

Running Theiagen's Snippy_V	ariants_PHB Workflow in Terra
Document TG-S	SNP-01, Version 1
Date:	Workflow Version:
01/21/2025	PHB v2.3.0

1. PURPOSE/SCOPE

This procedure describes the process of running Theiagen's Snippy_Variants_PHB pipeline using the Terra platform. Acceptable data types include single-end (SE) or paired-end (PE) reads in FASTQ format, or assembled sequences in FASTA format. A reference genome is provided as a workflow input for alignment to identify single-nucleotide polymorphisms (SNPs), multi-nucleotide polymorphisms (MNPs), and insertions/deletions (indels). An annotated GenBank file may be used as a reference genome together with user-specified genes of interest (as workflow inputs) to output mutations identified with the associated gene or annotation names.

2. REQUIRED RESOURCES

- Computer
- Internet connection: at least 10 and 5Mbps for download and upload speed, respectively
- Internet browser
 - Google Chrome, Firefox, or Edge
- Google account
- Terra account, linked to Google account
- Illumina SE or PE raw sequencing read files uploaded to Terra workspace
- Theiagen's Snippy_Variants_PHB workflow in Terra

3. RELATED DOCUMENTS

Document Number	Document Name
TC TED 02	Uploading Local or SRA NGS Data & Creating a
IG-IER-03	Results Metadata Table in Terra

4. PROCEDURE

Prior to running the Snippy_Variants_PHB workflow, the workflow must be imported into the Terra workspace and a reference genome file must be available to use for analysis. For details see **Appendix 10.1** and **10.2**, respectively.

4.1 Configure and Run the Snippy_Variants_PHB Workflow

- 1. Open Terra, navigate to the *workflows* tab, and select the *Snippy_Variants_PHB* workflow (Fig 1).
 - a. To import the workflow for the first time, refer to Appendix 10.1.

	Running Theiagen's Snippy_Variants_PHB Workflow in Terr			
	Document TG-SNP-01, Version 1			
	Date: Workflow Version:			
Z / [01/21/2025	PHB v2.3.0		

	Workspaces > theiagen-training-workspaces > Workflows	aces/Theiagen_Kropp_Sandbox >
DASHBOARD DATA ANAL	YSES WORKFLOWS JOB HISTOR	Y
WORKFLOWS		Ļ
Find a Workflow	Snippy_Streamline_PHB	Snippy_Variants_PHB
Figure 1.	V. v1.0.0 Source: Dockstore	V. v2.3.0 Source: Dockstore

- 2. In the version dropdown field, *select the workflow version* that was internally validated, or the latest version of the workflow (Fig 2).
- 3. Uncheck call caching (Fig 2).
- 4. Select the second bullet to *run workflow(s) with inputs defined by data table* (Fig 2, b).
- 5. Choose the relevant *data table* under the select data table dropdown (Fig 2, c).
- 6. Click *select data* (Fig 2, d).

← Back to list
Snippy_Variants_PHB
Version: v2.3.0 v 🦨 🧟
Source: github.com/theiagen/public_health_bioinformatics/Snippy_Variants_PHB:v2.3.0
Synopsis:
No documentation provided
O Run workflow with inputs defined by file paths
Run workflow(s) with inputs defined by data table
>1 Step 2
C PlatformAccuracy V Select Data
□ Use call caching ③ □ Use reference disks ③ ☑ Ignore emoty outpute ④ □ Delete intermediate outputs ④ □ Retry with more memory ④ □ Resource Figure 2.

- 7. In the pop-up window, *select each sample checkbox* to include in the analysis (Fig 3).
 - a. Click the down arrow and select all to process all specimens.
 - b. 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*.

	Document TG-SNP-01, Version 1		
	Date:	Workflow Version:	
Z / [01/21/2025	PHB v2.3.0	

Selec	t PlatformAccuracys to proce	ss 🔅 Settings 1 row selected	Advanced Search Search	٩
•	PlatformAccuracy_id	readl	read2	snippy_variants_bai
	SAbaetetubaControl	SAbaetetubaControl_S5_L001_R1	SAbaetetubaControl_S5_L001_R2	SAbaetetubaControl.bam.bai
	SAbaetetubaControl-2	SAbaetetubaControl_S5_L001_R1	SAbaetetubaControl_S5_L001_R2	SAbaetetubaControl-2.bam.ba
	SAbaetetubaControl-3	SAbaetetubaControl_S5_L001_R1	SAbaetetubaControl_S5_L001_R2	SAbaetetubaControl-3.bam.ba
	SAMN01813419	SRR957989_1.fastq.gz		SAMN01813419.bam.bai
		¢		

8. In the first value field for the reference_genome_file variable, *click the folder icon* (Fig 4).

SCRIPT •• INPUTS ••	OUTPUTS •• Run Analysis		Cancel Save
Hide optional inputs	Download json Drag or clic	k to upload json Clea	IT INPUTS
Task name 🌡	Variable	Туре	Input value
snippy_variants_wf	reference_genome_file	File	Required []
Figure 4	samplename	String	this.PlatformA Browse bucket files
- igure -			

- 9. *Select the appropriate reference genome file* from the workspace (Fig 5).
 - a. To import reference genomes for the first time, see Appendix 10.2.

