Running Influenza A, H3N2 Metagenomic Samples in Terra using Theiagen's Freyja FASTQ Workflow

Document TG-FREY-H3N2, Version 1

Date:	Workflow Versions:
4/4/2025	PHB v3

1. PURPOSE/SCOPE

To standardize the process of running Influenza A, H3N2 (H3N2) 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. Please note that this SOP should NOT be used to run Influenza A, H5N1 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
  - For workflow import, see Appendix 10.1.

REQUIRED WORKFLOW INPUTS FILES

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

*For Influenza A/H3N2, Freyja_FASTQ_PHB should not be run with the curated_lineages and usher_barcodes input files. The *update_db* input value should be set to *false*.

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 within the workspace containing wastewater data.
2. Select the *Freyja_FASTQ_PHB* workflow (Fig 1).
3. Uncheck *use call caching* (Fig 2).

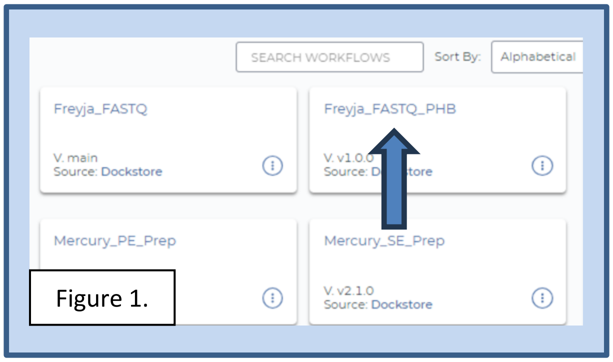
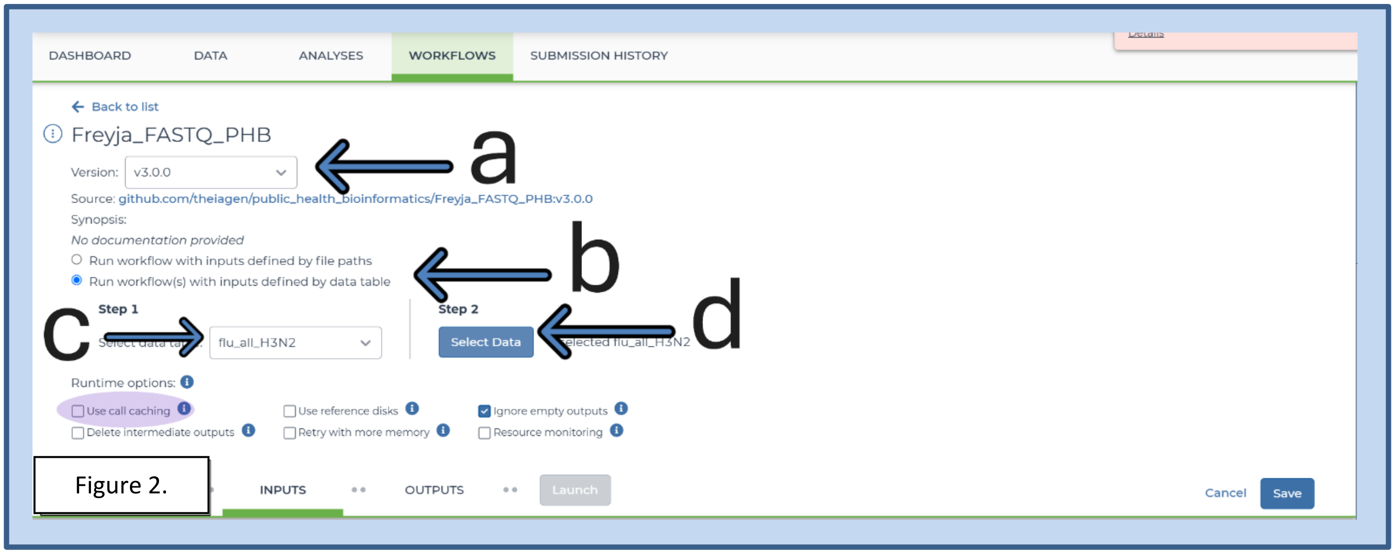


Figure 1.



