	Creating a Dashboard Visualization of SARS-CoV-2 Metagenomic Samples in Terra using Theiagen's Freyja Dashboard Workflow	
	Document TG-FREY-03, Version 2	
	Date:	Workflow Versions:
	4/16/2024	PHB v2

---

1. PURPOSE/SCOPE

To standardize the process of aggregating Freyja sample data to generate an interactive visualization using Theiagen's Freyja Dashboard workflow in Terra. Acceptable data types include Freyja demixed output files generated from the Freyja FASTQ workflow.

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
- Freyja demixed files uploaded to Terra workspace
- Theiagen's Freyja_Dashboard_PHB workflow in Terra

REQUIRED WORKFLOW INPUTS FILES
- Freyja demixed files (output from Freyja FASTQ
- Dashboard name
- Collection dates
- Viral load
- *Optional: freyja_dashboard_config json 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-01	Running SARS-CoV-2 Metagenomic Samples in Terra using Theiagen's Freyja FASTQ Workflow
TG-FREY-02	Plotting SARS-CoV-2 Metagenomic Sample Data in Terra using Theiagen's Freyja Plot Workflow

|	**Creating a Dashboard Visualization of SARS-CoV-2 Metagenomic Samples in Terra using Theiagen's Freyja Dashboard Workflow**	
	Document TG-FREY-03, Version 2	
	Date:	Workflow Versions:
	4/16/2024	PHB v2

4. PROCEDURE

4.1 RUNNING THE FREYJA DASHBOARD WORKFLOW

1. Open Terra and navigate to the *workflows* tab within the workspace containing metagenomic SARS-CoV-2 data
2. Select the *Freyja_Dashboard_PHB* workflow (Fig 1)
3. Choose the latest version of *version 2* in the version dropdown field, or the version used during internal assay validation (Fig 2, a)

Figure 1.
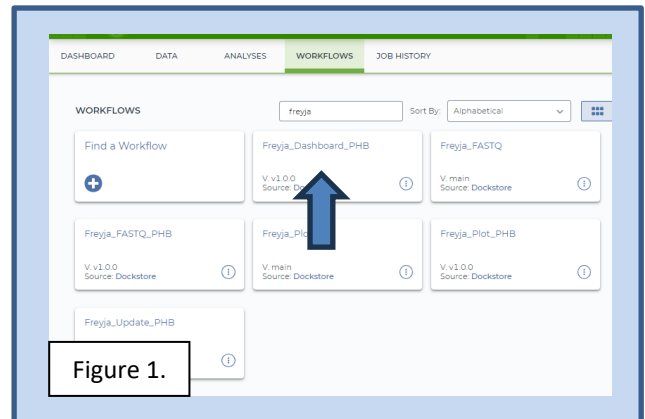
Freyja_Dashboard_PHB

Version: v2.0.0

Source: github.com/theiagen/public_health_bioinformatics/Freyja_Dashboard_PHB:v1.0.0

Synopsis:

No documentation provided

○ Run workflow with inputs defined by file paths

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

a

b

c → ww_specimen_set
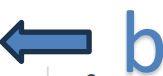
d ← SELECT DATA

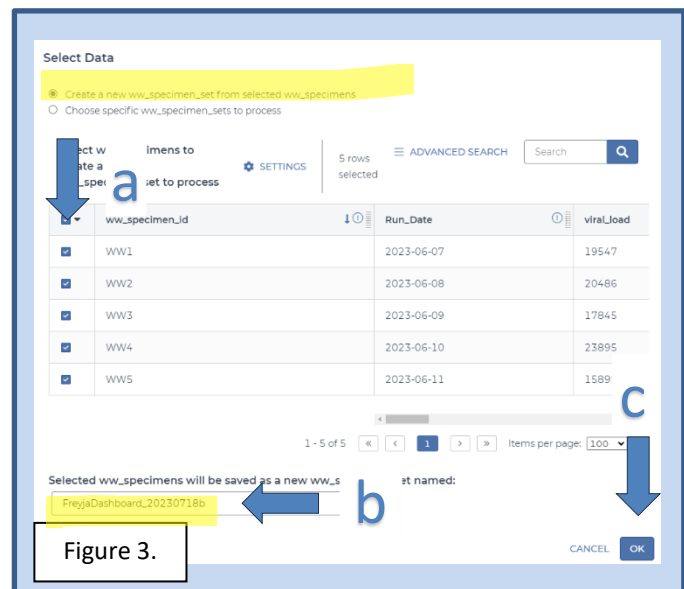☐ Use call caching ⓘ   ☐ Delete intermediate outputs ⓘ   ☐ Use reference disks ⓘ   ☐ Retry with more memory ⓘ   ☐ Ignore   Figure 2.

