	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date:	Workflow Version:
	10/7/2023	PHB v1

1. PURPOSE/SCOPE

To standardize the process of preparing viral next generation sequencing (NGS) sample assemblies and metadata using Theiagen's Augur_Prep_PHB workflow for analysis via Theiagen's Augur_PHB workflow in Terra to generate a phylogenetic tree and SNP matrix. Acceptable data types include assembly FASTA files and metadata imported via TSV file format. NOTE: For flu-specific augur procedures, refer to the influenza SOP, [TG-FLU-PE](#), on the Theiagen Public Health Resources website ([Theiagen Public Health Resources Notion site](#)).

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
- Sample assembly files uploaded to Terra workspace
- Theiagen's Augur_Prep_PHB, and Augur_PHB workflows in Terra workspace

2. REQUIRED AUGUR PREP INPUTS (5):

1. REQUIRED AUGUR PREP INPUTS (5):
 - Sample assembly (FASTA) files
 - Associated sample metadata fields (4) from (TSV) file* or Terra table


2. REQUIRED AUGUR INPUTS (both output from Augur Prep):

- Array file of FASTA files
- Array file of sample metadata

3. RELATED DOCUMENTS

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

NOTE: Text in gray indicated throughout this SOP may be customized, while black indicates required text. Remember, data table column names can serve as workflow inputs, but must align to do so. E.g. a table containing sample assemblies in a column titled assemblies would use this.assemblies as the Augur Prep assembly file input attribute (value). this. and workspace. prefixes are used in workflow input fields to link to data table columns and workspace files/data, respectively.

	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date:	Workflow Version:
	10/7/2023	PHB v1

4. PROCEDURE

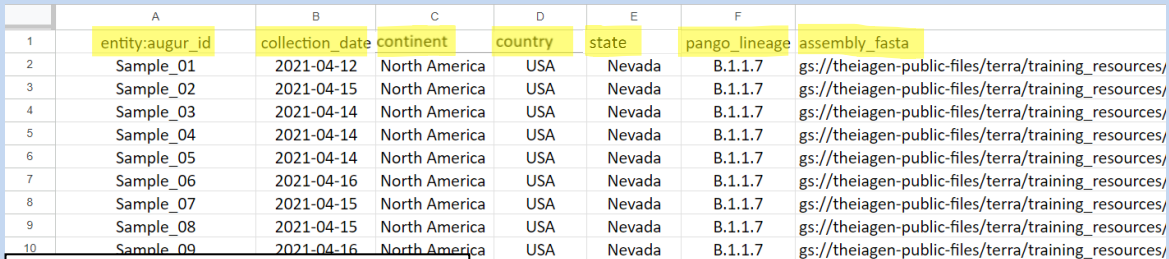
4.1 CREATE A TSV FILE FOR ASSOCIATED SAMPLE METADATA*

*Existing samples in Terra with the required inputs (5) listed below do not need a new TSV file created; proceed with [section 4.2](#)

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

Create a copy of [THIS](#) template and customize!
 (رسول الله) Don't forget to download in .tsv format.

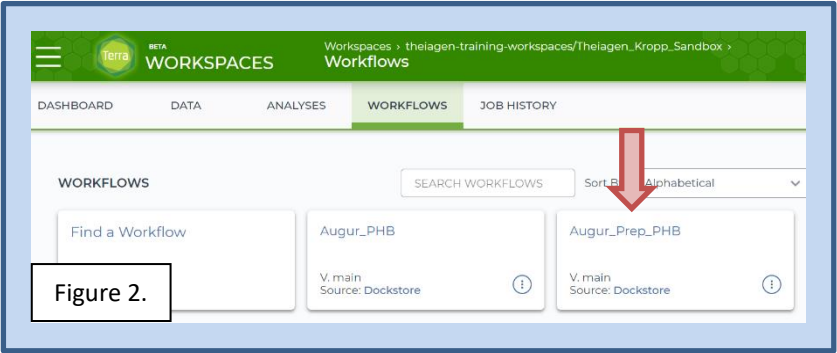
- a. In cell A1, specify the data table name, e.g. `entity:augur_id` (Fig 1)
 - i. List all `sample IDs` in column 1; **do not include spaces in sample names**
- b. Specify the following required sample metadata fields (5) as columns and fill in the relevant information for each sample:



	A	B	C	D	E	F	
1	entity:augur_id	collection_date	continent	country	state	pango_lineage	assembly_fasta
2	Sample_01	2021-04-12	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
3	Sample_02	2021-04-15	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
4	Sample_03	2021-04-14	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
5	Sample_04	2021-04-14	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
6	Sample_05	2021-04-14	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
7	Sample_06	2021-04-16	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
8	Sample_07	2021-04-15	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
9	Sample_08	2021-04-15	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/
10	Sample_09	2021-04-16	North America	USA	Nevada	B.1.1.7	gs://theiagen-public-files/terra/training_resources/


Figure 1: Example Metadata (TSV) File.

