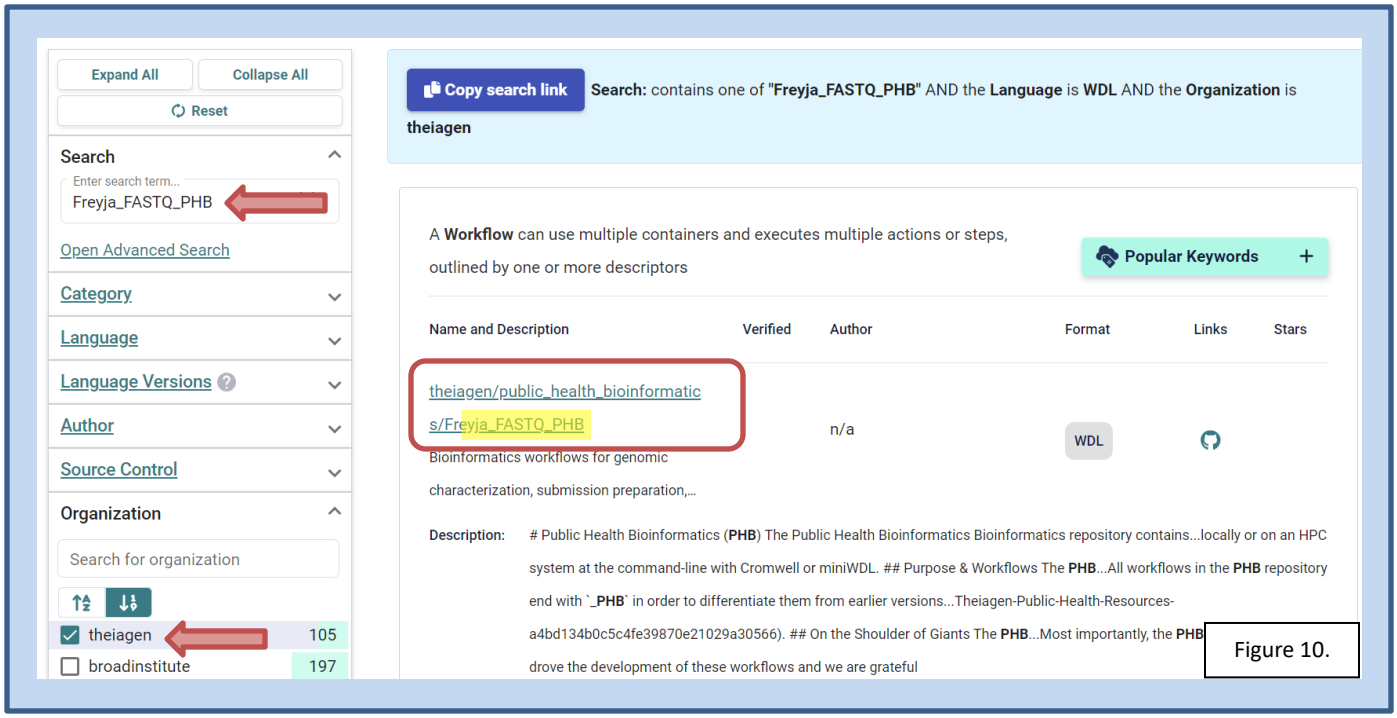

Figure 10.

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)