4. Select the first bullet to *run workflow with inputs defined by file paths* (Fig 2, b)
5. Select the relevant SET data table name under the select *root entity type* dropdown (Fig 2, c)
6. Click *select data* (Fig 2, d)
7. In the pop-up window select the first bullet to *create a new ww_specimen_set from selected ww_specimens* (Fig 3, highlighted text)
   a. *Select the checkbox* for each sample to be included in the analysis (Fig 3, a)

Select Data

● Create a new ww_specimen_set from selected ww_specimens

○ Choose specific ww_specimen_sets to process

	ww_specimen_id	Run_Date	viral_load
☑	WW1	2023-06-07	19547
☑	WW2	2023-06-08	20486
☑	WW3	2023-06-09	17845
☑	WW4	2023-06-10	23895
☑	WW5	2023-06-11	1589

a

c

1 - 5 of 5

Selected ww_specimens will be saved as a new ww_s   et named:

FreyjaDashboard_20230718b

b

Figure 3.

b. _Optional: name the output set name to differentiate it from other workflow runs, e.g. FreyjaDashboard_YYYYMMDDn (Fig 3, b); this populates a row to the SET data table_

c. Click ok

Figure 4.

8. In the inputs tab, set the first five attributes to the following (Fig 5, yellow highlights):

   a. _this.ww_specimens.Run_Date_

      i. _Notice an "s" is added to the end of the data table name: ww_specimen -> ww_specimens_

      ii. _"Run_Date" must match the sample data table column containing collection dates_

   b. _"Default_Dashboard"_; enter the dashboard name you want to display in quotes (Fig 8)

   c. _this.ww_specimens.freyja_demixed_

   d. _this.ww_specimens.ww_specimen_id_

   e. _this.ww_specimens.viral_load_

      i. _"viral_load" must match the corresponding sample data table column header_

9. _Optional: configure an additional input setting to create lineage groups if there are too many lineages to be visualized together (Fig 4, orange highlight)_

   a. _In the config variable, add "gs://theiagen-public-files-rp/terra/freyja-files/freyja_dash_config.json" to group lineage according to Figure 5; otherwise, create a customized json file specifying how to define lineage groups_

10. Specify outputs by clicking on the outputs tab and use defaults (Fig 6)

11. Click save, run analysis, add comments if desired, and click launch

Figure 5.

Figure 6.

Figure 7.

4.2 **ANALYZING DASHBOARDS**

1. Navigate to the data tab of the workspace containing metagenomic data and open the pertinent SET data table

2. Open the file located in the freyja_dashboard column for the relevant data set (Fig 7, red oval)
   a. *If the sample set name was specified (section 4.1, step 6b) the data set name will be indicated in the ww_specimen_set_id column (Fig 7, highlight)*

3. *Different dashboard visualizations are available by clicking on different blue boxes above the graph (Fig 8)*

Figure 8.

4. The variants visualization (Fig 8) displays the variant prevalence (y-axis) by collection date (x-axis)

5. In the lineages view (Fig 9), the lineage group prevalence by collection date is shown
   a. *Lineage groups are displayed according to the group definitions provided by the config variable in section 4.1, step 8; modify groups by changing definitions in the json file and re-running the dashboard workflow*

6. Viral load by collection date may be viewed by selecting the viral load tab of the graph (Fig 10)

7. Scaled variants (Fig 11) and scaled lineages (Fig 12) tabs depict the variant or lineage abundance viral loads by collection date
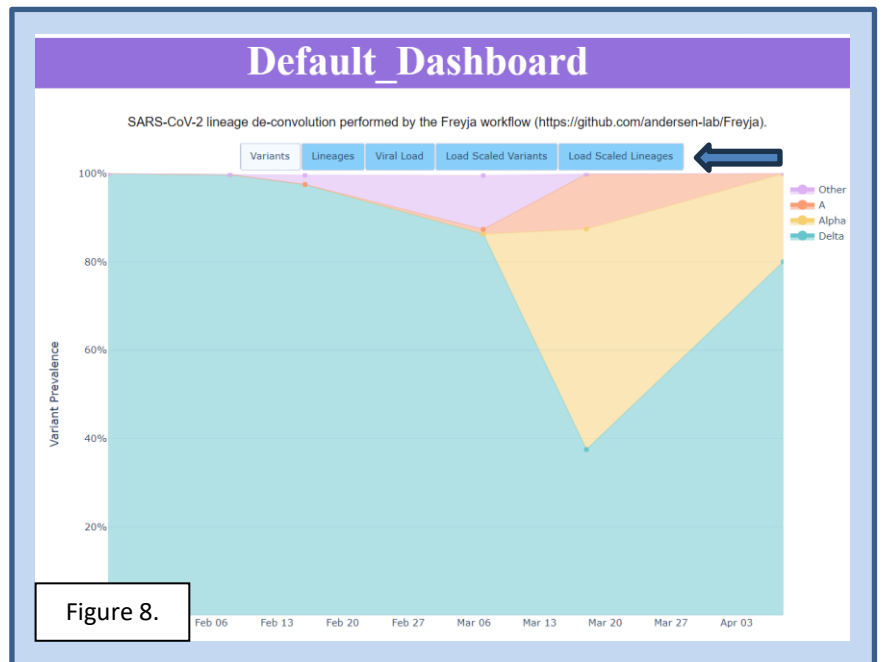
Figure 9.

