

Analyzing SARS-CoV-2 Da	ata in Terra using Theiagen's
TheiaCoV Clea	arLabs Workflow
Document TG-	SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

1. PURPOSE/SCOPE

To standardize the process of analyzing SARS-COV-2 (SC2) next generation sequencing (NGS) data using Theiagen's TheiaCoV_ClearLabs_PHB workflow in Terra to generate assemblies, quality control (QC) metrics, and determine Nextclade clade and Pangolin lineage assignments. Acceptable data types include ClearLabs raw read file format. Read the documentation <u>here</u>.

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
- ClearLabs raw sequencing read files uploaded to Terra workspace, see TG-TER-03
- Theiagen's TheiaCoV_ClearLabs_PHB workflow in Terra, see appendix 10.1

3. RELATED DOCUMENTS

IMPORTANT NOTES

- Metadata column headers and workflow input text indicated in gray in this SOP are customizable; black is required text
- Terra data table column headers become available as workflow inputs when running workflows, search for them in workflow input dropdowns using the prefix this. to filter
- Filter for workspace data and files in workflow input dropdowns using the prefix workspace.

Document Number	Document Name
TG-TER-03	Getting Started in Terra: Importing Reads, Metadata, Workflows, and More

4. PROCEDURE

4.1 RUNNING THE THEIACOV WORKFLOW

 Open Terra and navigate to the workflows tab within the workspace containing SC2 data (Fig 1)

2.	Select the
	TheiaCoV_ClearLabs_PHB
	(Fig 1)

DASHBOARD	DATA	ANALYSES	WORKFLOWS	JOB HISTOP	RY	
WORKFLOW	5		labs	× Sor	t By: Alphabe	-
Find a Wo	rkflow	Thei	aCoV_ClearLabs		TheiaCoV_ClearLabs_PI	нв
Figure 1		V. v2.	3.2 e: Dockstore	(i)	V. v1.0.0 Source: Dockstore	(;

1745 Shea Center Dr ● Suite 400 ● Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved.

Page	1	of	11	1



TheiaCoV_ClearLabs_PHB	
Version: v2.0.1	
Source: github.com/theiagen/public_health_bioinformatics/	FheiaCoV_ClearLabs_PHB:v1.1.0
Synopsis: No documentation provided	
O Run workflow with inputs defined by file paths	h
Run workflow(s) with inputs defined by data table	D
s	
s 🜔 t 🚛 🔆 KK_ClearLabs_Test 💙	
Use call caching 🚺 🗌 Delete intermediate outputs 🜖	Use reference disks 🟮 🗌 Retry with mo Figure 2.

- 3. Uncheck call caching (Fig 2)
- 4. Choose the *latest version* of the workflow, or the version used for internal 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 under the select data table dropdown (Fig 2, c)
- 7. Click select data (Fig 2, d)
- 8. In the pop-up window select the checkbox for each sample to be included in the analysis (Fig 3)



- 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
- c. Scroll to the bottom and click *ok*

Q				
	Search		SETTINGS 2 rows selected	ect Ki d s_Tests
as	ly_length_unambiguous	① assemb	1 O QC_Call	KK_ClearLabs_Test_id
1		1061	FAIL	-00003868
1.		715	FAIL	-00003870
	» Iter je: 100		FAIL	2)

Formatted: Intense Emphasis, Font: (Default) +Body (Calibri), Font color: Auto

1745 Shea Center Dr • Suite 400 • Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved. Page 2 of 11



Analyzing SARS-CoV-2 Da	ata in Terra using Theiagen's
TheiaCoV Clea	arLabs Workflow
Document TG-	SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

