



Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow

Document TG-FREY-04, Version 2

Date:

4/5/2024

Workflow Versions:

PHB v2

1. PURPOSE/SCOPE

To standardize the process of pulling the most up-to-date reference files for curated_lineages and usher_barcodes files used as inputs to the Freyja FASTQ workflow for lineage deconvolution and assignment. These files will be copied into the Terra workspace as static files where Freyja_Update_PHB is run and can be saved for subsequent Freyja FASTQ analyses to allow samples to be analyzed using the same reference 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
- Theiagen's Freyja_Update_PHB workflow in Terra

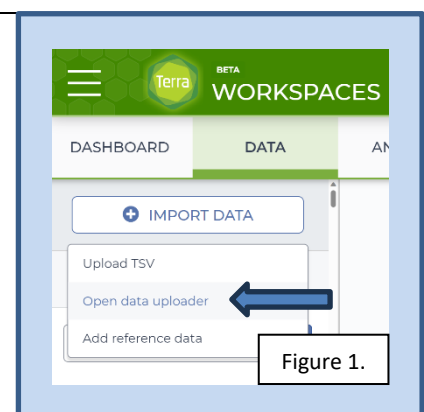
3. RELATED DOCUMENTS

Document Number	Document Name
TG-TER-03	Uploading Local or SRA NGS Data & Creating a Results Metadata Table in Terra
TG-FREY-01	Running SARS-CoV-2 Metagenomic Samples in Terra using Theiagen's Freyja FASTQ Workflow

4. PROCEDURE

4.1 CREATE A CLOUD LOCATION FOR FREYJA REFERENCE FILES

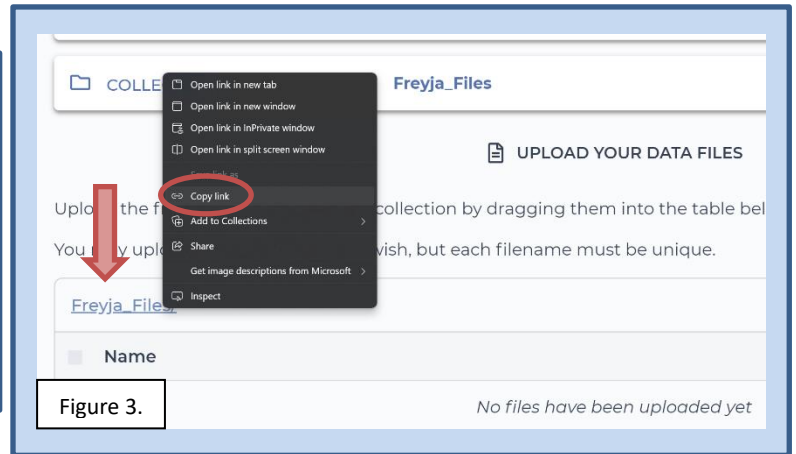
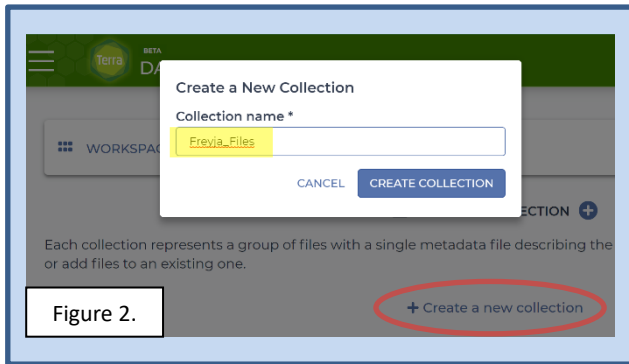
1. If a Freyja_Files folder has previously been created, but the filepath needs to be copied to enter into the Freyja_Update workflow inputs, in the Terra data tab, navigate to **Files > uploads/** and right click to **copy the FreyjaFiles/ folder filepath**
2. In the **data** tab of the Terra workspace containing metagenomic data, click **import data** and **open data uploader** (Fig 1)
3. Click **create a new collection** (Fig 2) – this creates a google bucket storage folder within the Terra workspace where files can be populated





