Document TG-FREY-01, Version 2

Date: 4/4/2024 Workflow Versions: 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
TC TER 02	Uploading Local or SRA NGS Data & Creating a
IG-IER-03	Results Metadata Table in Terra
	Creating Static Reference Files for Freyja
TG-FREY-04	Analysis in Terra using Theiagen's Freyja Update
	Workflow

4. PROCEDURE

4.1 RUNNING THE FREYJA FASTQ WORKFLOW

- 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)



*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



Document TG-FREY-01, Version 2

Date:	Workflow Versions:
4/4/2024	PHB v2

- 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)



- 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

ec Ce	t ww_specimens to ss	SETTINGS	4 rows selected	■ ADVANCED SEARCH	Sean	ch Q
•	ww_spe 10	aligned_bai	0	aligned_bam	0	alignment_method
	WWl	WW1.primertrim.sorted.t	<u>bam.bai</u>	WW1.primertrim.sorted.bam		BWA Version: 0.7.17-r1188; iVi
	WW2	WW2.primertrim.sorted.	<u>pam.bai</u>	WW2.primertrim.sorted.bam		BWA Version: 0.7.17-r1188; iV
	WW3	WW3.primertrim.sorted.t	oam.bai	WW3.primertrim.sorted.bam		BWA Version: 0.7.17-r1188; iV
	WW4	WW4.primertrim.sorted.	oam.bai	WW4.primertrim.sorted.bam		BWA Version: 0.7.17-r1188; iVi
	WW5					
	WW4 WW5	WW4.primertrim.sorted.	<u>bam.bai</u>	WW4, primertrim sorted barn		BWA Version: 0.7.17-r118

- 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)



Document TG-FREY-01, Version 2

Date:Workflow Versions:4/4/2024PHB v2

- 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
 - Find other SC2 primer bed files available on the Theiagen <u>Docker Image and Reference</u> <u>Materials for SARS-CoV-2 Genomic Characterization Notion page</u> 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*

SCRIPT •• INPUTS •	OUTPUTS •• RUN ANALYSIS		
Hide optional inputs			\checkmark
Task name 🖡	Variable	Туре	Attribute
freyja_fastq	primer_bed	File	workspace.SWIFT_primer_bed
freyja_fastq	read1_raw	File	thisreadl
freyja_fastq	read2_raw	File	thisread2
freyja_fastq	reference_genome	File	"gs://theiagen-public-files/terra/freyja-files/hCoV-2019.reference.fasta"
freyja_fastq	samplename	String	thisww_specimen_id
Figure 4.	cpu	Int	Optional

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

eps	FIOAT	Optional
freyja_lineage_metadata	File	"gs://theiagen-public-files/terra/freyja-files/curated_lineages_2023-08-28.json"
freyja_usher_barcodes	File	"gs://theiagen-public-files/terra/freyja-files/usher_barcodes_2023-08-28.csv"
Fill out top box or b	ottom box, no	t both; files indicated above are for example only
update_db	Boolean	true Figure 5.



Document TG-FREY-01, Version 2

Date:Workflow Versions:4/4/2024PHB v2

- i.<u>Option 1</u>: 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 developers¹, 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 <u>TG-FREY-04: Freyja_Update</u> SOP for details
- ii.<u>Option 2</u>: 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.<u>Option 3</u>: To use specific versions of the curated_lineages and usher_barcodes files that have been used previously for analysis, enter the corresponding <u>"gs://[FILENAME]..."</u> filepaths for <u>freyja_lineage_metadata</u> and <u>freyja_usher_barcodes</u> 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

SCRIPT •• IN	NPUTS ••	OUTPUTS •• RUN A	NALYSIS	
Output files will be saved to Tiles / submission unique ID /	[/] freyja_fastq / wor			
References to outputs will be wri Tables / ww_specimen Fill in the attributes below to add	itten to I or update colum	ns in your data table		
Task name ↓		Variable	Туре	Attribut Use defaults
Task name ↓ freyja_fastq		Variable aligned_bai	Type File	Attribut Use defaults this aligned_bai
Task name ↓ freyja_fastq		Variable aligned_bai aligned_bam	Type File File	Attribut Use defaults thisaligned_bai thisaligned_bam



Document TG-FREY-01, Version 2

PHB v2

DETERMINING LINEAGES, ABUNDANCES, AND COVERAGE METRICS 4.2

- 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:
 - freyja_ barcode_version a.
 - freyja_demixed b.
 - freyja_depths c.
 - d. freyja metadata version
 - e. freyja variants