- 9. Specify the desired *dataset tags* and *docker image* inputs
 - a. To run TheiaCoV_ClearLabs_PHB for the first time or configure with the newest dataset tags and docker images, upload the TheiaCov input json file on the inputs tab by navigating to the Key Resources Notion page titled Docker Image and Reference Materials for SARS-CoV-2 Genomic Characterization
 - i. **NOTE**: TheiaCoV PHB v2.0.0+ workflows are not backwards compatible with older versions of Nextclade; <u>use Nextclade Dataset Tag</u> 2024-04-15—15-08-222 or <u>newer</u>
 - ii. Expand the TheiaCoV in PHB (v2.0.0 or higher) section, followed by the Terra.Bio Input
 ISONs for PHB v2.0.0 or higher; click on the json file associated with the ClearLabs platform, TheiaCoV ClearLabs PHB 2025-04-02.json, or newer
 - iii. Right click and save the file (text does not have to be selected to save properly)
 - *iv.* Return to the workflow in Terra, click *upload json* (Fig 4, red circle), *select* the saved json file, and click *open*
 - b. To run the workflow with previously saved dataset tags and docker images, no changes are needed
 - c. To add docker images and dataset tags as workspace files for availability in input dropdowns, refer to appendix 10.2

Use call caching	e intermediate outputs Use reference dis		mory ①	22020-00-10-12-10-00-j
SCRIPT •• INPUT		ANALYSIS	Π	CANCEL
Hide optional inputs			Download json Drag or cick to upl	load json Gear inputs SEARCH INPU
Task name ↓	Variable	Туре	Attribute	
theiacov_clearlabs	clear_lab_fastq	File	this.reads	Ę
theiacov_clearlabs	primer_bed	File	workspace.Artic_V4-1_primer_bed	E.
theiacov_clearlabs	samplename	String	Required	
	cpu	Int	this.KK_ClearLabs_Test_id	
Figure 4.			this reads	

- 10. Set the first and third attributes in the table to *this.reads* and *this.KK_ClearLabs_Test_id*, respectively (Fig 4)
 - a. Where *this.KK_ClearLabs_Test_id* is the unique name of your data table in Terra
- 11. Manually choose the *primer_bed* file for the primer set used to sequence samples
 - Labs using the Artic V4-1 will choose workspace.Artic_V4-1_primer_bed; for other primer bed files, see <u>Docker Image and Reference Materials for SARS-CoV-2 Genomic Characterization</u> for available primer bed files

1745 Shea Center Dr ● Suite 400 ● Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved. Page 3 of 11



, .	ata in Terra using Theiagen's arLabs Workflow
	-SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

b. To add workspace files for availability in input dropdowns, refer to appendix 10.2

SCRIPT ** INPUTS * Output files will be saved to □ Files / submission unique ID / theiacov_cleart	OUTPUTS PUN ANALYSIS abs / work/ wunique ID			
References to outputs will be written to Tables / KK_ClearLabs, Test Fill in the attributes below to add or update colu	imns in you a table		nywniaed juon Dreg or cick to uplead juon Clear outputs SEADCH OUTPUTS	5
Task name 🖡	Variable	Туре	Attribute Use defaults	
theiacov_clearlabs	aligned_bai	File	thisaligned_bai	
Figure 5.	aligned_bam	File	thisaligned_bam	
	antia dantan	Carlos .	akin matin stanton	

12. Specify outputs by clicking on the *outputs* tab and *use defaults* (Fig 5)

13. Click save

14. Launch the workflow by clicking run analysis (Fig 5); enter desired comments and click launch

4.2 QUALITY ASSESSMENT OF THEIACOV OUTPUTS

- 1. Navigate to the *data* tab of the workspace containing SC2 data and open the pertinent data table
- 2. Click <u>Select Columns</u> (Fig 6, green rectangle) and select <u>none</u> to deselect all output columns (Fig 6, yellow highlight)
- 3. To simplify the table, select the three following outputs that will be used to make a QC assessment: *assembly_length_unambiguous*, *Number_N*, and *percent_reference_coverage*
 - a. <u>Optional</u>: save this selection by clicking in the save this column selection field and naming it (e.g. QC_assessment); do not include any spaces in the name (Fig 6, red rectangle)
 - b. Click *done*