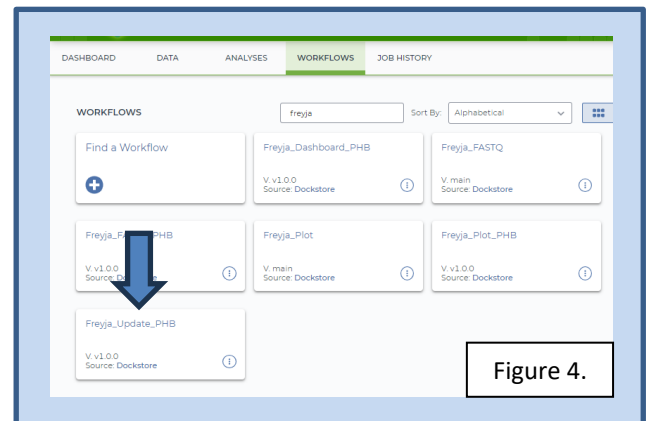
Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow	
Document TG-FREY-04, Version 2	
Date:	Workflow Versions:
4/5/2024	PHB v2

4. Name the collection `Freyja_Files` or something similar and click `create collection` (Fig 2)
5. Right click on the `Freyja_Files folder` and `copy link` (Fig 3); this will be pasted into a Freyja_Update workflow input field



4.2 RUNNING THE FREYJA UPDATE WORKFLOW

1. Navigate to the `workflows` tab in the workspace containing metagenomic SARS-CoV-2 data
2. Select the `Freyja_Update_PHB` workflow (Fig 4)
3. Choose the latest version of `version 2` in the version dropdown field (Fig 5, a)
4. Select the first bullet to `run workflow with inputs defined by file paths` (Fig 5, b)
5. `Uncheck call caching` (Fig 5)
6. In the `inputs` tab, `paste` in the first attribute field the link copied above in step 4.1, 4:



a. `"gs://fc-774455a0-292b-490f-9766-7b83bacbee59/uploads/Freyja_Files/"`

i. Remove the last `"/` from the pasted file path so that the first input value reads `"gs://fc-774455a0-292b-490f-9766-7b83bacbee59/uploads/Freyja_Files"`

7. Click `save`, `run analysis`, `add comments` if desired, and click `launch`
8. *This will populate files into the Terra workspace folder created in section 4.1 that can be used as Freyja reference files for the Freyja_FASTQ workflow to ensure continuity across Freyja analysis*



Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow

Document TG-FREY-04, Version 2

Date:

4/5/2024

Workflow Versions:

PHB v2

Freyja_Update_PHB

Version: ← a

Source: github.com/theiagen/public_health_bioinfo / `freyja_Update_PHB:v1.0.1`

Synopsis:

No documentation provided

Run workflow with inputs defined by file paths ← b

Run workflow(s) with inputs defined by data table

Use call caching ⓘ Delete intermediate outputs ⓘ Use reference disks ⓘ Retry with more memory ⓘ Ignore errors ⓘ

Figure 5.

SCRIPT ●● INPUTS ●● OUTPUTS ●● RUN ANALYSIS CANCEL SAVE

Hide optional inputs Download json | Drag or click to upload json | Clear inputs SEARCH INPUTS

Task name ↓	Variable	Type	Attribute
freyja_update	gcp_uri	String	"gs://fc-774455a0-292b-490f-9766-7b83bacbee59/uploads/Freyja_Files" [...]
freyja_update_refs	disk_size	Int	Optional [...]
freyja_update_refs	docker	String	Optional [...]

Figure 6.

4.3 ADDING/UPDATING WORKSPACE DATA ELEMENTS FOR REFERENCE FILES TO USE AS INPUTS FOR FREYJA_FASTQ_PHB ANALYSIS

1. In the Terra workspace, navigate to the `data tab` (Fig 7)
2. Scroll to the bottom of the left-hand sidebar and click `Files` (Fig 7)
3. Open the `uploads/` folder (Fig 7)
4. Open the `Freyja_Files/` folder created in [section 4.1](#) (Fig 8)
5. Open the most recent folder indicated by date, e.g. 2024-04-11 (Fig 9)
6. Hover the mouse over the `curated_lineages.json` file and click the clipboard to `copy the filepath` or right click on the file name and select copy link address (Fig 10)