Figure 10.

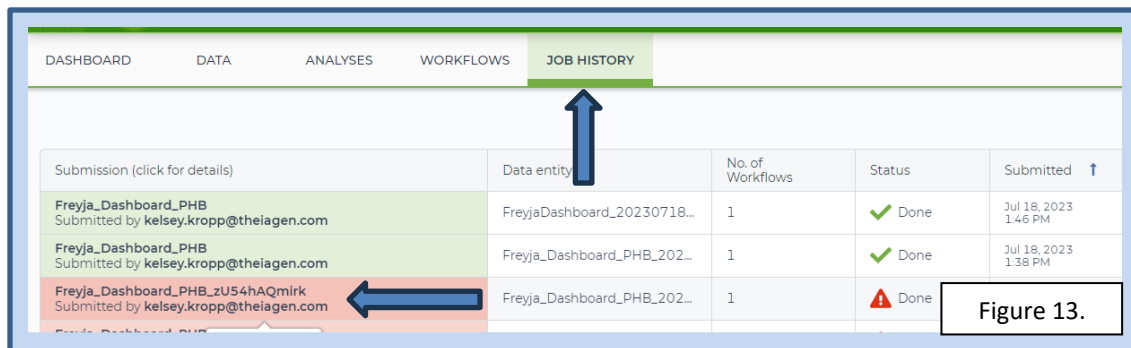Figure 11.

Figure 12.
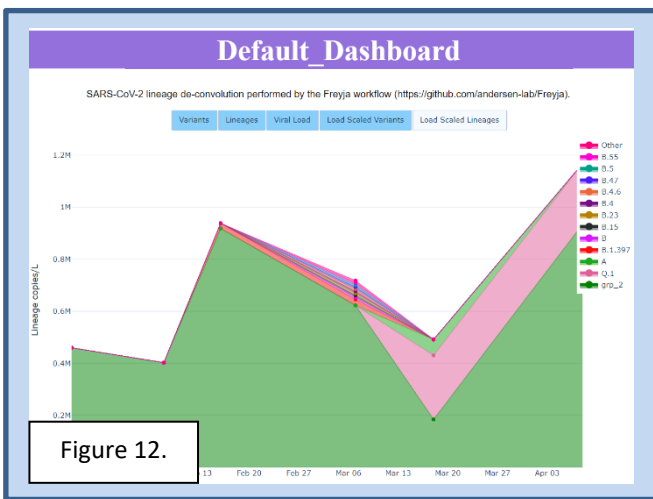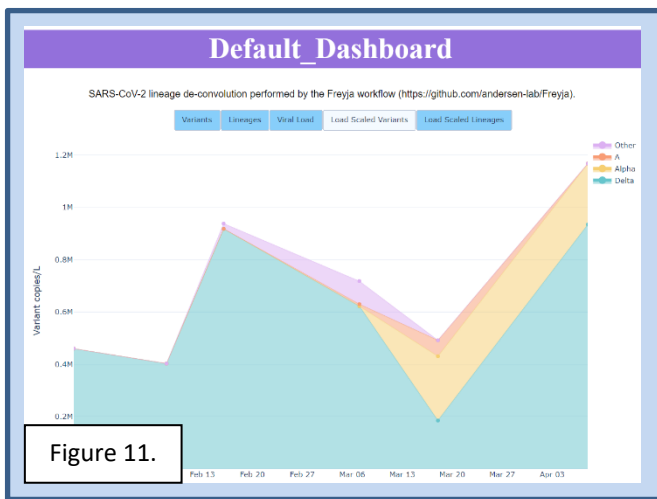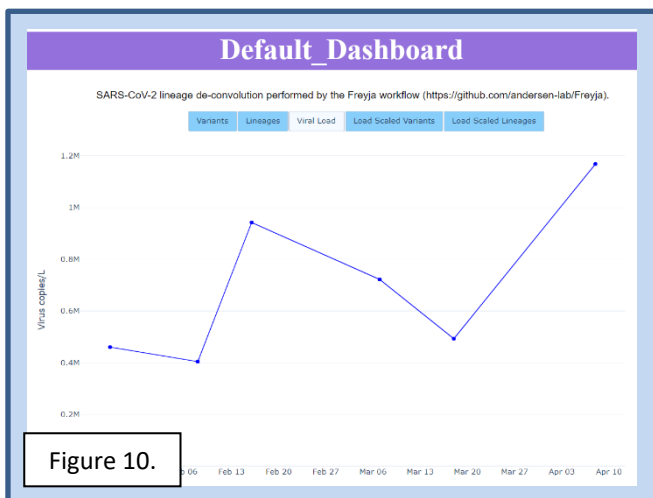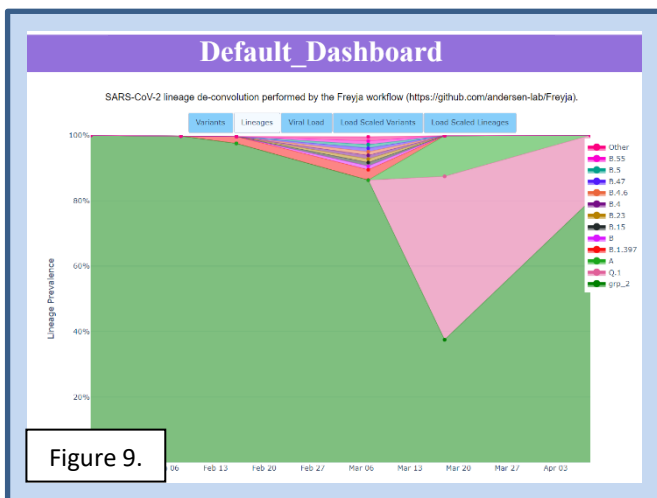
5. QUALITY RECORDS