Choose input file	\otimes
Files /	
Name	
.data-table-versions/	
Assemblies/	
ColumnTranslations/	
FASTAs/	
SD. PhoenixValidationCriteria txt	
Salmonella_enterica_ATCC_35640	Figure 5.



Running Theiagen's Snippy_Variants	_PHB Workflow in Terra	
Document TG-SNP-01,	Version 1	

Becament 10 Sit	
Date:	Workflow Version:
01/21/2025	PHB v2.3.0

- 10. Set the second input value for the samplename variable to this.PlatformAccuracy_id, where "PlatformAccuracy" is the name of the data table to be analyzed (Fig 6).
 - a. The samplename variable input should appear in the dropdown when clicking in the input value field. If it does not appear, check the data table name selected for analysis.

Task name 🖡	Variable	Туре	Input value
snippy_variants_wf	reference_genome_file	File	"gs://fc-774455a0- 292b-490f-9766- 7b83bacbee59/Salmo 🗁 () nella_enterica_ATCC_3 5640"
snippy_variants_wf	samplename	String	this.PlatformAccuracy_id {}
Figure 6. ^{Jery}	cpu		this.PlatformAccuracy_id

11. Scroll down through the input variable list and set the read1 and read2 variables to this.read1 and this.read2, respectively where read1 and read2 are the data table column names containing read files (Fig 7).

query_gene	String	"catB3" {}
readl	File	this.read1
Figure 7.	File	this.read2

- 12. Optional: In quotations, input a gene name or other annotation(s) of interest into the query_gene input variable; e.g. "catB3" (Fig 7).
 - a. The text must match the GenBank file exactly.
- 13. Specify outputs by clicking on the *outputs* tab and *use defaults* (Fig 8).
- 14. Click save (Fig 8).

Running Theiagen's Snippy_Variants_PHB Workflow in Terr	
Document TG-SNP-01, Version 1	
Date:	Workflow Version:
01/21/2025	PHB v2.3.0

SCRIPT •• INPUT Output files will be saved to D Files / submission unique ID / snip	rs •• OUTPUTS •• Run Analyzis	D	
References to outputs will be written Tables / PlatformAccuracy Fill in the attributes below to add or u	to ipdate columns in your data table		Download json Drag or click to upload json Clear outputs SEARCH OUTPUTS
Task name 🖡	Variable	Туре	Input value Use defaults
snippy_variants_wf	snippy_variants_bai	File	thissnippy_variants_bai []
snippy_variants_wf	snippy_variants_bam	File	thissnippy_variants_bam []
snippy variants wf	snippy_variants_coverage_tsv	File	Optional []
Figure 8		Chrise	

15. Click *run analysis*, *enter comments* as needed, and select *launch* (Fig 9).

Select Data 1 selected Plat reference disks Ignore empty of Resource monitors with more memory Resource monitors OUTPUTS Image: Run Analysis ts_wf / workflow unique ID	for Confirm launch Output files will be sa us us-central1 (lowa) Running workflows w How much does my v • You are launching Describe your submis	ved as workspace data in: ill generate cloud charges. workflow cost? g* 1 workflow run in this submission. sion (optional):
umns in your data table	Enter comment for	the submission
umns in your data table	Enter comment for	the submission
umns in your data table Variable snippy_variants_bai	Enter comment for	the submission
umns in your data table Variable snippy_variants_bai snippy_variants_bam	File File	the submission Launch this snippy_variants_bai this snippy_variants_bam



Running Theiagen's Snippy_Va	riants_PHB Workflow in Terra
Document TG-SI	NP-01, Version 1
Date:	Workflow Version:
01/21/2025	PHB v2.3.0

5. QUALITY RECORDS

- Raw read files
- Workflow version and input parameters
- Reference sequence and metadata, as applicable
- Sample read, assembly, and result-specific QC metrics
- All workflow outputs relevant to results, including tool and database versions