Terra WORKSPACES Workspaces > theiagen-training-wc Data

DASHBOARD DATA ANALYSES WORKFLOWS JOB HIS

theiaeuk (1) theiaprok_illum... theiaprok_illum... (1) theiaprok_ont (2) ww_specimen (5) ww_specimen_set (3)

REFERENCE DATA

No references have been added.
Add reference data

OTHER DATA

Workspace Data

Files uploads/

Files

Files

FASTAs/

FreyjaValidationCriteria.txt

GenbankFiles/

LotTrackerResults.csv

ReferenceSeqs/

SD_PhoenixValidationCriteria.txt

ValidationCriteria/

qc_check_task/

submissions/

uploads/

Figure 7.



Creating Static Reference Files for Freyja Analysis in Terra using TheiaGen's Freyja Update Workflow

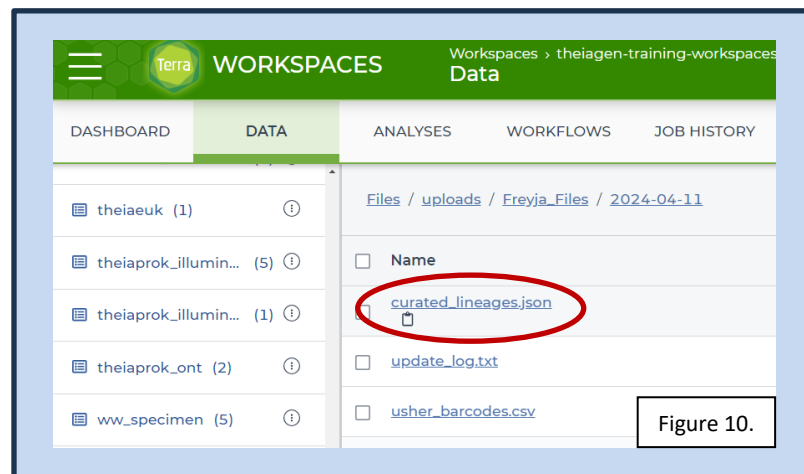
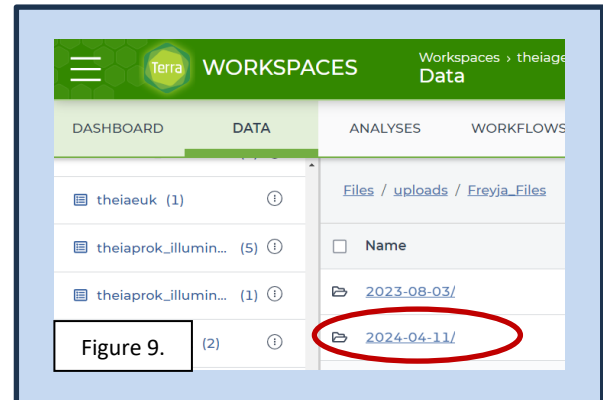
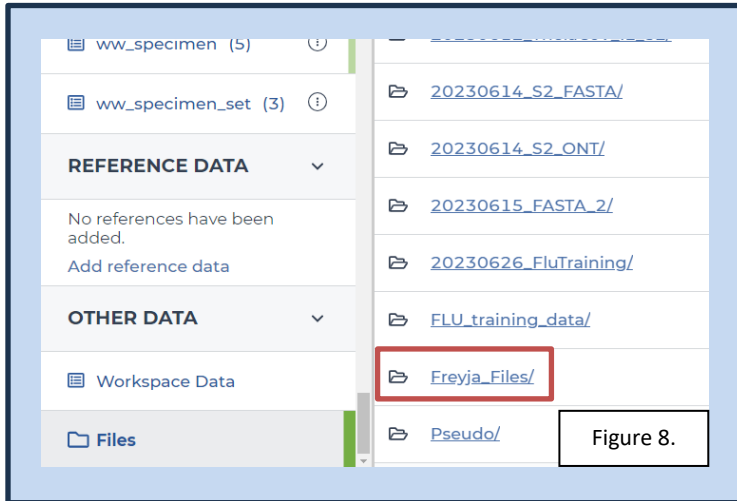
Document TG-FREY-04, Version 2