	CES	Workspaces + the Data	sagen-training-workspaces/APHL_OF
ASHBOARD DATA	ANA	LYSES WORKF	LOWS SUBMISSION HISTORY
O Import Data	/ E.O.	C 🗙 Open With.	Di Erpot 🏟 Select Columns
rables ~	0.	illumina_ 10	QC_assignment
Search all tables		Sample_01	PASS
acinetobacter_test (3)		Sample_02	DASS
E aspergillus_fumigatus (2) ①		Sample_03	FAIL
🖩 aspergillus_furnigatu (1) 🛈		Sample_04	FAIL.
gps_validation (500)		Sample_05	PASS
I IIIumina pe_speci (20) ①		Sample_06	FAIL
B illumina pe specime (3) ()	D	Sample_07	FAIL
Figure 6.		Sample_08	FAIL
inguic 0.		Sample,09	FAIL

1745 Shea Center Dr • Suite 400 • Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved. Page 4 of 11



Analyzing SARS-CoV-2 Da	ata in Terra using Theiagen's
TheiaCoV Clea	arLabs Workflow
Document TG-	SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

4. Optional: add a column to record QC PASS/FAIL by clicking edit, add a column (Fig 7) a. Name the new

spaces

column (e.g. QC_Call); do

TABLE Q (5) (6) not include any b. Set the value Figure 7. type as string

- c. Click save 5. Use table 1 to assess the quality of each sample's genome assembly (see next page) &/or labspecific quality metrics
- 6. Optional: notate in the QC_assessment field for each sample PASS or FAIL by clicking the pencil icon in the corresponding field (Fig 7, red circle)
- 7. For samples that pass the guidance thresholds, proceed to section 4.3
 - a. For samples that do not pass guidance thresholds, resequence
 - i. Samples not meeting guidance thresholds indicated here may proceed to analysis at the discretion of the laboratory

Table 1. Guidance thresholds for genome assembly QC

QC Metric	Data Table Column	Guidance Threshold* ¹
Number N	number_N	<5kbp
Assembly length unambiguous	assembly_length_unambiguous	>24kbp
Percent reference coverage	percent_reference_coverage	>83%

¹ Metrics and thresholds presented for guidance only as there are currently no standard assembly metric requirements; internal validation procedures will ultimately define acceptable assembly QC parameters

¹⁷⁴⁵ Shea Center Dr • Suite 400 • Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved.



Analyzing SARS-CoV-2 Da	ata in Terra using Theiagen's
TheiaCoV Cle	arLabs Workflow
Document TG	-SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

4.3 DETERMINING SARS-CoV-2 CLADES, LINEAGES, AND WHO VARIANTS OF CONCERN (VoC)

- 1. Navigate to the *data* tab of the Terra workspace containing SC2 data of interest
- 2. *Open the data table* by clicking on the name of the data table in the left sidebar
- 3. View settings above the data table (Fig 6), select none (Fig 7)
- 4. Select the following columns: *nextclade_clade* and *pango_lineage*
 - a. <u>Optional</u>: save this column group for future use by clicking the save this column selection field, naming it (e.g. SC2_Results), and clicking save
- 5. Click done
- 6. Determine the Nextclade clade for each sample
 - a. In the data table, find the column titled <u>nextclade_clade</u>; result formats will use the following nomenclature: <u>21L (Omicron)</u> where:
 - i. 211 indicates the sample clade and
 - ii. In parentheses, (Omicron), contains the WHO variant of concern classification
 1. Not every sample will belong to a WHO classification
 - b. Samples indicated as recombinant may indicate a case where multiple strains have combined during viral replication producing a new lineage
 - c. More information on SARS-CoV-2 recombinants can be found at the following Github site: <u>pipeline-resources/docs/sc2-recombinants.md at main · pha4ge/pipeline-resources · GitHub</u>
- 7. Identify the Pangolin lineage for each sample
 - a. In the data table, find the column titled *pango_lineage*; nomenclature will be similar to the following: B.1.167
 - b. For more information on each of the lineages, visit <u>https://cov-lineages.org/lineage_list.html</u>
- 8. Follow lab-specific QC, resulting, and reporting procedures, as applicable

5. QUALITY RECORDS

- Raw read files
- Workflow version and input parameters
- Reference sequence, if applicable
 - SC2: 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.
 - Sample read, assembly, and result-specific QC metrics
- Sample read, assembly, and result-specific QC metrics
 All workflow outputs relevant to results including to all and do
- All workflow outputs relevant to results, including tool and database versions