Freyja_FASTQ_PHB

Version: v3.0.0

Source: github.com/theiagen/public_health_bioinformatics/Freyja_FASTQ_PHB:v3.0.0

Synopsis:

No documentation provided

○ Run workflow with inputs defined by file paths

● Run workflow(s) with inputs defined by data table

Step 1 Select data table: flu_all_H3N2

Step 2 Selected flu_all_H3N2

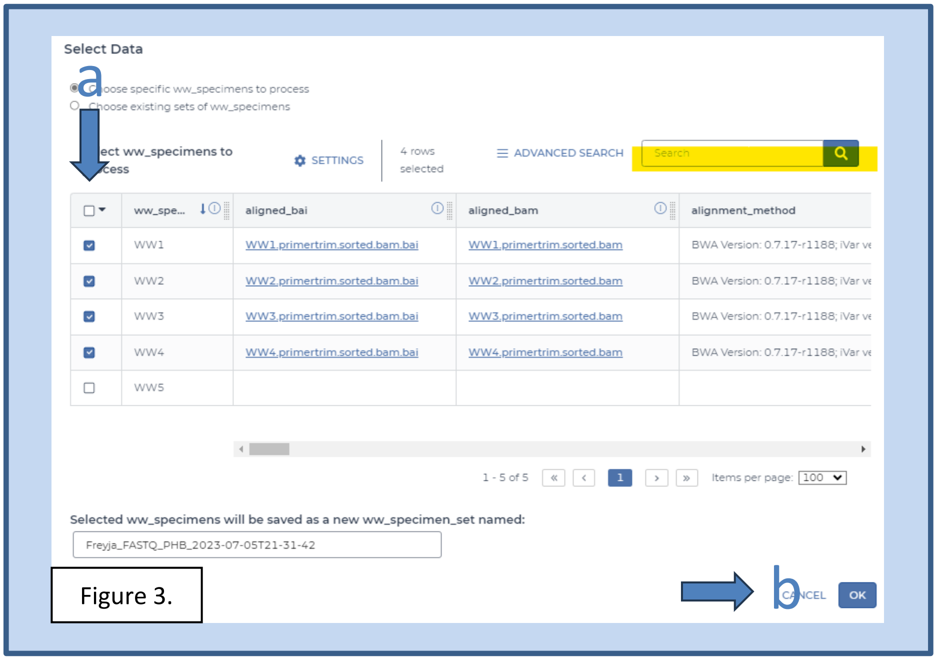
Runtime options:

☐ Use call caching ☐ Use reference disks ☑ Ignore empty outputs

☐ Delete intermediate outputs ☐ Retry with more memory ☐ Resource monitoring

Figure 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).
    a *Click* the 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*.



Select Data

● Choose specific ww_specimens to process
○ Choose existing sets of ww_specimens

ect ww_specimens to cess

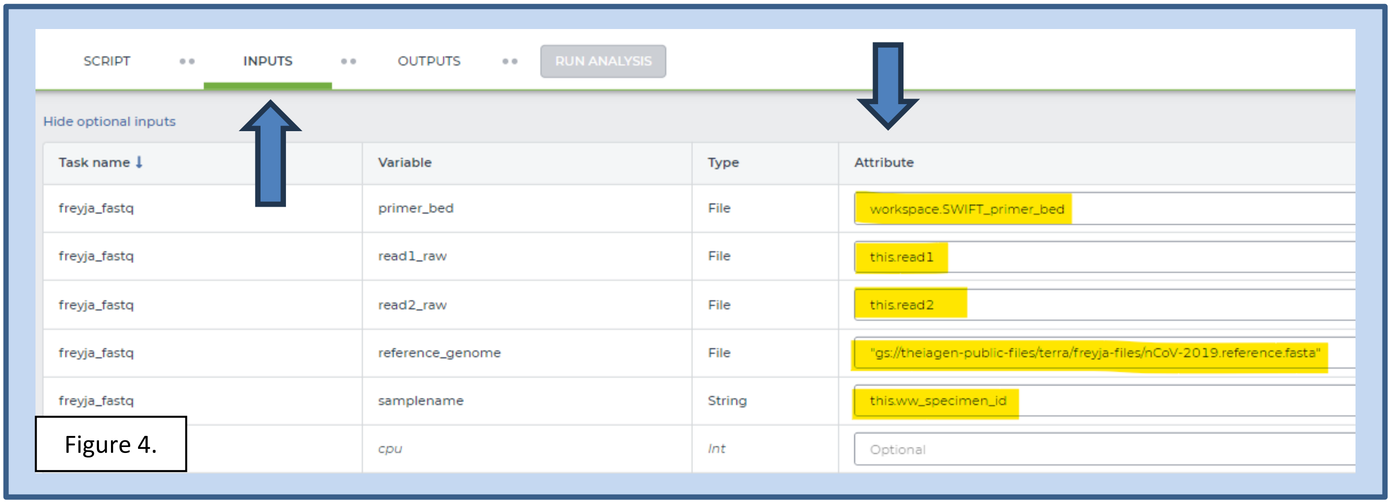
☐	ww_spe...	aligned_bai	aligned_bam	alignment_method
☑	WW1	WW1.primertrim.sorted.bam.bai	WW1.primertrim.sorted.bam	BWA Version: 0.7.17-r1188; iVar ve
☑	WW2	WW2.primertrim.sorted.bam.bai	WW2.primertrim.sorted.bam	BWA Version: 0.7.17-r1188; iVar ve
☑	WW3	WW3.primertrim.sorted.bam.bai	WW3.primertrim.sorted.bam	BWA Version: 0.7.17-r1188; iVar ve
☑	WW4	WW4.primertrim.sorted.bam.bai	WW4.primertrim.sorted.bam	BWA Version: 0.7.17-r1188; iVar ve
☐	WW5			