Date:

4/5/2024

Workflow Versions:

PHB v2



7. Open the **workspace data** tab in the bottom right of the left-hand sidebar (Fig 11, blue arrow)
8. If reference files have never been added as workspace data elements, click the **blue plus symbol** in the bottom right (Fig 11, red arrow)
 - a. Scroll to the bottom of the table where a new row will be populated
9. If reference files were previously added as workspace data elements, hover the mouse over the relevant entry and click the **pencil icon** that will appear to the right to edit the variable
10. Click in the **key field** and **name the element** being added (Fig 12)
 - a. For the Freyja curated_lineages file, it may be helpful to name the key **FreyjaLineageMetadata**
11. In the value field, choose **string** as the value type and paste the filepath copied above in step 6
 - a. No quotation marks need to be added to this string
12. In the description field, it may be helpful to specify **Updated YYYY-MM-DD, initials**



Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow	
Document TG-FREY-04, Version 2	
Date:	Workflow Versions:
4/5/2024	PHB v2

WORKSPACES | Workspaces > theiagen-training-workspaces/theiagen_kropp_sandbox > Data

Figure 11.

Key	Value	Description
Artic_V3_primer_bed	V3_nCoV_2019_primer.bed	
Artic_V4-1_primer_bed	V4-1_nCoV-2021_primer.bed	
Artic_V4_primer_bed	V4_nCoV_2021_primer.bed	
Midnight_primer_bed	Midnight_PrimerS_SARS-CoV-2.schem...	
SWIFT_primer_bed	SWIFT_SARS_CoV_2.scheme.bed	Updated 2023-07-05
freyja_dashboard_config	freyja_dash_config.json	Input 2024-07-18
kraken2_phoenix	k2_standard_080b_20230605.tar.gz	Updated by inés on 21/07/2023
nextclade_dataset_tag	2022-07-28/11200000/	Updated on 2022-08-12
nextclade_docker_image	nextstrain/nextclade:2.4.0	Updated on 2022-08-12
panini_docker_image	staphbp/panini4.1.2-pdata-1.1.4	Updated on 2022-08-12
vadr_docker_image	staphbp/vadr:1.4.2	Updated on 2022-07-28

←

→

ADD VARIABLE +

13. Repeat steps 2-12 to copy the most recent or desired usher_barcode file path and add or update a workspace data element for **FreyjaUsherBarcodes**
14. Proceed to run the Freyja_FASTQ_PHB workflow to (re-)analyze SC2 metagenomic samples
 - a. The corresponding workflow inputs for updated reference files will be as follows:
 - i. Freyja_lineage_metadata: **workspace.FreyjaLineageMetadata**
 - ii. Freyja_usher_barcode: **workspace.FreyjaUsherBarcodes**

WORKSPACES | Workspaces > theiagen-training-workspaces/theiagen_kropp_sandbox > Data

Figure 12.

Artic_V4-1_primer_bed	V4-1_nCoV-2021_primer.bed	
Artic_V4_primer_bed	V4_nCoV-2021_primer.bed	
<input type="text" value="FreyjaLineageMetadata"/>	gs://fc-58c86846-73: String	Updated 2024-4-11 kk ✓ ✕
FreyjaUsherBarcodes	usher_barcode.csv	Updated 2024-4-11 kk ↑
Freyja_ReferenceGenome	nCoV-2019.reference.fasta	MN908947.3



Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow

Document TG-FREY-04, Version 2

Date:

4/5/2024

Workflow Versions:

PHB v2

5. QUALITY RECORDS

- Freyja curated_lineages and usher_barcode reference files
- Workflow version and configuration settings (default and custom inputs)

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

None

8. REFERENCES

None

9. REVISION HISTORY

Revision	Version	Release Date
Document creation	1	8/2023
Updates for PHB v2.0.0 release (no major functionality changes), clarified process to re-use reference files (added section 4.3), updated quality records, added appendix 10.1	2	4/2024



Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow

Document TG-FREY-04, Version 2

Date:

Workflow Versions:

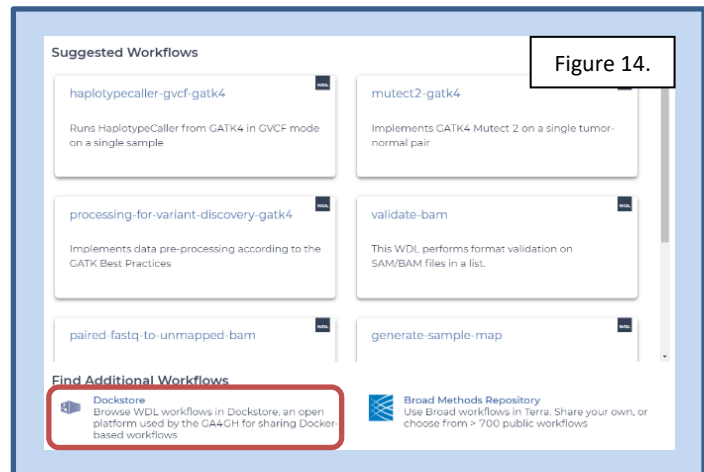
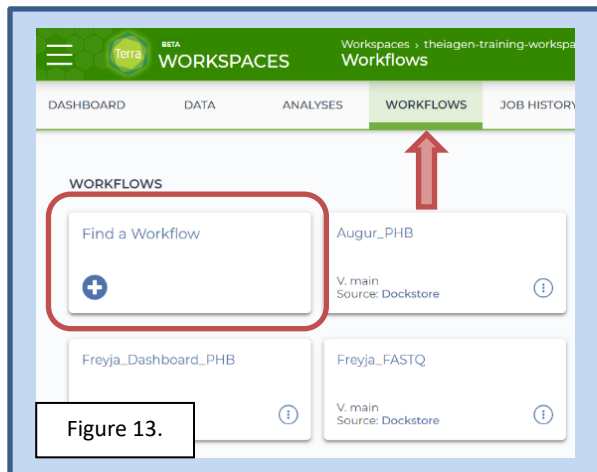
4/5/2024