- freyja_demixed files
- Dashboard visualizations

6. TROUBLESHOOTING

1. Identify if failed run is due to "too many lineages to show" error
   a. In *job history*, click on the *failed run* for freyja dashboard (Fig 13)
   b. Open *job manager* in the links column within the table by clicking on the three lines (Fig 14)
   c. Click *sign in* to authenticate sign in; open the *log files* indicated by the cloud on paper (Fig 15)
   d. Within the code, use the search function (*Ctrl+F* or *Cmd+M*) for "*lineage*;" if the error code indicates "*too many lineages to show*," *re-run the workflo*w with a configuration file included in the workflow input settings to simplify the number of lineage groups to display in the visualization
      i. See section 4.1 step 8 for details.
2. Consult with internal staff familiar with this procedure or contact support@theiagen.com for other troubleshooting inquiries
3. For document edit requests, contact support@theiagen.com

Figure 13.

Figure 14.

Figure 15.

Figure 16.

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 can only be used to analyze SC2 data from Illumina PE sequencing files

8. REFERENCES

None

9. REVISION HISTORY

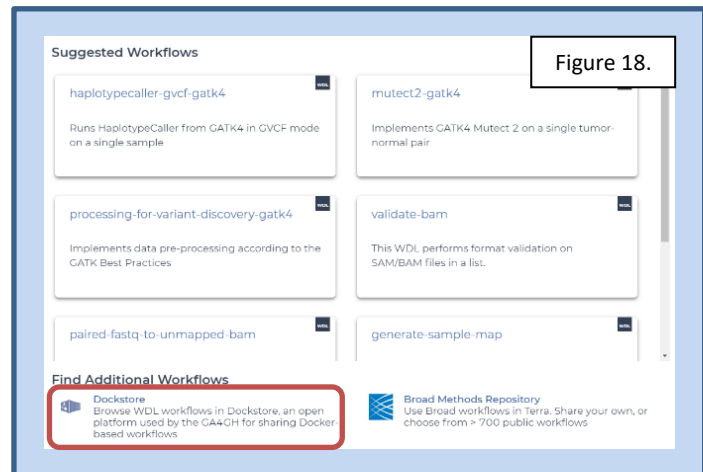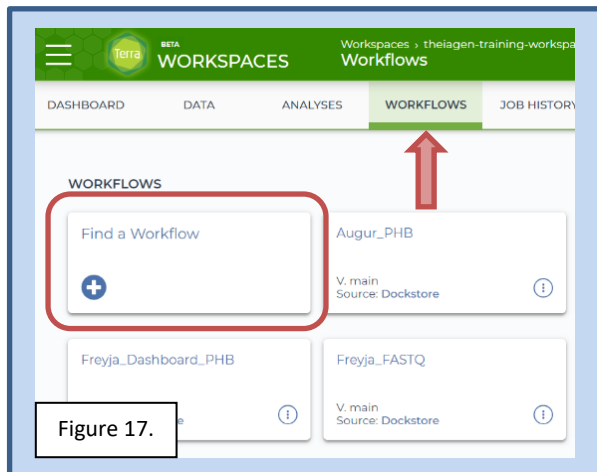| Revision	Version	Release Date
Document creation	1	8/2023
Minor document revisions for PHB v2.0.0 release; added appendix	2	4/2024