Selected ww_specimens will be saved as a new ww_specimen_set named:

Freyja_FASTQ_PHB_2023-07-05T21-31-42

Figure 3.

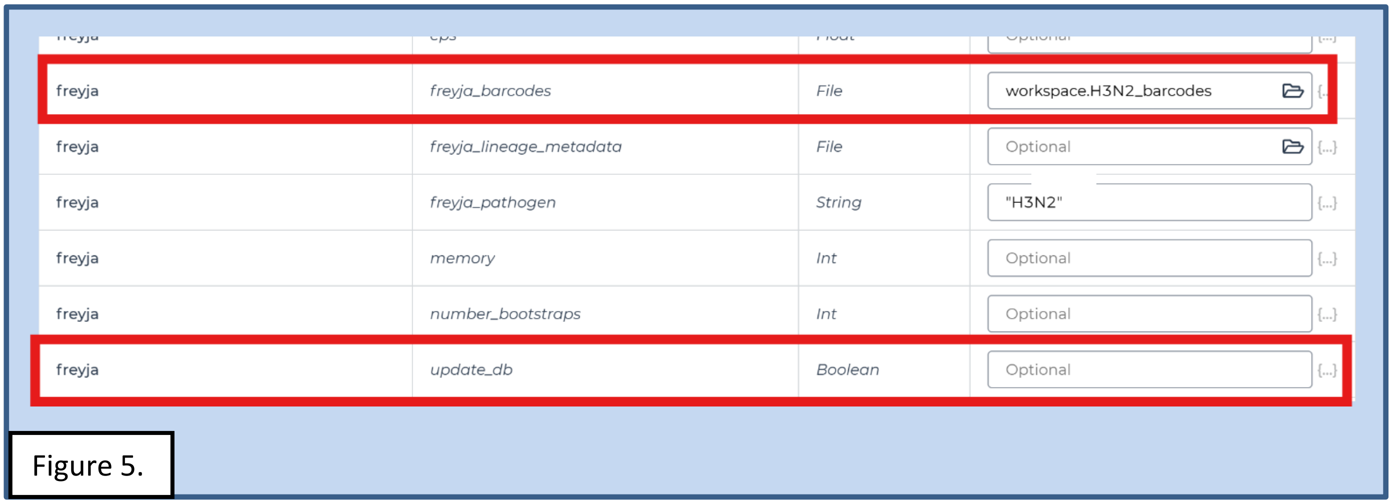
8. Click on the inputs tab to specify settings (Fig 4).
   a Manually set the first five attributes to the following, respectively.
      i. Primer bed file: *workspace.[FILENAME]*
         1. For appropriate H3N2 primer sets, ensure primer bed file (.bed file containing the primers used during sequencing) 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.*
      ii. Raw read1 file: *this.read1*
      iii. Raw read2 file: *this.read2*
      iv. Reference genome can be found here: https://github.com/andersen-lab/Freyja-barcodes/tree/main/H3N2/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).
      v. Unique Terra data table name: *this.sample_id*.



Task name ↓	Variable	Type	Attribute
freyja_fastq	primer_bed	File	workspace.SWIFT_primer_bed
freyja_fastq	read1_raw	File	this.read1
freyja_fastq	read2_raw	File	this.read2
freyja_fastq	reference_genome	File	"gs://theiagen-public-files/terra/freyja-files/nCoV-2019.reference.fasta"
freyja_fastq	samplename	String	this.ww_specimen_id
	cpu	*Int*	Optional

Figure 4.

b Specify the *barcodes file* used to assign H3N2 lineages (Fig 5). *The H3N2 barcodes file in use by the Andersen lab can be found here as the barcodes.csv file: [Freyja-barcodes/H3N2/latest at main · andersen-lab/Freyja-barcodes](). 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 H3N2*.



freyja	*freyja_barcodes*	*File*	workspace.H3N2_barcodes
freyja	*freyja_lineage_metadata*	*File*	Optional
freyja	*freyja_pathogen*	*String*	"H3N2"
freyja	*memory*	*Int*	Optional
freyja	*number_bootstraps*	*Int*	Optional
freyja	*update_db*	*Boolean*	Optional

Figure 5.

c Verify the *update_db* entry is empty (Fig 5).