1745 Shea Center Dr • Suite 400 • Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved. Page 6 of 11



1 0	ata in Terra using Theiagen's arLabs Workflow
Document TG-	SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

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

- This SOP is written for the analysis of SC2 data; v2+ of the TheiaCoV_ClearLabs_PHB workflow is also compatible with the following pathogens: monkeypox virus (MPXV), human immunodeficiency virus (HIV), west nile virus (WNV), influenza virus, and respiratory syncytial viruses A and B (RSV). Refer to <u>Theiagen Public Health Resources documentation</u> for organismspecific parameters and details.
- TheiaCoV PHB v2.0.0+ workflows are not backwards compatible with older versions of Nextclade; use Nextclade Dataset Tag 2024-04-15-15-08-222 or newer

8. REFERENCES

- Smith, E., Wright, S., & Libuit, K. (2022, June 28). *Identifying SARS-CoV-2 Recombinants*. Github. Retrieved June 16, 2023, from <u>https://github.com/pha4ge/pipeline-</u> resources/blob/main/docs/sc2-recombinants.md#identifying-sars-cov-2-recombinants
- O'Toole, Áine et al. "Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch." Wellcome open research vol. 6 121. 17 Sep. 2021, doi:10.12688/wellcomeopenres.16661.2
- Libuit, Kevin G., Emma L. Doughty, James R. Otieno, Frank Ambrosio, Curtis J. Kapsak, Emily A. Smith, Sage M. Wright, et al. 2023. "Accelerating Bioinformatics Implementation in Public Health." Microbial Genomics 9 (7). <u>https://doi.org/10.1099/mgen.0.001051</u>
- 4. Theiagen Genomics Public Health Bioinformatics Workflow Documentation

9. REVISION HISTORY

Revision	Version	Release Date
Document creation	1	7/2023
Added TG-TER-04 reference, uncheck call caching, updated input json, figures, and formatting	2	9/2023
Removed section 4.1 for creating a metadata tsv file (refer to TG- TER-03 and TG-TER-04 for details); updated quality records and limitations sections; added primer bed file upload instructions; added appendices 10.1 and 10.2	3	5/2024
Updates for version release, aligning with Terra interface, formatting	4	5/2025

1745 Shea Center Dr • Suite 400 • Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved. Page 7 of 11



10. APPENDICES

10.1 IMPORTING THE THEIACOV_CLEARLABS_PHB WORKFLOW FROM DOCKSTORE

WORKSPACES	Workspaces + theiagen-training-workspa Workflows	Suggested Workflows	Figure 9
HBOARD DATA A	NALYSES WORKFLOWS JOB HISTORY		mutect2-gatk4
WORKFLOWS		Runs HaplotypeCaller from GATK4 in GVCF mode on a single sample	Implements GATK4 Mutlect 2 on it single turnor- normal pair
Find a Workflow	Augur_PHB	processing for variant-discovery-gatk4	validate-barn
0	V. main Source: Dockstore	Implements data pre-processing according to the CATK Best Practices	This WDIL performs format validation on SAM/BAM files in a list.
Freyja_Dashboard_PHB	Freyja_FASTQ	paired fastq to unmapped barn	generate-sample-map
	V main	Find Additional Workflows	Broad Methods Repository
Figure 8.	Source: Dockstore	Browse WDL workflows in Dockstore, an open platform used by the GA4CH for sharing Docker- based workflows.	Broad Methods Repository Use Broad workflows in Terra. Share your own choose from > 700 public workflows

- 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)
- 3. In the top banner click *Organizations*; then click *Theiagen Genomics* (Fig 10)
- 4. Open the Public Health Bioinformatics (PHB) collection (Fig 11)



1745 Shea Center Dr ● Suite 400 ● Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved.

Analyzing SARS-CoV-2 Data in Terra using Theiagen's TheiaCoV ClearLabs Workflow	
Document TG-SC2-CL, Version 4	
Date:	Workflow Versions
5/1/2025	PHB v3.0.0

Theiagen GENOMICS	Theiagen Genomics	
CO GENOMICS	Public health bioinformatics for pathogen surve	illance
Collections 4	Members 2 C Updates 10	About the Organization
n Public Health Bioin	formatics (PHB)	This organization does not have a description
-	public health pathogen genomics	G https://www.theiagen.com

