

Document TG-FLU-PE, Version 1				
Date:	Effective Date:	Workflow Version:		
8/4/2023	8/2023	PHB v1		

### 1. PURPOSE/SCOPE

To standardize the process of analyzing influenza (Flu) next generation sequencing (NGS) data using Theiagen's TheiaCoV\_Illumina\_PE\_PHB, Augur\_Prep\_PHB, and Augur\_PHB workflows in Terra to determine typing, subtyping, and lineage designation. Acceptable data types include Illumina paired end (PE) raw read file format.

### 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
- Illumina paired end (PE) raw sequencing read files uploaded to Terra workspace
- Theiagen's TheiaCoV\_Illumina\_PE\_PHB, Augur\_Prep\_PHB, and Augur\_PHB workflows in Terra

### 3. RELATED DOCUMENTS

Document Number	Document Name
	Uploading Local or SRA NGS Data & Creating a
TG-TER-03	Results Metadata Table in Terra

### 4. PROCEDURE

### 4.1 CREATE A SAMPLE METADATA FILE (TSV FILE) FOR RAW READS

1. In Excel, *create a list* containing the following sample information:

### a. For all analyses:

- i. Column 1 header (Fig 1): *entity:kilifi\_H3N2\_id* where *kilifi\_H3N2* is the name of the data table/group of samples to be analyzed
- ii. List all *sample IDs* in column 1

entity <mark>:kilifi_H3N2_</mark> id	accession	collection_date	continent	country	state	read1	read2
100734	A/Kilifi/131/2010	8/5/2010	Africa	Kenya	Kilifi	100734_R	100734_R
100954	A/Kilifi/132/2010	8/17/2010	Africa	Kenya	Kilifi	100954_R	100954_R
SRR11445941	SRR11445941	1/24/2017	Europe	Belgium	Brussels-C		
SRR13443360	SRR13443360	6/8/2018	Europe	Belgium	Brussels-C		
Figure 1: Raw Read Me	tadata File. 76	3/11/2019	Europe	Spain	Catalonia		



Document TG-FLU-PE, Version 1				
Date: Effective Date:		Workflow Version:		
8/4/2023 8/2023		PHB v1		

# b. For analysis from raw sequencing reads (Fig 1)

- i. Use column headers: read1 and read2
- ii. List the *full cloud file paths* to read1 and read2 files
- c. For analysis using SRA fetch (Fig 2):
  - i. Column header: *accession*, or similar
- d. <u>Optional</u>: remaining columns may be used to add metadata like run\_id, additional lab results, sample collection information, demographic data, etc
- e. Do not include spaces in the headers
- 2. Save as a txt or tsv file
- Upload to Terra workspace; see TG-TER-03 for details

# 4.2 RUNNING THE THEIACOV WORKFLOW



- 1. Open Terra and navigate to the workflows tab in the workspace containing flu data
- 2. Select the *TheiaCoV\_Illumina\_PE\_PHB* workflow (Fig 2)
- 3. Choose the latest version of *version 1* in the version dropdown field, or the workflow version that was used during internal assay validation (Fig 3, a)
- 4. Select the second bullet to run workflow(s) with inputs defined by data table (Fig 3, b)
- 5. Select the data table name under the select *root entity type* dropdown (Fig 3, c)
- 6. Click select data (Fig 3, d)

TheiaCoV_IIIumina_PE_PHB	Figure 3.
Version: v1.0.1 • <b>2</b>	
Source: github.com/theiagen/public_health_bioinformatics/TheiaCoV_Illumina_PE_PHB:v1.0.1	
Synopsis:	
No documentation provided	
O Run workflow with inputs defined by file paths	
Run workflow(s) with inputs defined by data table	
Use call caching 🕄 🗌 Delete intermediate outputs 🕄 🗌 Use reference disks 🕄 🗌 Retry with more memory 🖉	Ignore empty outputs



Document TG-FLU-PE, Version 1				
Date:	Effective Date:	Workflow Version:		
8/4/2023	8/2023	PHB v1		
8/4/2023	8/2023	PHB v1		