	CES	Workspaces > th Data	eiagen-training-workspaces/Theiage	n_Kropp_Sandbox >		
				Select columns		
DASHBOARD DATA	ANAL	YSES WORKFL	OWS JOB HISTORY	Show: al none	Sort: alphabetical	
🗉 TheiaCoV_IIIumi 🌾 🚺 🔺	A 500			📖 🗌 freyja_bootstrap_lineages	^	SAVE THIS COLUMN SELECTION
	/ EDI	I 🕺 OPEN WITH	B EXPORT & SETTINGS	🚥 🗌 freyja_bootstrap_lineages_pdf		
	□ -	ww_spe 1	freyja_demixed	IIII freyja_bootstrap_summary		Your saved column selections:
🗐 illumina_pe_sp (2				I freyja_bootstrap_summary_pdf		FreyjaFASTQ (i)
		WW1	WW1_freyja_demixed.tsv	m ✓ freyja_demixed		
		WW2	WW2_freyja_demixed.tsv	Intervia_fasto_wf_analysis_date		
🗐 kilifi_H3N2 (39) 🕕	_	14/14/7	MART free in descined by	I freyja_fastq_wf_version		
🖩 kilifi H3N2 set (5) 🕕		VVVV3	wws_freyja_demixed.tsv	🚥 🗹 freyja_metadata_version		
		WW4	WW4_freyja_demixed.tsv	🚥 🗹 freyja_variants		
🔳 kleb_training (23) 🔅		WW5	WW5 frevia demixed tsv	🔤 🗹 freyja_version		
kleb training set (5)			www.jneyja_demixed.tsv	📟 🗹 freyja_barcode_version		
				ivar_version_primtrim	-	
🔲 p_aeruginosa_v (35) 🕕						
Figure 7						CANCEL

- 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 *freyig depths column file* to determine the relative depth of coverage for every variant identified



Document TG-FREY-01, Version 2

Date:	
4/4/2024	

Workflow Versions: PHB v2

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 <u>support@theiagen.com</u> for troubleshooting inquiries
- For document edit requests, contact <u>support@theiagen.com</u>

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

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

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

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)

	Workspaces > theiagen-training-workspa Workflows	Suggested Workflows	Figure 9.
DASHBOARD DATA ANAL	LYSES WORKFLOWS JOB HISTORY	haplotypecaller-gvcf-gatk4 Runs HaplotypeCaller from CATK4 in GVCF mode on a single sample	mutect2-gatk4 Implements GATK4 Mutect 2 on a single tumor- normal pair
Find a Workflow	Augur_PHB V. main Source: Dockstore	processing-for-variant-discovery-gatk4 Implements data pre-processing according to the GATK Best Practices	validate-bam
Freyja_Dashboard_PHB	Freyja_FASTQ	paired-fastq-to-unmapped-barm	generate-sample-map
Figure 8.	V. main Source: Dockstore	Dockstore Dockstore Browse WDL workflows in Dockstore, an open platform used by the CA4GH for sharing Docker- based workflows	Broad Methods Repository Use Broad workflows in Terra, Share your own, or choose from > 700 public workflows

- 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)

Expand All Collaps	se All	Copy copy	seh link Search: oo	atains one of " Frovi	- EASTO PHP" AND th) the Organiza t	lon is
♥ Reset		theiagen	Search. co	italiis one or riey				
Search Enter search term Freyja_FASTQ_PHB								
Open Advanced Search	A Workflow outlined by o	can use multiple cont one or more descripto	ainers and execute	s multiple actions or s	teps, 🧄 🍖 Pop	oular Keywords	s +	
<u>Category</u>	~							
Language	~	Name and De	scription	Verified	Author	Format	Links	Stars
Language Versions 🕜	~	theiagen/pu	blic_health_bioinform	atic				
Author	~	<u>s/Fr<mark>eyja_FA</mark></u>	<u>STQ_PHB</u>		n/a	WDL	0	
Source Control	~	Bioinformatic	s workflows for genomic on, submission preparatio	on,				
Organization	^			,				
Search for organization		Description:	# Public Health Bioinfo system at the comman	rmatics (PHB) The Pul d-line with Cromwell o	blic Health Bioinformatics r miniWDL. ## Purpose &	Bioinformatics repository co Workflows The PHB All wor	ntainslocally of kflows in the PH I	^r on an HPC 3 repository
1 <u>₽</u> ↓ <u>₽</u>			end with `_ PHB ` in orde	er to differentiate them	from earlier versionsThe	eiagen-Public-Health-Resour	ces-	
🗸 theiagen 🧹	105		a4bd134b0c5c4fe3987	'0e21029a30566). ##	On the Shoulder of Giants	The PHB Most importantly,	the PHB	
broadinstitute	107		drove the development	- f also	al		FI	sure 10.