- i. `assembly_fasta` - add the cloud file path for each sample assembly file (begin with "gs://")
 1. If this output exists in the Terra data table from a prior workflow, this can be skipped
 - ii. `collection_date` - must use the format YYYY-MM-DD
 - iii. `continent`
 - iv. `country`
 - v. `state`
- c. **Optional: additional columns may be used to add metadata like run_id, lineage, demographic data, etc**



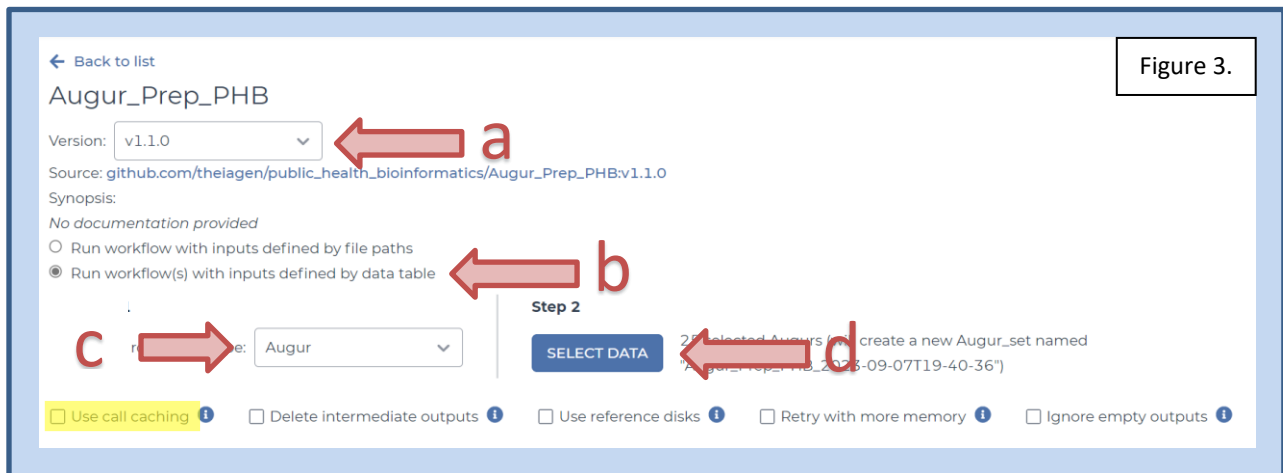
d. **Do not include spaces in the column headers**

2. **Save as** a txt or tsv file
3. **Upload** to Terra workspace; see [TG-TER-03](#) for details