- 7. In the pop-up window select the checkbox for each sample to be included in the analysis (Fig 4)
  - 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
  - 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 4, highlight)
- c. Scroll to the bottom and click ok
  d. Optional: Check the box to ignore empty outputs (Fig 3)
  8. Set the first three attributes in the table to this.kilifi\_H3N2\_id (Fig 5), respectively
  - a. Where *kilifi\_H3N2* is the unique name of your data table in Terra
  - b. In the *organism* attribute field enter *"flu"* in quotation marks (Fig 5)

ide optional inputs			Download json   Drag or click to upload json   Cle
Task name ↓	Variable	Туре	Attribute
theiacov_illumina_pe	read1_raw	File	this.read1
theiacov_illumina_pe	read2_raw	File	this.read2
theiacov_illumina_pe	samplename	String	this.kilifi_H3N2_id d
abricate_flu	cpu	Int	Optional

- 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

	g Influenza Data in Terra CoV Illumina PE and Aug	• •	
	Document TG-FLU-PE, Version 1		
Date:	Effective Date:	Workflow Version:	
8/4/2023	8/2023	PHB v1	

SCRIPT •• INPUTS	•• OUTPUTS •• RUN A	NALYSIS		
Output files will be saved to  Files / submission unique ID / theiaco	/_illumina_pe / w			
References to outputs will be written to 国 Tables / illumina_pe_specimen Fill in the attributes below to add or upda	te columns in you ta table			CANCEL SAVE
		Download json   Drag or cl	ick to upload json   Clear outputs	SEARCH OUTPUTS
Task name ↓	Variable	Туре	Attribute   Use defaults	
Figure 6.	abricate_flu_database	String	Optional	{}

## 4.3 RUNNING THE AUGUR PREP WORKFLOW FOR HA OR NA FLU PROTEINS

- Confirm the TheiaCoV worklow has successfully completed by viewing the relevant job submission status in the *job history* workspace tab; green and red indicate successful and failed jobs, respectively, while blue represents unfinished jobs
- 2. Perform quality assessment per internal protocols and proceed with Augur Prep for samples passing quality control (QC) metrics
  - a. For samples that do not meet QC thresholds, resequence
    - *i.* Samples not meeting thresholds may proceed to Augur Prep at the discretion of the lab
- 3. Navigate to the *workflows* tab and select the *Augur\_Prep\_PHB* workflow
- 4. Choose the latest version of the workflow in the *version dropdown field*, or the workflow version that was used during internal assay validation
- 5. Select the second bullet to *run workflow(s) with inputs defined by data table* (Fig 3, b)
- 6. Select the relevant data table name under the select **root entity type** dropdown (Fig 3, c)
- 7. Click *select data* (Fig 3, d)
- 8. In the pop-up window select the checkbox for each sample to be included in the analysis (Fig 7)
  - a. Do not include any samples that are missing either *irma\_ha\_segment* or *irma\_na\_segment*
  - b. Click the down arrow and select all to process all samples
  - c. Additionally, the search bar may be used to narrow down the sample list to only those matching search criteria (e.g. only "SRR" sample names or just "Type\_A" samples, etc)
  - *d.* <u>Optional</u>: name the output file something to differentiate it from other runs, e.g. HA\_AugurPrep\_YYYYMMDD.# or NA\_AugurPrep\_YYYYMMDD.# (Fig 6, c)
  - e. Scroll to the bottom and click ok

	g Influenza Data in Terra CoV Illumina PE and Aug	0
	Document TG-FLU-PE, V	ersion 1
Date:	Effective Date:	Workflow Version:
8/4/2023	8/2023	PHB v1

□•	kilifi_H3N2s to process	↓ ①scan_version	irma_ha_segment	irma_na_segment	(
	99056	-scan 0.4.4	99056_HA.fasta	99056_NA.fasta	
	<b>6</b> h	-scan 0.4.4	SRR11445892_HA.fasta	SRR11445892_NA.fasta	
	SRR11445940	-scan 0.4.4	SRR11445940_HA.fasta	SRR11445940_NA.fasta	
	SRR11445941	-scan 0.4.4	SRR11445941_HA.fasta	SRR11445941_NA.fasta	
	SRR13443360	-scan 0.4.4	SRR13443360_HA.fasta	SRR13443360_NA.fasta	
	SRR19881876				
	SRR19881876	4			