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*.

Task name	Variable	Type	Attribute
freyja_fastq	aligned_bai	File	this.aligned_bai
freyja_fastq	aligned_bam	File	this.aligned_bam
	alignment_method	String	this.alignment_method

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*

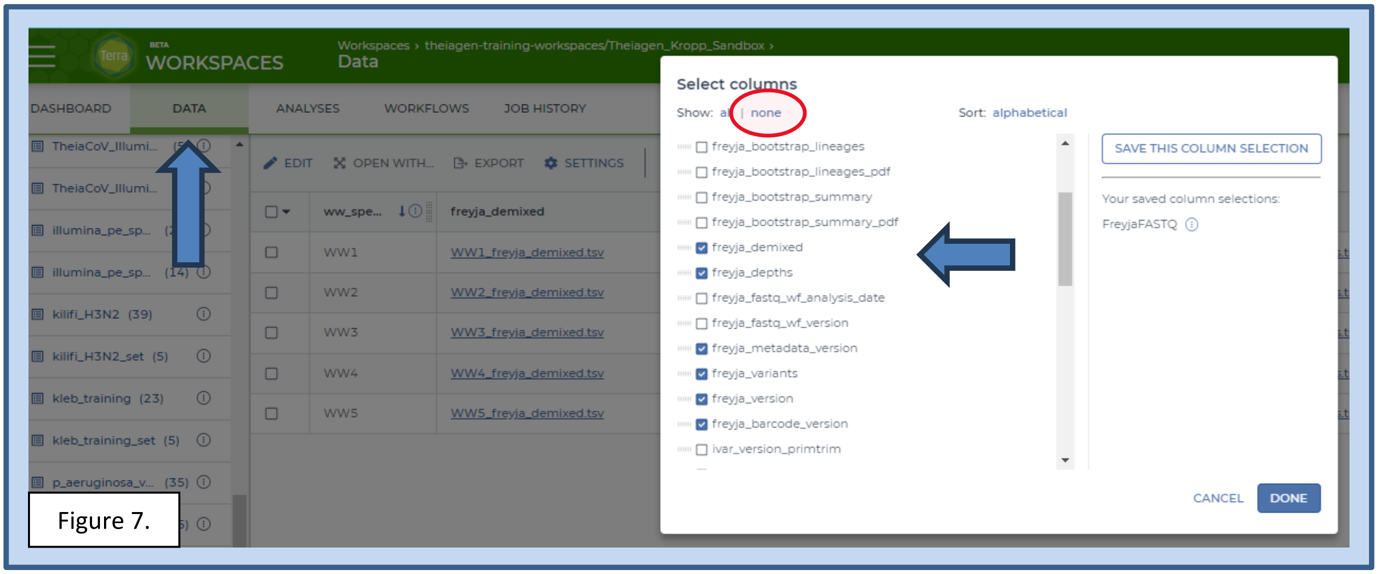


Figure 7.

4. Click on the *freyja_demixed column file* to determine the following sample information:
   a. Lineages identified

  b. 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

- Wentworth, D.E., et al. (2014). Influenza A virus (A/Wisconsin/67/2005(H3N2)) hemagglutinin (HA) gene, complete cds. CY163680.1.. NCBI. https://www.ncbi.nlm.nih.gov/nuccore/575499275.
- 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.
2. This procedure is not intended for analysis of Influenza A, H5N1 samples.

8. REFERENCES

1. Andersen Lab Github. https://github.com/andersen-lab/Freyja. Accessed on 3/11/2025.

9. REVISION HISTORY

Revision	Version	Release Date
Document creation	1	05/2025

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).

