

Nov 12, 2021

Running the Titan_ClearLabs Workflow on Terra.bio

Titan Clear Labs SARS-CoV-2 Strain Characterization Workflow for the Terra Platform

Frank J Ambrosio¹, Jill V Hagey², Kevin Libuit¹,
Technical Outreach and Assistance for States Team²

¹Theiagen Genomics; ²Centers for Disease Control and Prevention

1



dx.doi.org/10.17504/protocols.io.bzygp7tw

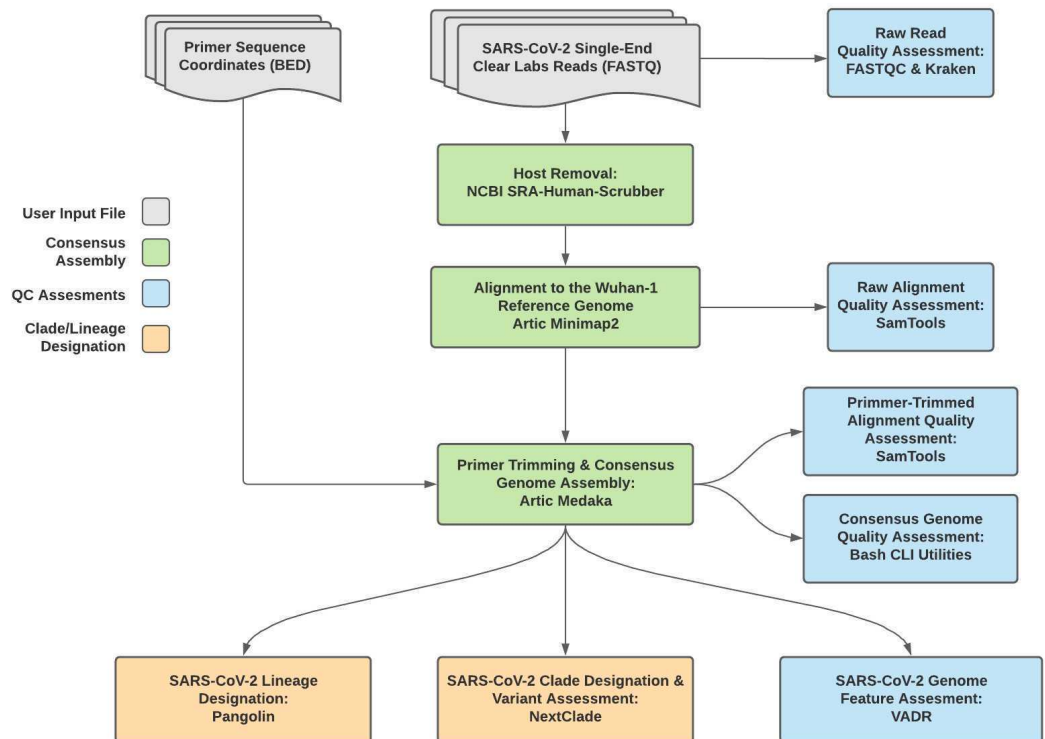
Theiagen

Francis Ambrosio

The opinions expressed here do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. The protocol content here is under development and is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

The Titan_ClearLabs workflow is a part of the Public Health Viral Genomics Titan series for SARS-CoV-2 genomic characterization. Titan_ClearLabs was written to process Clear Labs read data for SARS-CoV-2 ARTIC V3 amplicon sequencing. Upon initiating a Titan_ClearLabs run, input read data provided for each sample will be processed to perform consensus genome assembly, infer the quality of both raw read data and the generated consensus genome, and assign lineage or clade designations as outlined in the Titan_ClearLabs data workflow below.

Additional technical documentation for the Titan_ClearLabs workflow is available at: https://public-health-viral-genomics-theiagen.readthedocs.io/en/latest/titan_workflows.html#titan-clearlabs



Titan workflow for use with Clear Labs sequencing read data

Required input data for Titan_ClearLabs:

Cear Labs FASTQ read files (single FASTQ file per sample)

Primer sequence coordinates of the PCR scheme utilized in BED file format

Video Instruction:

Theiagen Genomics: Titan Genomic Characterization

<https://www.youtube.com/watch?v=zP9I1r6TNrw>

Theiagen Genomics: Titan Outputs QC

<https://www.youtube.com/watch?v=Amb-8M71umw>

For technical assistance, please contact us at: **TOAST@cdc.gov**

Titan_ClearLabs.png

DOI

[dx.doi.org/10.17504/protocols.io.bzygp7tw](https://doi.org/10.17504/protocols.io.bzygp7tw)