- 9. Set the first five input attributes in the table to the following, respectively (Fig 8)
  - a. this.irma\_ha\_segment
  - i. When running Augur Prep on the NA segment, enter this.irma\_na\_segment
  - b. this.collection\_date
  - c. this.continent
  - d. this.country
  - e. this.state
    - i. Input text shown in grey indicates this is variable; text must your Terra data table column headers, excluding the prefix "this." (Fig 1)
- 10. In the *organism* attribute field enter *"flu"* in quotation marks (Fig 5)
- 11. Specify outputs by clicking on the *outputs* tab and *use defaults* (Fig 6); click *save*
- 12. Launch the workflow by clicking run analysis; enter desired comments and click launch
- 13. Repeat section 4.3 to run Augur Prep for the NA protein with the following changes:
  - a. When selecting data to run, *enter a different output file name* to distinguish the NA Augur run from the HA run, e.g. NA\_AugurPrep\_YYYYMMDD.# (Fig 7, c)



		Document TG-FLU-PE, Ve	ersion 1
	Date: Effective Date:		Workflow Version:
8/4/2023 8/2023		8/2023	PHB v1

b. Specify <u>this.irma\_na\_segment</u> in the first input field rather than using this.irma\_ha\_segment as before

	<u> </u>		
Hide optional inputs	Variable	Туре	Attribute
		iype	
augur_prep	assembly	File	this.irma_ha_segment
augur_prep	collection_date	String	this.collection_date
augur_prep	continent	String	thiscontinent
augur_prep	country	String	thiscountry
	state	String	this.state

	e a new kilifi_H3N2_set from selected kilifi_H3N2s se specific kilifi_H3N2_sets to process		
Selec	t kilifi_H3N2_sets to process 💠 SETTINGS 1 row selected =	ADVANCED SEARCH	Search Q
•	kilifi_H3N2_set_id	kilifi_H3N2s	
	HA_AugurPrep_20230627.1	92804, 93547, 944	(37 entities)
	NA_AugurPrep_20230627.1	92804, 93547, 944	(37 entities)
	SRA_Fetch_PHB_2023-06-27T18-24-30	SRR11445892, SRR1	(5 entities)
	TheiaCoV_IIIumina_PE_PHB_2023-06-27T18-29-37	92804, 93547, 944	(70 entities)



Document TG-FLU-PE, Version 1				
Date: Effective Date: Workflow Version:				
8/4/2023	8/2023	PHB v1		

# 4.4 RUNNING THE AUGUR [ANALYSIS] WORKFLOW

- 1. Open the Augur\_PHB workflow in the workspace containing flu data
- 2. Choose latest version of the workflow in the *version dropdown field*, or the workflow version that was used during internal assay validation
- 3. Select the second bullet to *run workflow(s) with inputs defined by data table* (Fig 3, b)
- 4. Select the relevant SET data table under the select root entity type dropdown (Fig 3, c)
  - a. E.g. Kilifi\_H3N2\_set
- Click select data and choose the output file name from the Augur Prep workflow previously ran (e.g. HA\_AugurPrep\_YYYYMMDD.#); if Augur Prep was run twice for the HA and NA protein segments, the Augur Analysis workflow must be run twice, as well (they cannot be run together)
- 6. Click on the *inputs* tab to specify settings
- Set the first three attributes in the table to *this.kilifi\_H3N2s.irma\_ha\_segment*, *"kilifi\_H3N2\_HA"*, and *this.kilifi\_H3N2s.augur\_metadata*, respectively (Fig 10)
  - a. Where *kilifi\_H3N2* is the unique name of your data table in Terra for all three attribute fields
- Set the flu\_segment, flu\_subtype, and organism optional fields to "HA", "H3N2", and "flu", respectively (Fig 10)
- 9. Specify outputs by clicking on the outputs tab and use defaults (Fig 6); click save
- 10. Launch the workflow by clicking run analysis; enter desired comments and click launch