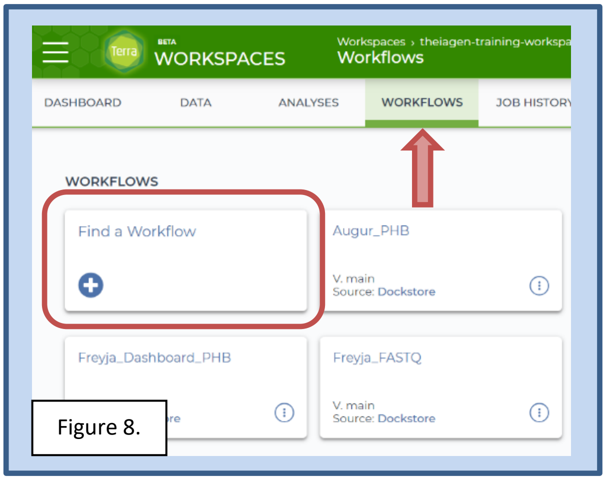
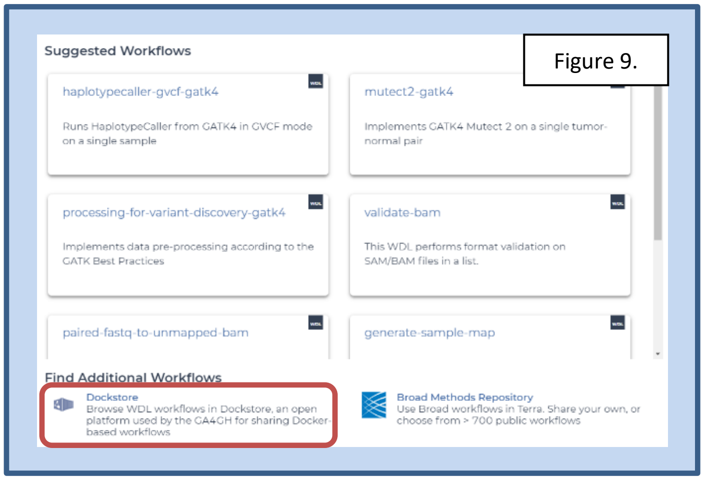


Figure 8.



Runs HaplotypeCaller from GATK4 in GVCF mode on a single sample

Implements GATK4 Mutect 2 on a single tumor-normal pair

Implements data pre-processing according to the GATK Best Practices

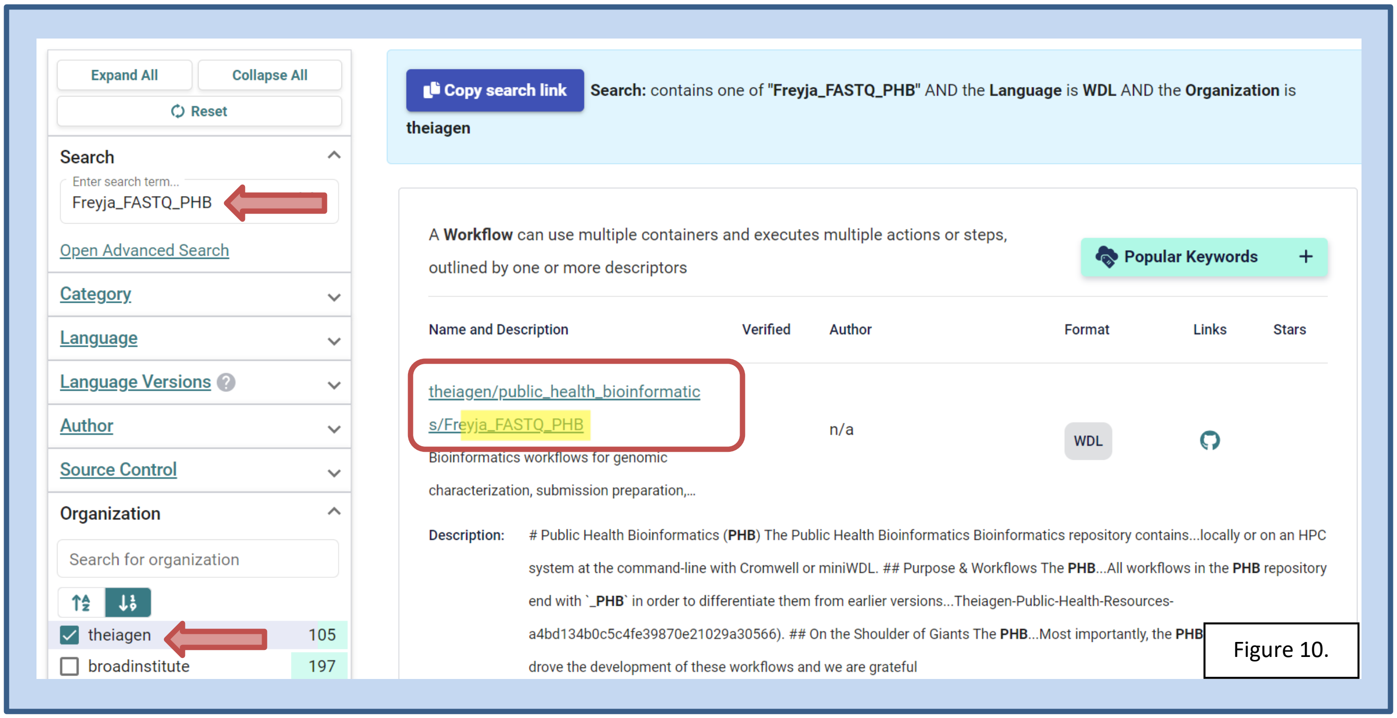
This WDL performs format validation on SAM/BAM files in a list.