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

- "Mutations identified by the Snippy_Variants_PHB workflow are highly dependent on the choice of reference genome. Mutations cannot be identified in genomic regions that are present in your query sequence and not the reference."¹
- "The outputs from samtools coverage (found in the snippy_variants_coverage_tsv file) may differ from the snippy_variants_percent_ref_coverage due to different calculation methods. samtools coverage computes genome-wide coverage metrics (e.g., the proportion of bases covered at depth ≥ 1), while snippy_variants_percent_ref_coverage uses a user-defined minimum coverage threshold (default is 10), calculating the proportion of the reference genome with a depth greater than or equal to this threshold."¹

8. **REFERENCES**

¹"Phylogenetic Construction – Snippy Variants." Theiagen Public Health Bioinformatics, Theiagen Genomics, 3 Dec. 2024, <u>https://theiagen.github.io/public_health_bioinformatics/latest/workflows/phylogenetic_cons</u> <u>truction/snippy_variants/?h=snip</u>. Accessed 23 Jan. 2025.

9. **REVISION HISTORY**

Revision	Version	Release Date
Document creation	1	1/2025



Running Theiagen's Snippy_Va	riants_PHB Workflow in Terra
Document TG-SN	P-01, Version 1
Date:	Workflow Version:

Date: 01/21/2025

PHB v2.3.0

10. APPENDICES

10.1 Find and Import the Snippy_Variants_PHB Workflow

- 1. Navigate to the *workflows* tab of the workspace (Fig 10).
- Users that already have the Snippy_Variants_PHB workflow in their workspace can <u>select the</u> workflow (Fig 10) and proceed to the Configure and Run the Snippy_Variants_PHB Workflow section of this SOP.

	KSPACES Worl	kspaces > theiagen-tra prkflows	aining-workspa	ices/Theiagen_Kropp_Sandbox :	
DASHBOARD DAT	A ANALYSES	WORKFLOWS	JOB HISTOR	(
WORKFLOWS		\smile			
Find a Workflow	Snip	py_Streamline_PHE	3	Snippy_Variants_PHB	
Figure 10.	V.vl. Source	0.0 :e: Dockstore	:	V. v2.3.0 Source: Dockstore	:

- a. To import the workflow, click *find a workflow* (Fig 10).
- b. In the pop-up window, click *Dockstore.org* in the bottom left (Fig 11).
- c. Click *Organizations* in the banner at the top and search for *Theiagen* using the search box (Fig 12).

iagen-training-workspaces/Theiagen_Kropp_Sandbox >	arch 🕂 Organizations 📾 Abor	ut 👔 Docs 🖳 Forum Login Register
Find a workflow		Sort by Search Organizations
A community repository of best practice workflows A report that offers integration with GitHub. Figure 11.	* 35 Now Comission ate a registry of best practice Figure 12.	C E N O M I C S G E N O M I C S Theiagen Genomics Public health bioinformatics for pathogen surveillance bitps://www.theiagen.com Denver, co USA jeol.sevinsky@theiagen.com



Running Theiagen's Snippy_	_Variants_PHB Workflow in Terra
Document TG	G-SNP-01, Version 1
Date:	Workflow Version:
01/21/2025	PHB v2.3.0

d. Click the *Public Health Bioinformatics (PHB)* collection (Fig 13)and using *ctrl* + f on Windows search for *snippy* (Fig 14).



- e. Click on the *Terra icon* (Fig 15) to import the workflow into a Terra workspace.
- f. Select the workspace in the destination workspace dropdown field and click *Import* (Fig 16).

nippy_Variants_PHB:v	1.0. 🚖 0
DAG Metrics	Launch with
Export this workflow version to Terra.	Terra
riants_PHB	- Geolandi
Figure 15.	http://www.second.com/second/second

Importing fro	m Dockstore	Workflow Name
github.com/theiag v.v1.0.0	en/public_health_bioinformatics/Snippy_Variants_PHB	Snippy_Variants_PHB
Please note: Dock this Workflow will	itore cannot guarantee that the WDL and Docker image referenced by not change. We advise you to review the WDL before future runs.	Destination Workspac
	-	Select a workspace



Running Theiagen's Snippy_Variants_PHB Workflow in Terra		
Document TG-SNP-01, Version 1		
Date:	Workflow Version:	
01/21/2025	PHB v2.3.0	

10.2 IMPORT THE REFERENCE GENOME FASTA FILE

1. In the Terra workspace data tab, click on the *folder* icon on the right-hand side (Fig 17).



- 2. Click *upload* and select the corresponding reference genome file in fasta (e.g. .fa, .fasta) or full GenBank (.gbk) file format (Fig 18).
- 3. Choose the relevant file and click *open* to import the file.

<u>Files</u> / <u>ReferenceSeqs</u>	
□ Name	Size
Figure 18	3 KB
ligure 10.	