SCRIPT •• INPUTS	OUTPUTS     OUTPUTS     RUN ANALYSIS						
Hide optional inputs							
Task name 🖡	Variable	Туре	Attribute				
augur	assembly_fastas	Array[File]+	this.kilifi_H3N2sirma_ha_segment				
augur	build_name	String	"kilifi_H3N2_HA"				
augur	sample_metadata_tsvs	Array[File]+	this.kilifi_H3N2s.augur_metadata				
augur	auspice_config	File	Optional				
augur	clades_tsv	File	Optional				
augur	distance_tree_only	Boolean	Optional				
augur	flu_segment	String	"НА"				
augur	flu_subtype	String	"H3N2"				
augur	lat_longs_tsv	File	Optional				
augur	min_num_unambig	Int	Optional				

	g Influenza Data in Terra CoV Illumina PE and Aug	0			
Document TG-FLU-PE, Version 1					
Date: Effective Date: Workflow Version:					
8/4/2023	8/2023	PHB v1			

	•	kilifi_H3N2_set_id	t 🛈	kilifi_H3N2s		
		HA_AugurPrep_20230627.1		92804, 93547, 944	(37 entities)	
		NA_AugurPrep_20230627.1		92804, 93547, 944	(37 entities)	
	Figure	11A. <sup>:ch_PHB_2023-06-27</sup>	T18-24-30	SRR11445892, SRR1	(5 entities)	
SCRIPT ••	INPUTS	•• OUTPUTS ••	RUN ANALYSIS			
de optional inputs						
Task name ↓		Variable		Туре	Attribute	
augur	ugur ass			Array[File]+	this.kilifi_H3N2s.irr	ma_na_segment
augur build_name			String	"kilifi_H3N2_NA"		
augur sample_metadata_tsvs			5V5	Array[File]+	this.kilifi_H3N2s.au	igur_metadata
augur						
augur		flu_segment		String	"NA"	

- 11. Repeat section 4.4 for the NA protein with the following changes:
  - a. When selecting data to run, choose the *NA Augur Prep output file* (Fig 11A)
  - b. Specify *this.kilifi\_H3N2s.irma\_na\_segment*, *"kilkifi\_H3N2\_NA"* and *"NA"* in the first, second, and seventh input fields (Fig 11B)



- Navigate to the data tab of the Terra workspace containing flu data
- 2. Open the data table by clicking on the name of the data table in the left sidebar





Document TG-FLU-PE, Version 1					
Date: Effective Date: Workflow Version:					
8/4/2023	8/2023	PHB v1			

- 3. View *settings* above the data table (Fig 12), select *none* (Fig 13)
- 4. Select the following columns: *abricate\_flu\_subtype*, *abricate\_flu\_type*, *irma\_subtype*,

*irma\_subtype*, *nextclade\_clade*, and nextclade\_lineage

- a. <u>Optional</u>: save this column group for future use by clicking the save this column selection field, naming it (e.g. FluTyping), and clicking save
- IRMA is used to produce a consensus and variants assembly; abricate is used to confirm IRMA typing results
- 5. Click done
- 6. Determine the type, subtype, and

lineage for each sample by viewing the corresponding columns

- 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
- 8. Follow lab-specific resulting and reporting procedures, as applicable

ASHBOARD DATA	AN	ALYSES WORKFLOWS JOB HISTORY	Select columns	
			Show: all   none Sort: alphabetical	
	1 1 60	DIT 🔀 OPEN WITH 🕒 EXPORT 🏟 SETTINGS	meanbaseq_trin	N SELECTION
			nextclade_aa_dels	
ABLES	~ □•	kilifi_H3N2_id 10 ivar_v prop	nextclade_aa_subs	
a		100734	- 🛛 nextclade_clade	
Search all tables	2	C	- nextclade_docker	
1		100954	nextclade_ds_tag	
🛙 kilifi_H3N2 (39)		0 109275	🔲 nextclade_json	
		109275	🗌 nextclade_lineage	e e
kilifi_H3N2_set (2)		109292	nextclade_tamiflu_resistance_aa_subs	
REFERENCE DATA	~		- nextclade_tsv	
EFERENCE DATA		109342	- nextclade_version	
No references have been adde	d.	109562	num_reads_clean1	
dd reference data		109562	- Dinum reads clean?	
	. 0	109630	CANC	DONI