Browse WDL workflows in Dockstore, an open platform used by the GA4GH for sharing Docker-based workflows

Use Broad workflows in Terra. Share your own, or choose from > 700 public workflows

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).



Search: contains one of "**Freyja_FASTQ_PHB**" AND the **Language** is **WDL** AND the **Organization** is **theiagen**

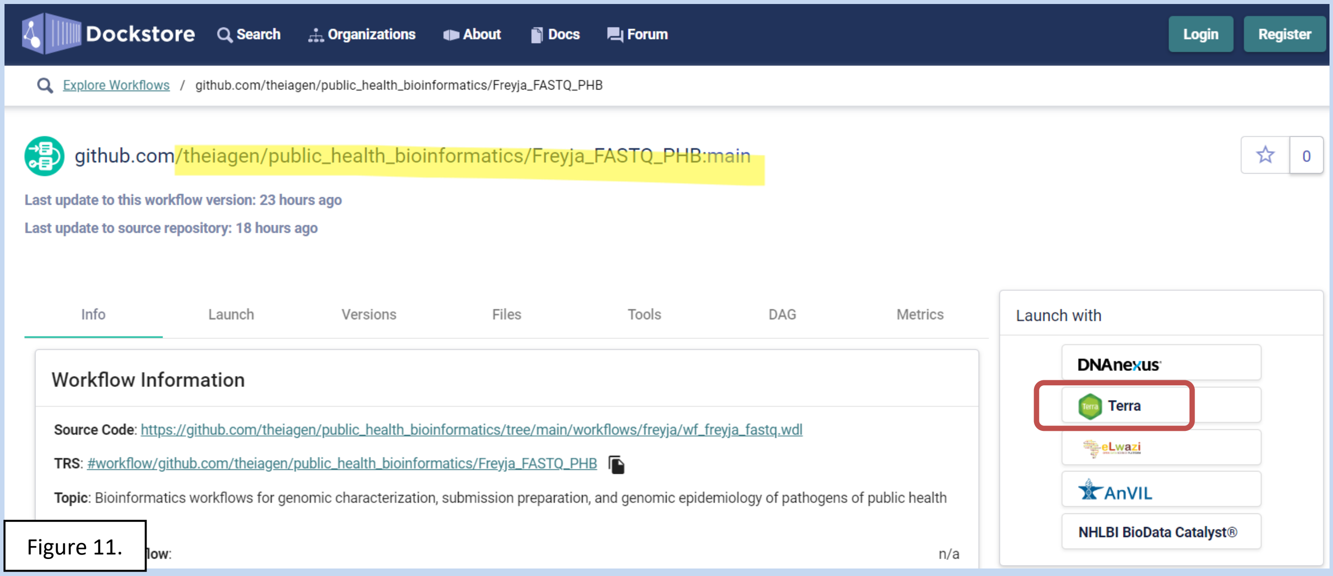
A **Workflow** can use multiple containers and executes multiple actions or steps, outlined by one or more descriptors

Bioinformatics workflows for genomic characterization, submission preparation,...

Description: # Public Health Bioinformatics (**PHB**) The Public Health Bioinformatics Bioinformatics repository contains...locally or on an HPC system at the command-line with Cromwell or miniWDL. ## Purpose & Workflows The **PHB**...All workflows in the **PHB** repository end with `_**PHB**` in order to differentiate them from earlier versions...Theiagen-Public-Health-Resources-a4bd134b0c5c4fe39870e21029a30566). ## On the Shoulder of Giants The **PHB**...Most importantly, the **PHB** drove the development of these workflows and we are grateful

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).