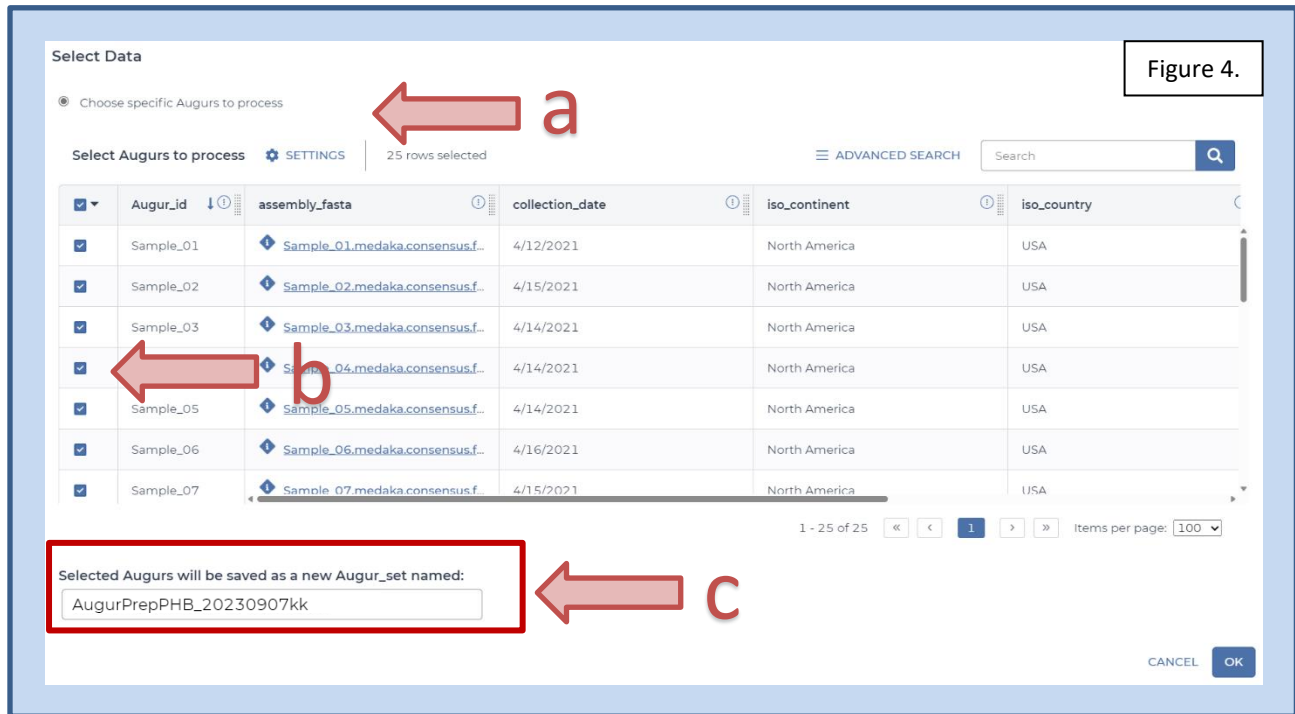
	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date:	Workflow Version:
	10/7/2023	PHB v1

4.2 RUNNING THE AUGUR PREP WORKFLOW


1. Ensure sample data passes quality control (QC) metrics per internal protocols before proceeding with Augur Prep
 - a. For samples that do not meet QC thresholds, resequence or proceed with Augur Prep at the discretion of the laboratory
2. Navigate to the **workflows** tab and open the **Augur_Prep_PHB** workflow (Fig 2)
3. Uncheck use call caching (Fig 3, highlight)
4. Choose the latest version of **v1** in the **version dropdown field** (Fig 3, a) 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 4)
 - a. Do not include any samples that are missing assemblies
 - b. Click the down arrow in the top left of the sample table 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 "B.1.1.7" samples, etc)
 - d. Optional: the output of this workflow will populate a row in the **SET** data table for the Terra table selected in Figure 3c. This SET row (job) must be selected when running the Augur workflow; as such, it may be helpful to name the output file so the user knows which prep output to select, e.g. *AugurPrepPHB_YYYYMMDD[initials]* or similar (Fig 4, c)
 - e. Scroll to the bottom and click **ok**



2. Set the first five input attributes in the table to the following, respectively (Fig 5)
 - a. `this.assembly_fasta`
 - b. `this.collection_date`
 - c. `this.continent`
 - d. `this.country`
 - e. `this.state`
9. *Optional: if sample metadata for other inputs already exist in the Terra data table, they may be specified using the optional input attribute fields (Fig 5, e.g. pango_lineage, nextclade_clade, etc)*
10. SARS-CoV-2 (SC2) is the organism set by default; if samples other than SC2 are being analyzed, enter the organism name in the `organism` attribute field in quotation marks (Fig 5)
 - a. Optional inputs are `"flu"`, `"MPXV"`, and `"WNV"`
 - i. Note that for MPXV and WNV, additional inputs will be required for the Augur workflow
11. Specify outputs by clicking on the `outputs` tab and `use defaults`; click `save` (Fig 6)
12. Launch the workflow by clicking `run analysis`; enter desired comments and click `launch`

	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date: 10/7/2023	Workflow Version: PHB v1




Task name ↓	Variable	Type	Attribute
augur_prep	assembly	File	<input type="text" value="this.assembly_fasta"/>  {...}
augur_prep	collection_date	String	<input type="text" value="this.collection_date"/> {...}
augur_prep	continent	String	<input type="text" value="this.continent"/> {...}
augur_prep	country	String	<input type="text" value="this.country"/> {...}
augur_prep	state	String	<input type="text" value="this.state"/> {...}
augur_prep	county	String	<input type="text" value="Optional"/> {...}
augur_prep	nextclade_clade	String	<input type="text" value="Optional"/> {...}
augur_prep	pango_lineage	String	<input type="text" value="this.pango_lineage"/> {...}
prep_augur_metadata	disk_size	Int	<input type="text" value="Optional"/> {...}
prep_augur_metadata	organism	String	<input type="text" value="Optional"/> {...} 
	timezone	String	<input type="text" value="Optional"/> {...}

Figure 5.