PHB v2

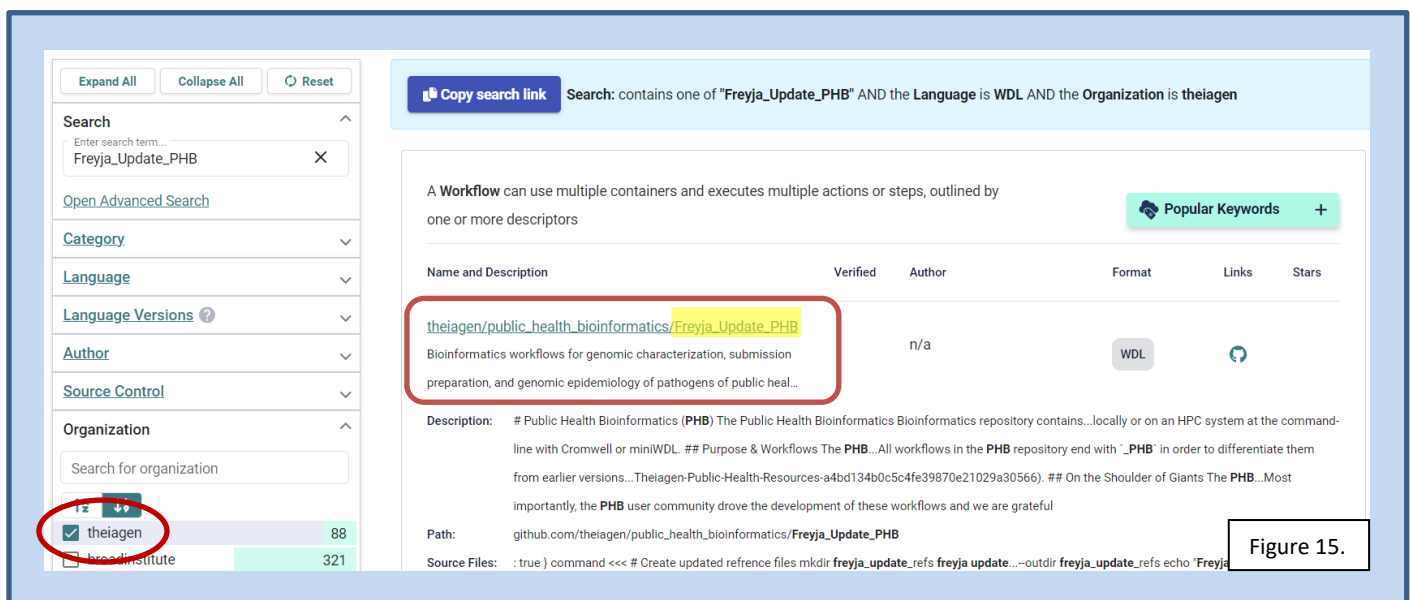
10. APPENDICES


10.1 IMPORTING FREYJA_UPDATE_PHB WORKFLOW FROM DOCKSTORE

1. In the **Terra workspace** of interest, open the **workflows** tab and click **find a workflow** (Fig 13)
2. In the pop-up window, click **dockstore** in the bottom left (Fig 14)