Explore Workflows / github.com/theiagen/public_health_bioinformatics/Freyja_FASTQ_PHB

github.com/theiagen/public_health_bioinformatics/Freyja_FASTQ_PHB:main

Last update to this workflow version: 23 hours ago

Last update to source repository: 18 hours ago

Workflow Information

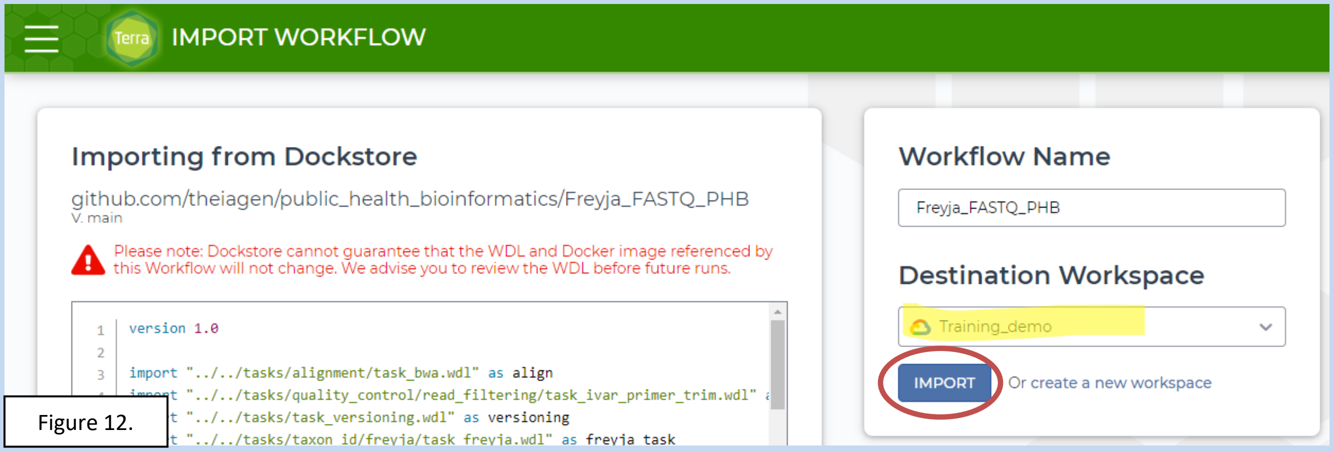
Source Code: https://github.com/theiagen/public_health_bioinformatics/tree/main/workflows/freyja/wf_freyja_fastq.wdl

TRS: #workflow/github.com/theiagen/public_health_bioinformatics/Freyja_FASTQ_PHB

Topic: Bioinformatics workflows for genomic characterization, submission preparation, and genomic epidemiology of pathogens of public health

low: n/a

Figure 11.



Importing from Dockstore

github.com/theiagen/public_health_bioinformatics/Freyja_FASTQ_PHB
V. main

Please note: Dockstore cannot guarantee that the WDL and Docker image referenced by this Workflow will not change. We advise you to review the WDL before future runs.

```
version 1.0

import "../../tasks/alignment/task_bwa.wdl" as align
import "../../tasks/quality_control/read_filtering/task_ivar_primer_trim.wdl" a
t "../../tasks/task_versioning.wdl" as versioning
t "../../tasks/taxon_id/freyja/task_freyja.wdl" as freyja_task
```

Workflow Name

Freyja_FASTQ_PHB

Destination Workspace

Training_demo

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).