SCRIPT .. INPUTS .. **OUTPUTS** .. **RUN ANALYSIS** 

Output files will be saved to
 Files / *submission unique ID* / augur_prep / *workflow unique ID*


References to outputs will be written to
 Tables / Augur

Fill in the attributes below to add or update columns in your data table

[Download json](#) | [Drag or click to upload json](#) | [Clear outputs](#) |

Task name ↓	Variable	Type	Attribute <input type="checkbox"/> Use defaults
augur_prep	augur_metadata	File	<input type="text" value="this.augur_metadata"/> {...}
augur_prep	augur_prep_phb_analysis_date	String	<input type="text" value="this.augur_prep_phb_an alysis_date"/> {...}
	augur_prep_phb_version	String	<input type="text" value="this.augur_prep_phb_ver sion"/> {...}

Figure 6.

	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date:	Workflow Version:
	10/7/2023	PHB v1

4.3 RUNNING THE AUGUR [ANALYSIS] WORKFLOW

1. Open the `Augur_PHB` workflow in the same workspace where `Augur_Prep_PHB` was run
2. Choose latest version of `v1` in the `version dropdown field`, or the version used for 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)
5. Click `select data` and choose the output file name from the Augur Prep workflow previously ran, e.g. `AugurPrepPHB_YYYYMMDD[initials]` (Fig 7)

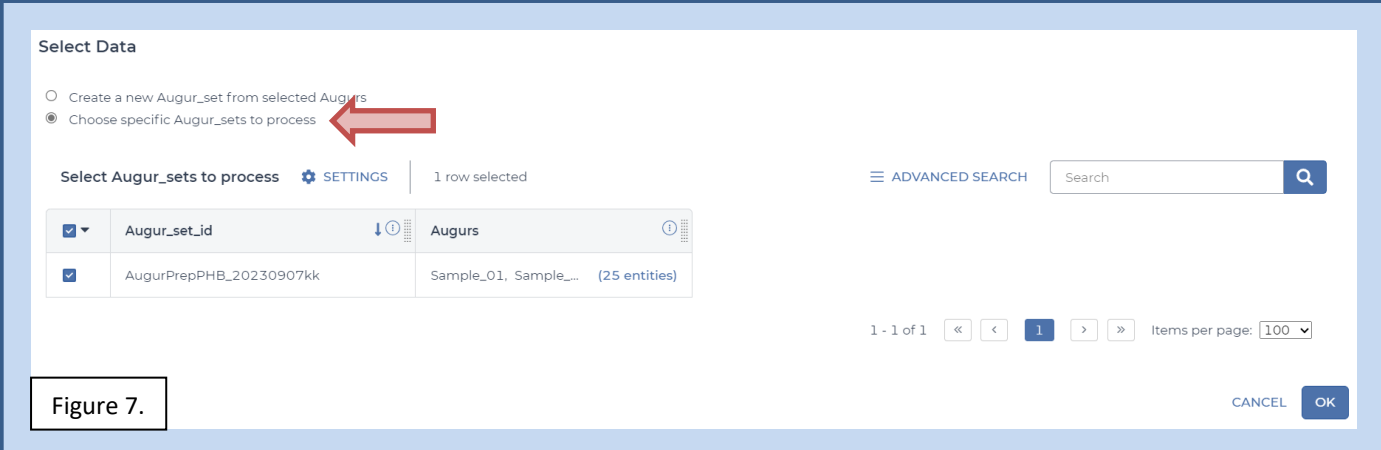



Figure 7.

6. Click on the `inputs` tab and set attributes as follows (Fig 8):
 - a. `this.Augurs.assembly_fasta`
 - b. `this.Augur_set_id`
 - c. `this.Augurs.augur_metadata`
7. If running SC2 or flu, set the organism field to `"sars-cov-2"` or `"flu"`, respectively (Fig 8)
 - a. *This sets the appropriate default values to mimic NextStrain builds for the following fields:*
 - i. See [appendix 10.1](#) for default values
 - b. For organisms other than SC2 or flu, users will need to provide the following data:
 - i. `min_num_unambig`: specify the minimum acceptable number of ambiguous basepairs in the sample consensus assembly; think of this in terms of percentage of the genome size, but input the desired value in basepairs (e.g. 150000)
 - ii. `reference_fasta`
 1. For "MPXV" (monkeypox virus): `gs://theiagen-public-files-rp/terra/augur-mpox-references/NC_063383.1.reference.fasta`
 2. For "WNV" (west nile virus): `gs://theiagen-public-files-rp/terra/augur-wnv-references/NC_063383.1.reference.fasta`
 - iii. `reference_genbank`
 1. For MPXV: `gs://theiagen-public-files-rp/terra/augur-mpox-references/NC_063383.1_reference.gb`
 2. For WNV: `gs://theiagen-public-files-rp/terra/augur-wnv-references/NC_063383.1_reference.gb`
8. Specify outputs by clicking on the `outputs` tab and `use defaults` (Fig 9); click `save`
9. Launch the workflow by clicking `run analysis`; enter desired comments and click `launch`