# 4.6 VISUALIZING THE AUGUR TREE IN AUSPICE

- 1. Navigate to the *workspace data tab* and select the *data table* containing flu data
- 2. Click *settings* and *none*; then select only the nextclade\_clade column (Fig 14)





Document TG-FLU-PE, Version 1					
Date: Effective Date: Workflow Version:					
8/4/2023	8/2023	PHB v1			

- 3. <u>Select the checkbox</u> for each sample containing a value in the nextclade\_clade column, including "unassigned" values (Fig 15, g)
- 4. Click *export* and *download as TSV* to download the metadata file (Fig 15, h)
- 5. *Save the file* as "[DATA\_TABLE\_NAME]auspice\_metadata\_YYMMDD.tsv"
- 6. In the flu Terra workspace, *open the "set" data table* with flu data; nomenclature is the same as the original data table plus the suffix "\_set" (e.g. flu\_H3N2\_set)
- Under the column titled auspice\_input\_json, click to download the file corresponding to the sample set of interest
  - a. A new window will open in the browser; *right click* and *save* this file
- In a new browser window open https://auspice.us/; drag and drop the auspice input json file onto the webpage
- Drag and drop the metadata file onto the webpage
  - a. The sample names in the auspice output json and the metadata file must match; if they don't, open the metadata file and manually edit the names
- 10. In auspice, click the *color by* dropdown and choose *nextclade clade* 
  - a. Unassigned clades may be older clades that are not currently assigned a clade by nextclade; these are displayed in black in the auspice timetree
- 11. *View the timetree and map* with geographical coordinates of the analyzed strains (Fig 17)
- 12. Repeat section 4.6 with the NA auspice tree and corresponding metadata

DASHBOARD DA	ТА	ANAL	YSES WORKFLOWS	JOB HISTORY
IMPORT DATA		/ EDI	T 🔀 OPEN WITH 🕒 EX	
TABLES	~	•	kilifi_H3N2_id ↓	nextclade_clade
Search all tables	۹		100734	h <sup>3C</sup>
🗉 kilifi_H3N2 (39)	1		954	unassigned
			109275	3В
kilifi_H3N2_set (2)	(!)		109292	3В
REFERENCE DATA	~		109342	3C
No references have been as Add reference data	dded.		109562	3C
OTHER DATA	~		109630	
Workspace Data			109974	3В
Figure 15.			110108	3C



	g Influenza Data in Terra CoV Illumina PE and Aug	0 0			
	Document TG-FLU-PE, Version 1				
Date:	Date: Effective Date: Workflow Version:				
8/4/2023	8/2023	PHB v1			



### 5. QUALITY RECORDS

- Raw read files
- Sample read and assembly QC metrics
- Nextclade\_clade and pango\_lineage determinations
- Auspice timetree and map

### 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. INTERFERENCES

N/A

#### 8. REFERENCES

None

#### 9. **REVISION HISTORY**

Revision	١.	Version	Release Date
Document creation	1	1	7/2023



Document TG-FLU-PE, Version 1		
Date:	Effective Date:	Workflow Version:
8/4/2023	8/2023	PHB v1

### **10. APPENDICES**

## 10.1 USING EXCEL TO ALIGN TSV SAMPLE NAMES TO AUSPICE.JSON FILE NAMES

- 1. Open the metadata tsv file titled "[DATA\_TABLE\_NAME]auspice\_metadata\_YYMMDD.tsv" in excel
- 2. Select column B, right click, and insert a column between columns A and B
- 3. In cell B2, enter the following formula relevant to the protein segment being analyzed
  - a. For HA protein segments: =A2&"\_A\_HA\_H3"
  - b. For NA protein segments: =A2&"\_A\_NA\_N2"
- 4. With cell B2 selected, click and drag the green plus at the bottom right of the green box down to copy the formula to the remaining samples
- 5. Select column C, right click, and insert a new column between columns B and C
- 6. Select and copy column B contents
- 7. Select column C and right click to paste values into column C
- 8. Copy and paste cell A1 into C1
- 9. Delete columns A and B leaving only the pasted sample name values and nextclade\_clade
- 10. Save the file and return to auspice; drag and drop or upload the new metadata file into auspice to add corresponding sample metadata