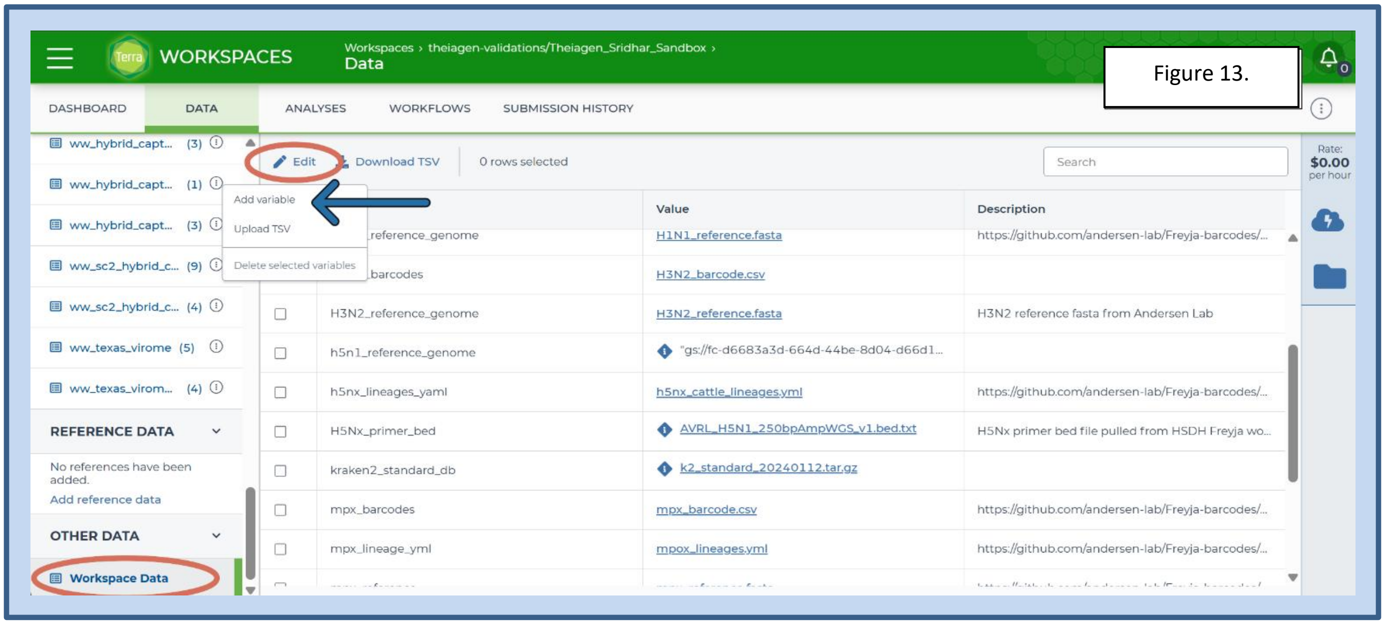


Figure 13.

	Value	Description
reference_genome	H1N1_reference.fasta	https://github.com/andersen-lab/Freyja-barcodes/...
barcodes	H3N2_barcode.csv	
H3N2_reference_genome	H3N2_reference.fasta	H3N2 reference fasta from Andersen Lab
h5n1_reference_genome	"gs://fc-d6683a3d-664d-44be-8d04-d66d1...	
h5nx_lineages_yaml	h5nx_cattle_lineages.yml	https://github.com/andersen-lab/Freyja-barcodes/...
H5Nx_primer_bed	AVRL_H5N1_250bpAmpWGS_v1.bed.txt	H5Nx primer bed file pulled from HSDH Freyja wo...
kraken2_standard_db	k2_standard_20240112.tar.gz	
mpx_barcodes	mpx_barcode.csv	https://github.com/andersen-lab/Freyja-barcodes/...
mpx_lineage_yml	mpox_lineages.yml	https://github.com/andersen-lab/Freyja-barcodes/...

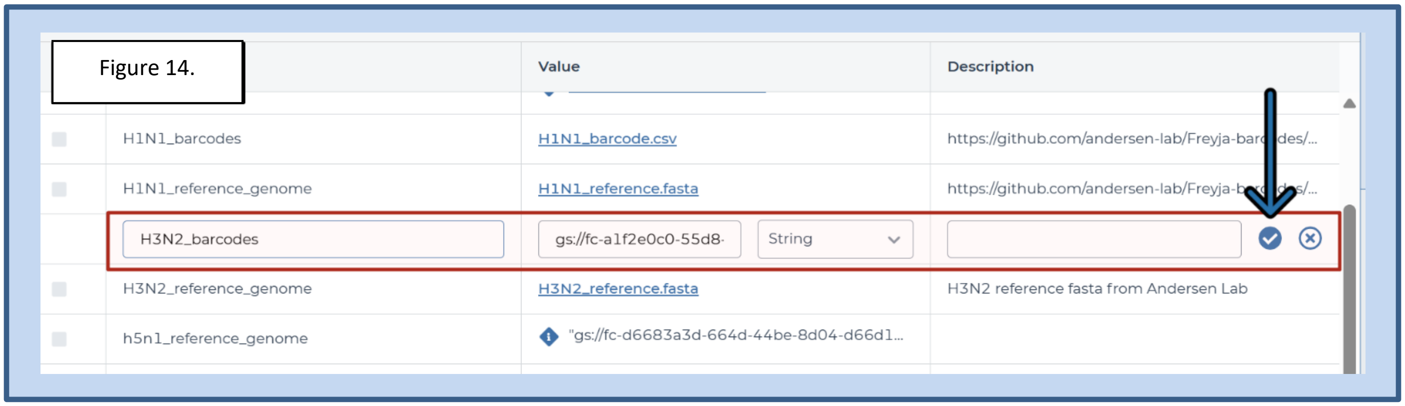


Figure 14.

	Value	Description
H1N1_barcodes	H1N1_barcode.csv	https://github.com/andersen-lab/Freyja-bar des/...
H1N1_reference_genome	H1N1_reference.fasta	https://github.com/andersen-lab/Freyja-b des/...
H3N2_barcodes	gs://fc-a1f2e0c0-55d8· String	
H3N2_reference_genome	H3N2_reference.fasta	H3N2 reference fasta from Andersen Lab
h5n1_reference_genome	"gs://fc-d6683a3d-664d-44be-8d04-d66d1...	