	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date: 10/7/2023	Workflow Version: PHB v1

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
augur	assembly_fastas	Array[File]+	<input type="text" value="this.Augurs.assembly_fasta"/> [...]
augur	build_name	String	<input type="text" value="SC2_202104_B.1.1.7s"/> [...]
augur	sample_metadata_tsvs	Array[File]+	<input type="text" value="this.Augurs.augur_metadata"/> [...]
augur	auspice_config	File	<input type="text" value="Optional"/> [...]
augur	clades_tsv	File	<input type="text" value="Optional"/> [...]
augur	distance_tree_only	Boolean	<input type="text" value="Optional"/> [...]
augur	flu_segment	String	<input type="text" value="Optional"/> [...]
augur	flu_subtype	String	<input type="text" value="Optional"/> [...]
augur	lat_longs_tsv	File	<input type="text" value="Optional"/> [...]
augur	min_num_unambig	Int	<input type="text" value="Optional"/> [...]
	organism	String	<input type="text" value="Optional"/> [...]

Figure 8.

Download json | Drag or click to upload json | Clear outputs SEARCH OUTPUTS

Task name ↓	Variable	Type	Attribute
augur	aligned_fastas	File	<input type="text" value="this.aligned_fastas"/> [...]
augur	augur_phb_analysis_date	String	<input type="text" value="this.augur_phb_analysis_date"/> [...]
	augur_phb_version	String	<input type="text" value="this.augur_phb_version"/> [...]

Figure 9.

DASHBOARD **DATA** ANALYSES WORKFLOWS JOB HISTORY

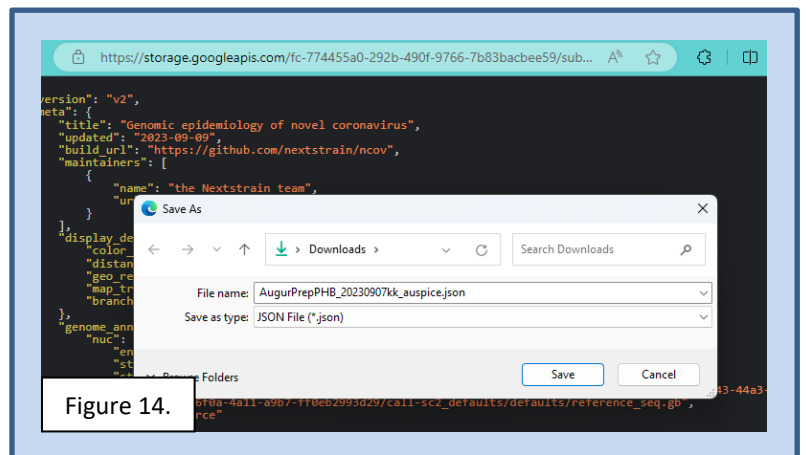
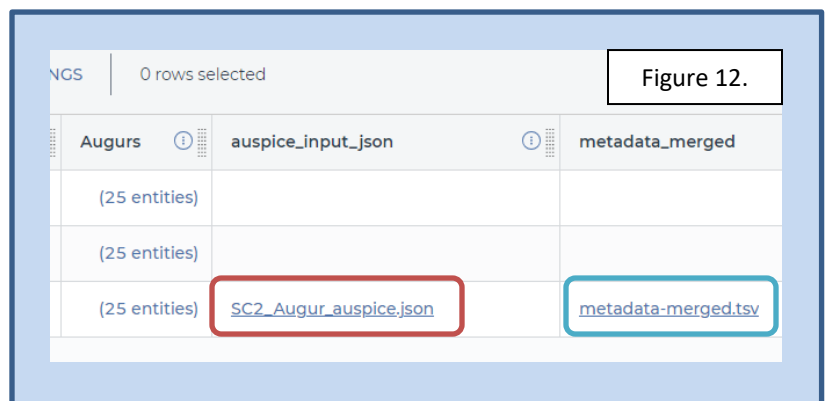
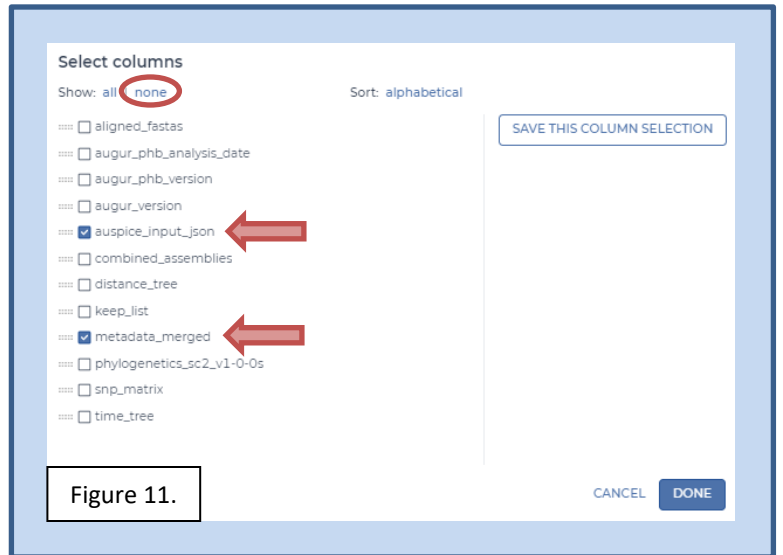
IMPORT DATA EDIT OPEN WITH... EXPORT SETTINGS 0 rows selected ADVANCED SEARCH Search