Frank J Ambrosio, Jill V Hagey, Kevin Libuit, Technical Outreach and Assistance for States Team 2021. Running the Titan_ClearLabs Workflow on Terra.bio.

protocols.io

<https://dx.doi.org/10.17504/protocols.io.bzygp7tw>



Titan Clear Labs SARS-CoV-2 Strain Characterization Workflow for the Terra Platform, Frank Ambrosio

ONT, Nanopore, SARS-CoV-2, MinION, GridION, PromethION, MK1C, Pangolin, Genomics, Analysis, Virology, Bioinformatics, RNA, DNA, Clear Labs, Covid, Computational Biology, Sequencing, Clear Labs, Hamilton

protocol ,

Nov 10, 2021

Nov 12, 2021

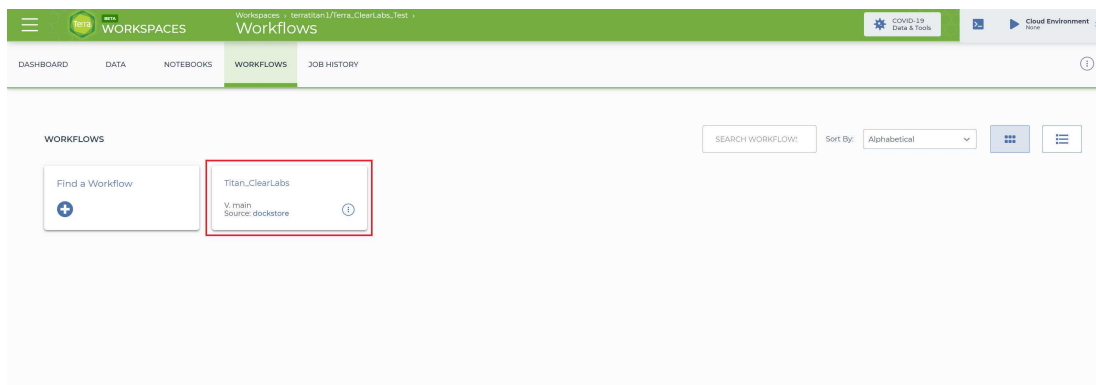
55016

:

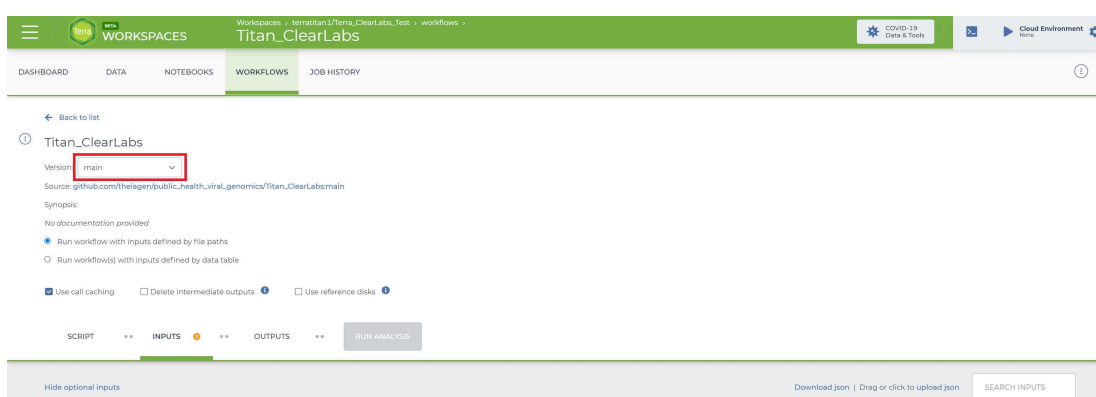
The opinions expressed here do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. The protocol content here is under development and is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](https://dx.doi.org/10.17504/protocols.io) is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with [protocols.io](https://dx.doi.org/10.17504/protocols.io), can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

Running the Titan_ClearLabs Workflow

- 1 To run the Titan_ClearLabs workflow, click on the 'Workflows' panel in the newly created workspace. It should bring you to your workflow page. Click on the 'Titan_ClearLabs' tile to bring up the Titan_ClearLabs assembly workflow page (if you do not see the Titan_ClearLabs workflow in your workspace, please see our video on importing a workflow to Terra.bio: <https://youtu.be/ZRpQyIIDMzo>).