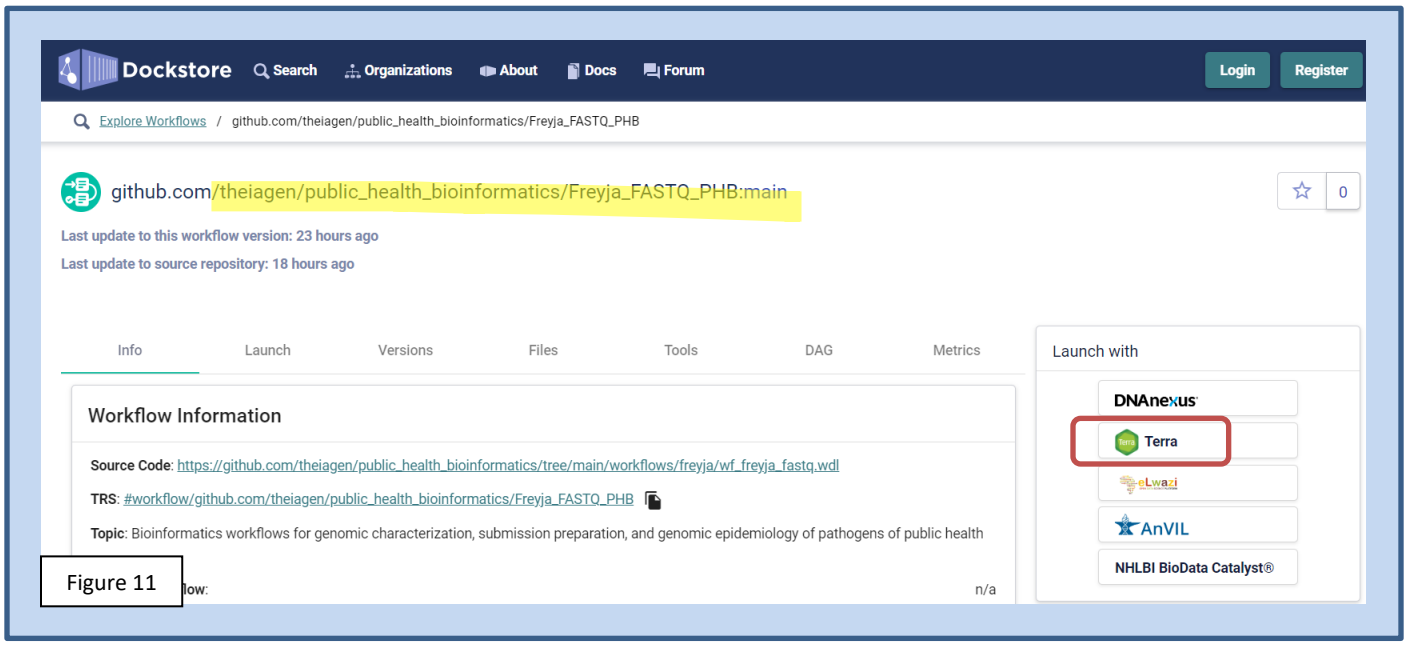

Figure 11

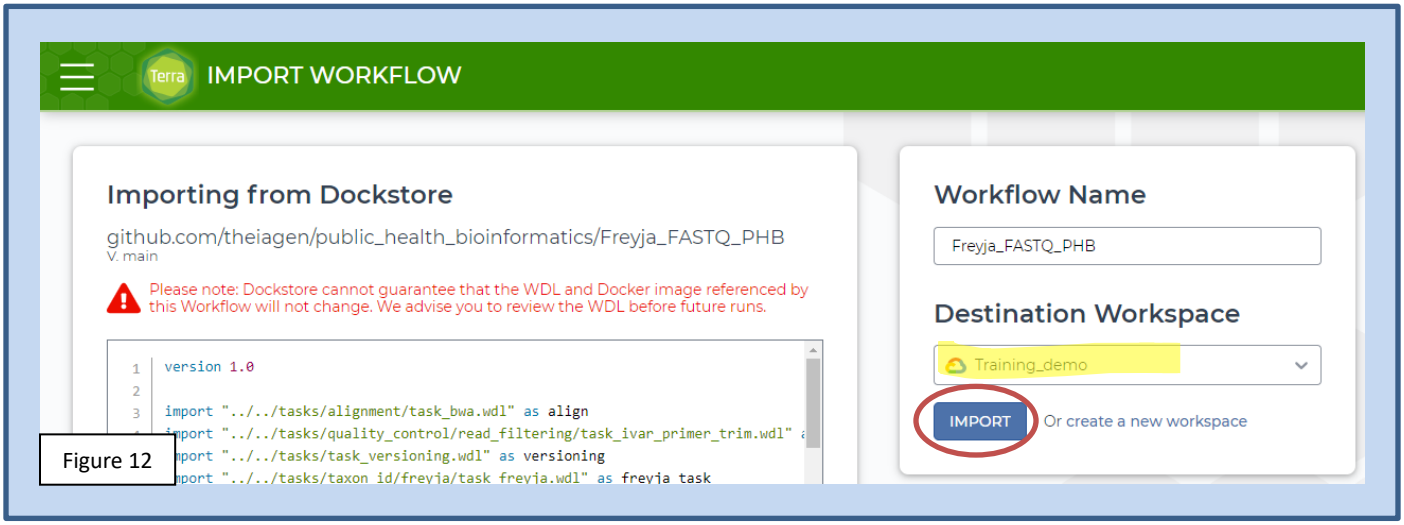

Figure 12

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