TABLES	Augur (25)	Augur_set (2)
<input type="checkbox"/>	Augur_set_id	auspice_input_json
<input type="checkbox"/>	AugurPrepPHB_20230907kk	Augur_Prep_PHB_2023-09-09T13-...

Figure 10.

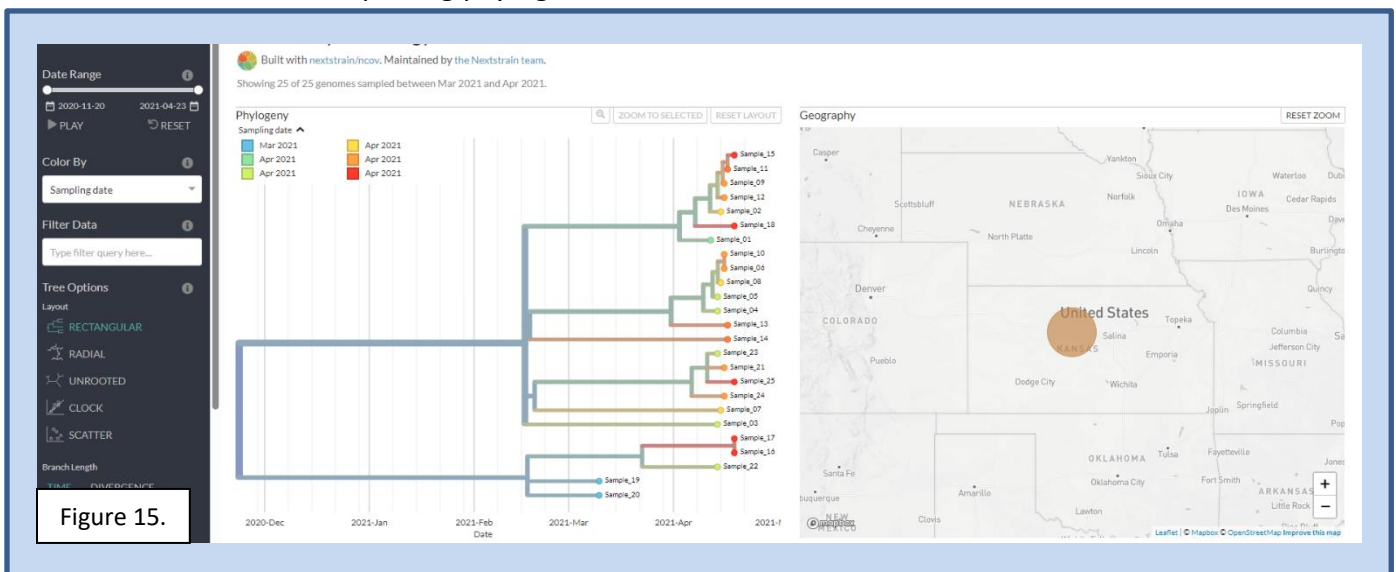
4.4 VISUALIZING THE AUGUR TREE IN AUSPICE