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 17)
2. In the pop-up window, click *dockstore* (Fig 18)

Figure 17.

Figure 18.

3. Type "*Freyja_Dashboard_PHB*" in the search bar (Fig 19)
4. In the left hand sidebar, scroll down to Organization and select "*theiagen*" (Fig 19)
5. Find the workflow by looking at the file path suffix; click the name to *open the workflow* (Fig 19)
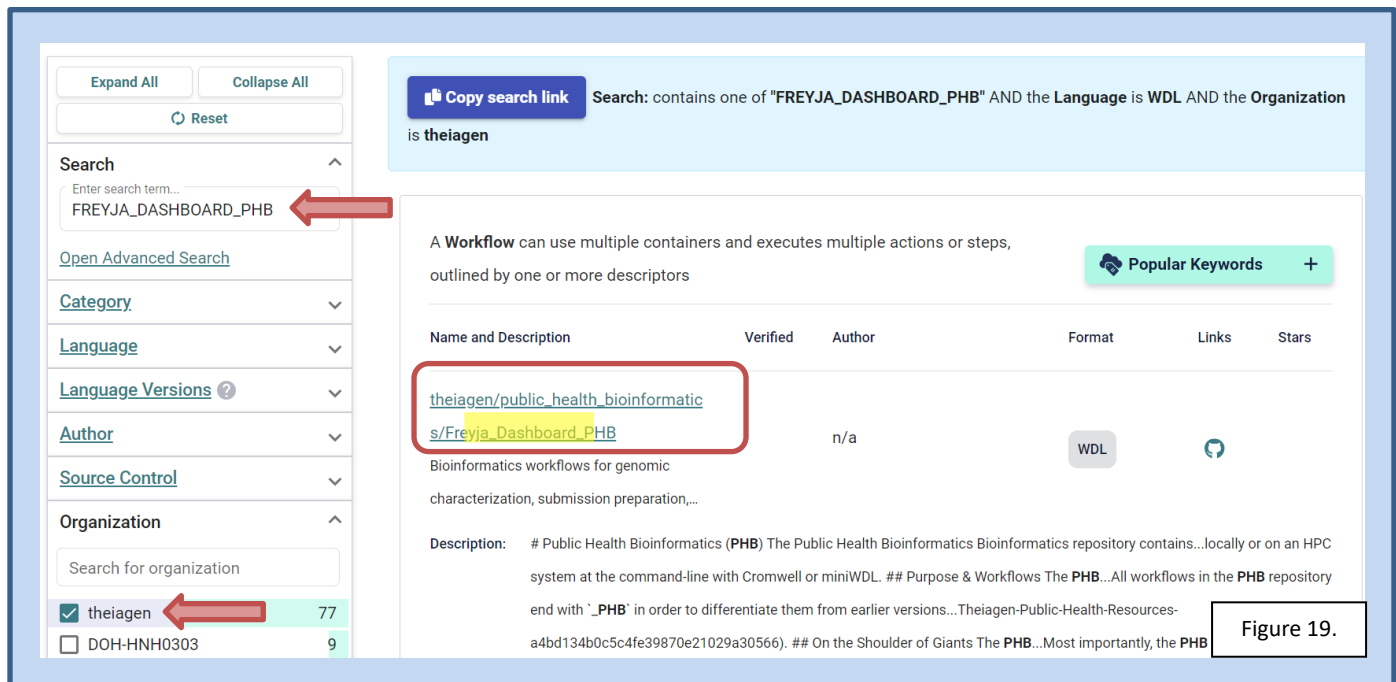
Figure 19.

6. Click *Terra* to launch the workflow in Terra (Fig 20)
7. Choose the *destination workspace* in the dropdown and click *import* or create a new workspace (Fig 21)

Figure 20

Figure 21