)

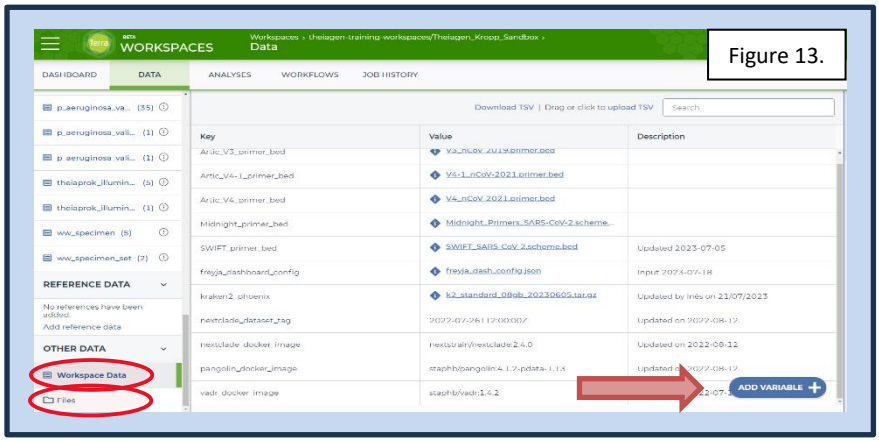

Figure 13.

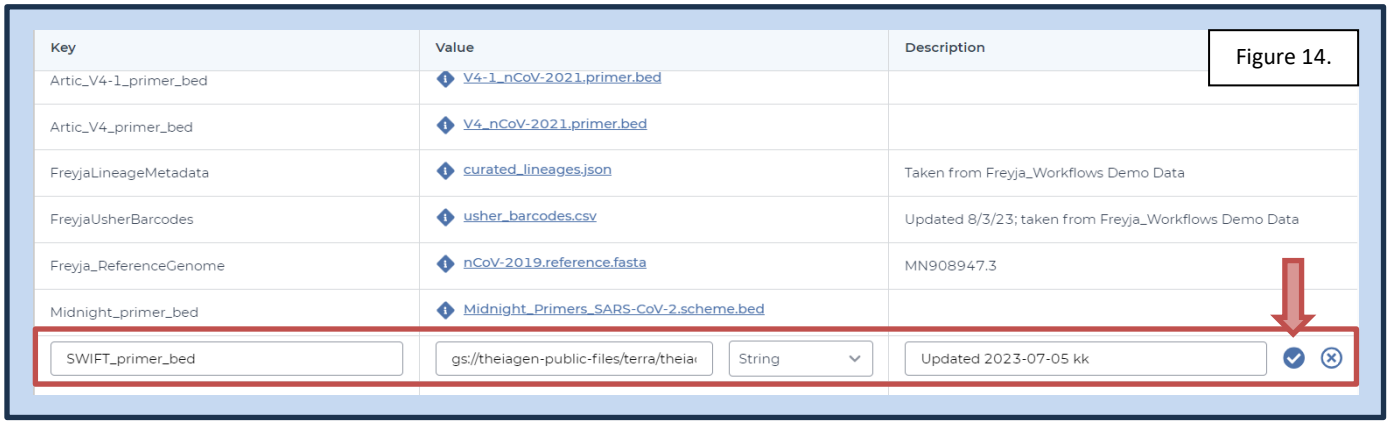

Figure 14.