1. Navigate to the **workspace data tab** and select the **SET data table** containing phylogenetic data
2. Click **settings** (Fig 10) and **none** (Fig 11)
3. Select the **auspice_input_json** and **metadata_merged** columns (Fig 11)
4. **Click the auspice_input_json file** for the desired Augur Prep sample set (Fig 12)
5. In the file details pop-up, **click download for <\$0.01*** (Fig 13); this will open a new window
6. **Save** the file (Fig 14)
 - a. In windows, **either right click > save as** or **Ctrl + S**; note the destination folder and click **save**
 - b. Follow relevant commands for macintosh or linux environments
7. Return to the workspace and **click the metadata_merged file** for the relevant Augur Prep sample set (Fig 12)



	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date: 10/7/2023	Workflow Version: PHB v1

8. In the file details pop-up, **click download for <\$0.01*** (Fig 13); this will open a new window
9. **Save** the file (Fig 14)
10. In a new browser window **open <https://auspice.us/>**; **drag and drop** the auspice input json file onto the webpage
11. **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
12. In auspice, click the **color by** dropdown and **select the desired metadata attribute** to visually highlight sample groups by color
 - 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*
13. **View the timetree and map** with geographical coordinates of the analyzed strains (Fig 15)
 - a. *Guidance on interpreting phylogenetic trees²*



14. To view phylogeny relationships via SNP differences, click the `snp_matrix` hyperlink for the relevant prep sample set and download for <\$0.01
15. In excel, highlight all cells with data, click conditional formatting, color scales, and choose the color scale that makes sense to you (Fig 16)
16. *The same sample set is listed along both axis so the the SNP difference between all sample combinations is visible*
17. The fewer SNP difference between samples, the more closely related the samples
 - a. *Guidance on interpreting SNP matrices³*

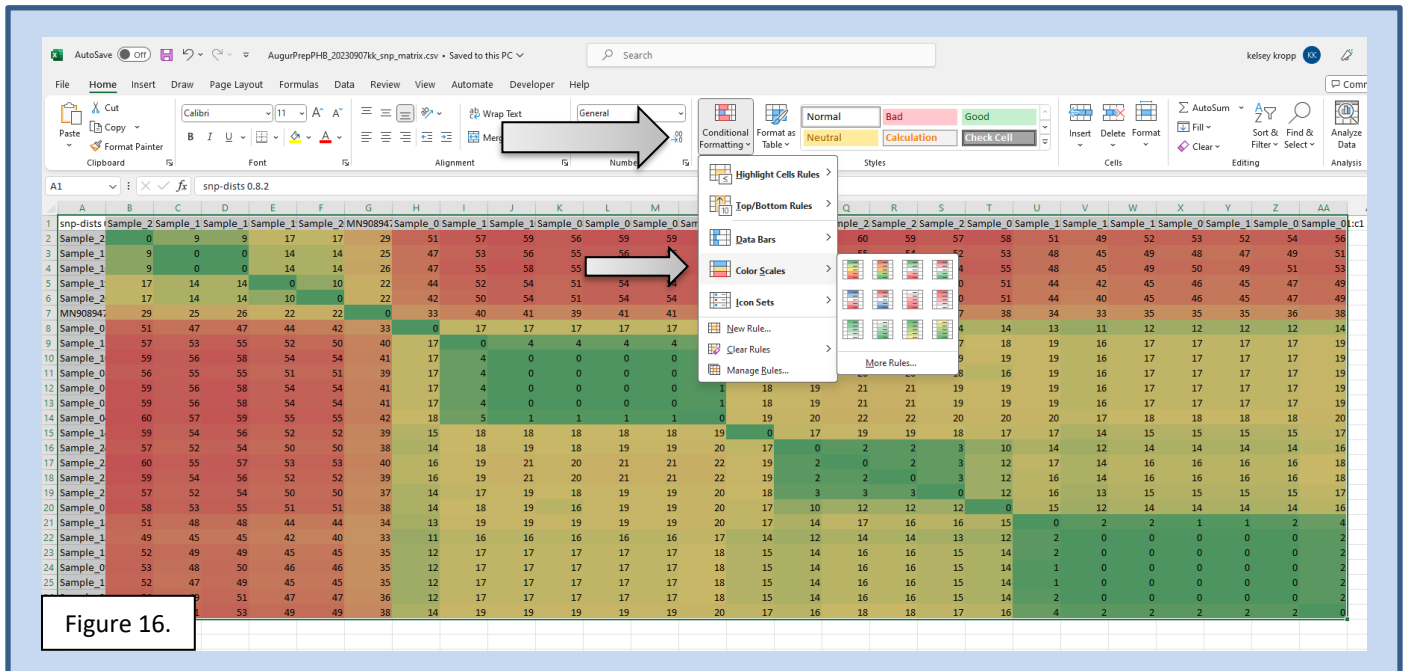


Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows

Document TG-AUGUR-01, Version 1

Date:
10/7/2023

Workflow Version:
PHB v1



5. QUALITY RECORDS

- Sample assemblies and metadata
- Pango_lineage determinations
- Auspice timetree and map
- SNP matrix