Document TG-FREY-01, Version 2

Date:	Workflow Versions:	
4/4/2024	PHB v2	

- 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 Workflor	ws / github.com/theia	gen/public_health_bioinfo	matics/Freyja_FASTQ_	РНВ			
github.cc	or <mark>theiagen/pul</mark> orkflow version: 23 ho e repository: 18 hours	olic_health_bioinf urs ago ago	ormatics/Freyj	a_FASTQ_PHB:n	nain		\$
Info	Launch	Versions	Files	Tools	DAG	Metrics	Launch with
Workflow In	formation						DNAnexus [®]
	tps://github.com/theia	gen/public_health_bioinf	ormatics/tree/main/v	<u>vorkflows/freyja/wf_fre</u>	<u>yja_fastq.wdl</u>		eLwazi
Source Code: htt TRS: <u>#workflow</u>	/github.com/theiagen/	public_health_bioinforma	ubmission preparatio	n and genomic epider	niology of pathogens	of public health	X AnVIL

Importing from Dockstore	Workflow Name
github.com/theiagen/public_health_bioinformatics/Freyja_FASTQ_PHB _{V. main}	Freyja_FASTQ_PHB
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.	Destination Workspace
<pre>1 version 1.0 2 3 import "//tasks/alignment/task_bwa.wdl" as align 4 taport "//tasks/quality control/read filtering/task ivar primer trim.wdl" a</pre>	Training_demo
gure 12 port "//tasks/task_versioning.wdl" as versioning port "//tasks/taxon id/frevja/task frevja.wdl" as frevja task	



Document TG-FREY-01, Version 2

Date: 4/4/2024

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)

ASHBOARD DATA	ANALYSES WORKFLOWS JO	BHISTORY	0
🛿 p_aeruginosa_va (35) 🛈	*	Download TSV Drag or click to up	aload TSV Search
p_aeruginosa_vali (1) ①	Koy	Value	Description
p seruginosa val (1) (1)		Y3_nLov_ZUI3.primer.bed	
g theiaprok_illumin_ (5) ①	Artic_V4-1_primer_bed	V4-1_nCoV-2021.primer.bed	
theiaprok illumin. (1) (1)	ArLic_V4_primer_bed	V4_nCoV-2021.primer.bed	
www.specimen (5)	Midnight_primer_bed	Midnight_Primers_SARS-CoV-2.scheme_	
	SWIFT_primer_bed	SWIFT_SARS_CoV_2.schome.bed	Updated 2023-07-05
ww_specimen_set (2) (7	freyja_clashboard_config	frevia_dash_config.ison	Input 2023-07-18
EFERENCE DATA V	kraken2 phoenix	k2_standard_08gb_20230605.tar.gz	Updated by Ines on 21/07/2023
o references have been ddwd. dd reference data	nextclade_dataset_tag	2022-07-26112:00:00Z	Updated on 2022-08-12
THER DATA ~	nextclade docker image	nextstrain/nextclade:2.4.0	Updated on 2022-08-12
Workspace Data	pangolin_docker_image	staphb/pangolin:4.1.2-pdata-1.1.6	Updated of 2022-08-12
	warte chuckeer inviacea	stardth/vartr1 6.3	ADD VARIABLE +

Кеу	Value	Description Figure 14.		
Artic_V4-1_primer_bed	V4-1_nCoV-2021,primer.bed			
Artic_V4_primer_bed	V4_nCoV-2021.primer.bed			
FreyjaLineageMetadata	curated_lineages.json	Taken from Freyja_Workflows Demo Data		
FreyjaUsherBarcodes	usher_barcodes.csv	Updated 8/3/23; taken from Freyja_Workflows Demo Data		
Freyja_ReferenceGenome	nCoV-2019.reference.fasta	MN908947.3		
Midnight_primer_bed	Midnight_Primers_SARS-CoV-2.scheme.bed			
SWIFT_primer_bed	gs://theiagen-public-files/terra/theia	Updated 2023-07-05 kk 🥑 🛞		