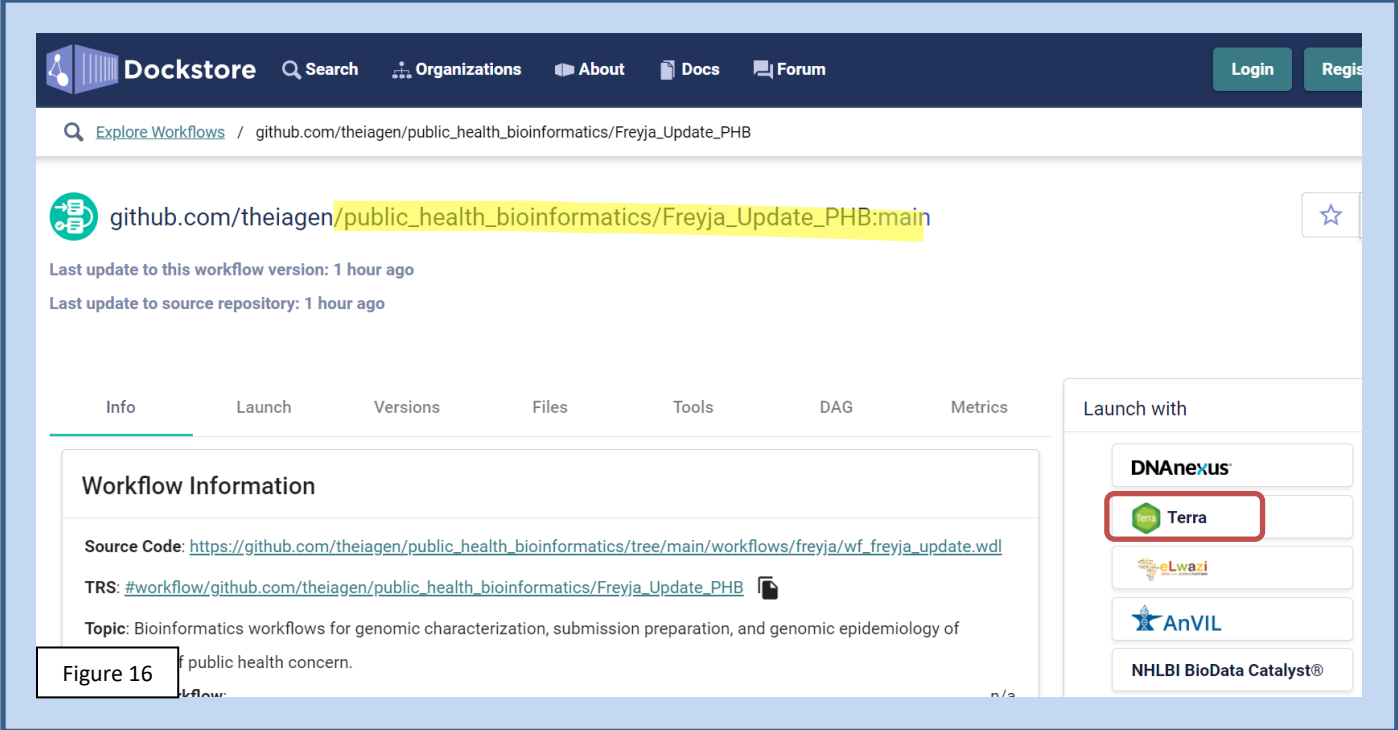


3. Type **"Freyja_Update_PHB"** in the search bar (Fig 15)
4. In the left-hand sidebar, scroll down to "Organization" and select **"theiagen"** (Fig 15)
5. Find the workflow by looking at the file path suffix; click the name to **open the workflow** (Fig 15)



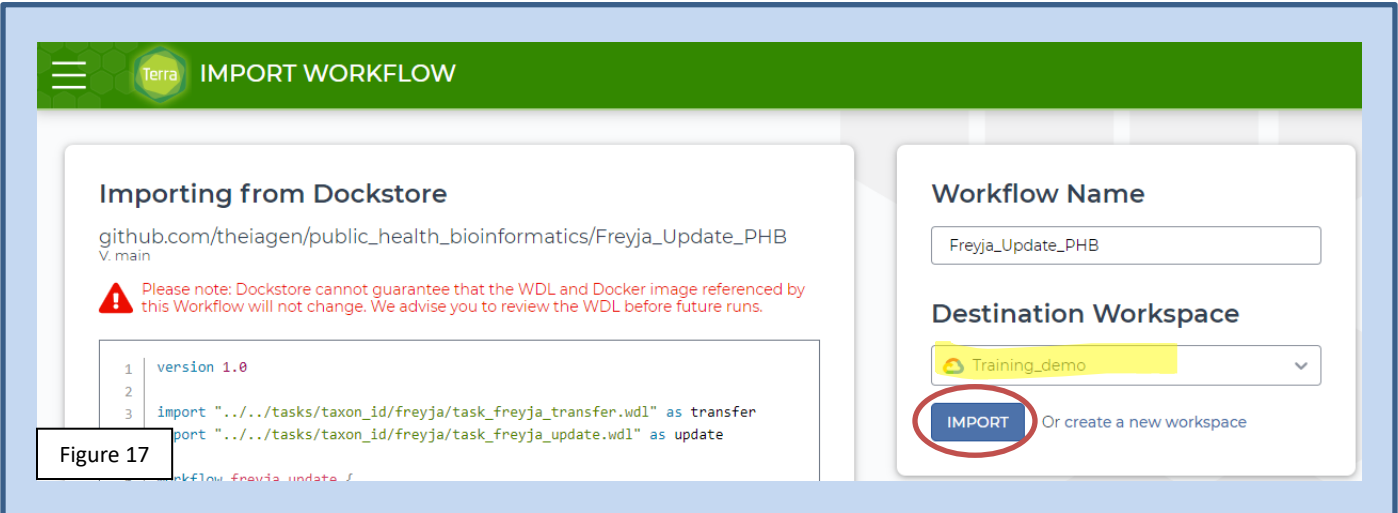
	Creating Static Reference Files for Freyja Analysis in Terra using Theiagen's Freyja Update Workflow	
	Document TG-FREY-04, Version 2	
	Date:	Workflow Versions:
	4/5/2024	PHB v2

6. Click **Terra** to launch the workflow in Terra (Fig 16)



The screenshot shows the Dockstore interface for a workflow. The workflow name is `github.com/theiagen/public_health_bioinformatics/Freyja_Update_PHB:main`. The 'Launch with' section on the right contains several options: DNAexus, Terra (highlighted with a red box), eLwazi, AnVIL, and NHLBI BioData Catalyst. The 'Workflow Information' section on the left provides details about the source code, TRS, and topic.

7. Choose the **destination workspace** in the dropdown and click **import** or create a new workspace (Fig 17)



The screenshot shows the Terra 'IMPORT WORKFLOW' interface. The workflow name is `Freyja_Update_PHB`. The 'Destination Workspace' dropdown is set to `Training_demo`. The 'IMPORT' button is highlighted with a red box. The interface also displays a warning message and a code snippet for the workflow definition.