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. INTERFERENCES

N/A

8. REFERENCES

1. Wright, S., Doughty, E., and Otieno, J. (August 2023). Augur. Retrieved from <https://theiagen.notion.site/Augur-33c1307d9f28427fbed95aeeba254446> on September 7, 2023
2. "Module 1: Introduction to COVID-19." Centers for Disease Control and Prevention, URL: <https://www.cdc.gov/amd/training/covid-toolkit/module1-3.html>.
3. "NGS Result Report Interpretation Guide" Centers for Disease Control and Prevention (CDC), URL: <https://www.cdc.gov/labquality/qms-tools-and-resources.html?textFilter=result>.

	Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows	
	Document TG-AUGUR-01, Version 1	
	Date:	Workflow Version:
	10/7/2023	PHB v1

9. REVISION HISTORY

Revision	Version	Release Date
Document creation	1	10/2023



Analyzing Phylogenetic Relationships in Terra using Theiagen's Augur Workflows

Document TG-AUGUR-01, Version 1

Date:

10/7/2023

Workflow Version:

PHB v1

10. APPENDICES

10.1 AUGUR DEFAULT VALUES FOR SARS-CoV-2 AND FLU

▼ Default Values for SARS-CoV-2

- min_num_unambig = 27000
- clades_tsv = [defaults/clades.tsv](#)
- lat_long_tsv = [defaults/lat_longs.tsv](#)
- reference_fasta = [defaults/reference_seq.fasta](#)
- reference_genbank = [defaults/reference_seq.gb](#)
- auspice_config = [defaults/auspice_config.json](#)
- min_date = 2020.0
- pivot_interval = 1
- pivot_interval_units = "weeks"
- narrow_bandwidth = 0.05
- proportion_wide = 0.0

1. Default values and additional information are provided on the Augur workflow page within the Theiagen Public Health Resource Notion Site, <https://theiagen.notion.site/Augur-33c1307d9f28427fbed95aeeba254446>
 - a. Check the Notion site for the most up-to-date values and information

▼ Default Values for Flu

- lat_longs_tsv = "gs://theiagen-public-files-rp/terra/flu-references/lat_longs.tsv"
- min_num_unambig = 900
- min_date = 2020.0
- pivot_interval = 1
- narrow_bandwidth = 0.166667
- proportion_wide = 0.0
- ▼ H1N1
 - ▼ HA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_h1n1pdm_ha.gb"
 - clades_tsv = "gs://theiagen-public-files-rp/terra/flu-references/clades_h1n1pdm_ha.tsv"
 - ▼ NA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_h1n1pdm_na.gb"
- auspice_config = "gs://theiagen-public-files-rp/terra/flu-references/auspice_config_h1n1pdm.json"
- ▼ H3N2
 - ▼ HA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_h3n2_ha.gb"
 - clades_tsv = "gs://theiagen-public-files-rp/terra/flu-references/clades_h3n2_ha.tsv"
 - ▼ NA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_h3n2_na.gb"
- auspice_config = "gs://theiagen-public-files-rp/terra/flu-references/auspice_config_h3n2.json"
- ▼ Victoria
 - ▼ HA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_vic_ha.gb"
 - clades_tsv = "gs://theiagen-public-files-rp/terra/flu-references/clades_vic_ha.tsv"
 - ▼ NA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_vic_na.gb"
- auspice_config = "gs://theiagen-public-files-rp/terra/flu-references/auspice_config_vic.json"
- ▼ Yamagata
 - ▼ HA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_yam_ha.gb"
 - clades_tsv = "gs://theiagen-public-files-rp/terra/flu-references/clades_yam_ha.tsv"
 - ▼ NA
 - reference_fasta = "gs://theiagen-public-files-rp/terra/flu-references/reference_yam_na.gb"
- auspice_config = "gs://theiagen-public-files-rp/terra/flu-references/auspice_config_yam.json"