 To find the TheiaCoV_ClearLabs_PHB workflow in Windows environments, hold <u>Ctrl + F</u> and <u>search TheiaCoV_ClearLabs</u>, then click on the link (Fig 12)



6. Click Terra to launch the workflow in Terra (Fig 13)

Last update to source repository: 3 hours ago Labels Thelagen phb Info Launch Versions Files Tools DAG Metrics Launch with	
Info Launch Versions Files Tools DAG Metrics Launch with	
DNAn	exus
Workflow Information	Ta
Source Code:	
TRS: #workflow/github.com/theiagen/public_health_bioinformatics/TheiaCoV_ClearLabs_PHB	<u>8.21</u>
Topic Bioinformatics workflows for genomic characterization, submission preparation, and genomic enidemiology of	

7. Choose the *destination workspace* in the dropdown and click *import* (Fig 14)

1745 Shea Center Dr • Suite 400 • Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved.



, .	ata in Terra using Theiagen's arLabs Workflow
Document TG-	-SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0



10.2 ADDING WORKSPACE DATA ELEMENTS

- 1. Navigate to the *Terra workspace* where analysis will be run
- To upload local files, open the *Files* tab in the right-side panel of the workspace (Fig 15, box)
 a. Click *upload* (Fig 16)
 - b. Once the upload is complete, *right click* on the file name and click *copy link*

DASHBOARD DATA	ANALYSES WORKFLOWS	SUBMERSION HISTORY Select a data type from the navgation panel on the left	() Sata
TABLES ~			\$0.00 particur
Scarch all tables Q		-	
REFERENCE DATA ~ No references have been addred. Add reference data			
OTHER DATA ~			Figure 15.
= 向 w	ORKSPACES	Workspaces > theiagen-training-workspaces/Theiagen_Kropp_Sandbox > Data	

DASHBOARD DATA	ANALYSES WORKFLOWS JOB HISTORY	
sd_theiaprok (103)	Elles Co UPL	OAD
sd_theiaprok_set (2)	□ Name	
🗏 sra_fetch (28)	Assemblies/	
sra_fetch_set (4)	Candida_albicans_ATCC_10231.fasta	
Figure 16. 👝 👩 🕐	D FASTAS/	

1745 Shea Center Dr ● Suite 400 ● Highlands Ranch, CO 80129 © Theiagen Consulting LLC, 2025. All Rights Reserved.



Analyzing SARS-CoV-2 Da	ata in Terra using Theiagen's
TheiaCoV Clea	arLabs Workflow
Document TG-	SC2-CL, Version 4
Date:	Workflow Versions
5/1/2025	PHB v3.0.0
	· · · · · · · · · · · · · · · · · · ·

- 3. Open the *workspace data* tab (Fig 15, circle)
- 4. Click Edit and Add variable in the top tool bar (Fig 17)



- 5. Click in the key field and name the element being added (Fig 18)
 a. E.g. to add the Artic v4-1 primer bed file, the key Artic_v4-1_primer_bed may be used
- 6. In the value field, choose *string* as the value type
 - a. Paste the file path; the value should start with gs://
 - b. NOTE: For other string elements like dataset tags and docker images paste the ID value i.E.g. for the nextclade docker image, add nextstrain/nextclade:2.14.0
 ii.Always ensure the docker images and dataset tags are aligned with versions used for internal validation procedures
- 7. <u>Optional</u>: A description may be added to denote the date updated with staff initials
- 8. Click the blue check mark on the right-hand side of the variable to save it
 - a. The variable will now be available as a workflow input which can be found by typing the prefix *workspace*. plus the key name *artic_v4-1_primer_bed*
 - i.e.g. *workspace.*artic_v4-1_primer_bed

Key	Value	Description	e 18
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 Dat	a
Freyja_ReferenceGenome	• nCoV-2019.reference.fasta	MN908947.3	1
Midnight_primer_bed	Midnight_Primers_SARS-CoV-2.scheme.bed		Ł
SWIFT_primer_bed	gs://theiagen-public-files/terra/theia	V Updated 2023-07-05 kk	9 (