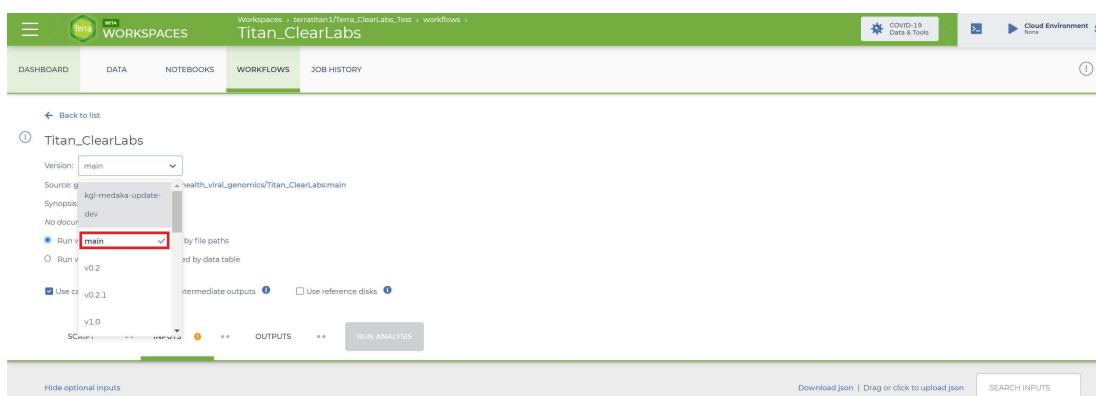


This will bring up the Titan_ClearLabs workflow page:



The Titan ClearLabs assembly workflow page

Select the version of the workflow you would like to run. **Double check that you are using the latest version of the workflow.** Alternately, you may specify another version, **but should only pick a stable version (with numbers) NOT a 'main' or 'dev' version.**



Ensure that "Run workflow(s) with inputs defined by data table" is selected and the "Use call caching" is checked and then select the root entity type for the data you wish to analyze

NOTE: Call caching allows Terra to identify and skip jobs that have been run previously; this option is by default enabled to avoid unnecessary compute costs. More information on Terra call caching, including examples of when you may want to disable this feature, is available through the [Terra Support Documentation](#).

← Back to list

ⓘ Titan_ONT

Version: v1.4.2

Source: github.com/theiagen/public_health_viral_genomics/Titan_ONT:v1.4.2

Synopsis:

No documentation provided

☐ Run workflow with inputs defined by file paths

☒ Run workflow(s) with inputs defined by data table

Step 1

Select root entity type: Test_ONT_sample

☒ Use call caching ☐ Delete intermediate outputs ⓘ ☐ Use reference disks ⓘ

Step 2

SELECT DATA No data selected

Click "SELECT DATA" and choose the samples you wish to analyze

← Back to list

ⓘ Titan_ONT

Version: v1.4.2

Source: github.com/theiagen/public_health_viral_genomics/Titan_ONT:v1.4.2

Synopsis:

No documentation provided

☐ Run workflow with inputs defined by file paths

☒ Run workflow(s) with inputs defined by data table

Step 1

Select root entity type: Test_ONT_sample

☒ Use call caching ☐ Delete intermediate outputs ⓘ ☐ Use reference disks ⓘ

Step 2

SELECT DATA No data selected

SCRIPT INPUTS OUTPUTS RUN ANALYSIS

Complete the INPUTS form with the appropriate attributes (if you have not yet uploaded any Clear Labs data to Terra.bio please reach out to the Theiagen Genomics bioinformatics support team [support@terrapublichealth.zendesk.com] and we will help you to establish Terra.bio integration with the Clear Labs portal")

The top two rows represent variables that have to be provided by the user. This was the information that we populated the sample data table with in the previous step.

In our example, for the first row, the 'clear_lab_fastq' variable, we clicked on the 'Attribute' text box and wrote 'this.clearlabs_fastq_gz' to indicate that the 'clear_lab_fastq' we wish to analyze are under the 'clearlabs_fastq_gz' column of our selected datatable. In the second row we input our primer_bed file of choice, which should be saved as a workspace element in Terra (if you do not have this saved as a workspace element please see our video on uploading a workspace element: <https://youtu.be/Orbz7uRtwmQ>). In the third row, the 'samplename' variable, we selected on the 'Attribute' text box and wrote 'this.clearlabs_specimen_id' to indicate the 'samplename' of each sample we are analyzing can be found in the 'Test_ONT_sample_id' column of our selected datatable.

NOTE: If you named your columns something other than reads then just type "this." followed by whatever the column name is. We would advise naming your reads column "reads" for clarity.

Task name	Variable	Type	Attribute
titan_clearlabs	clear_lab_fastq	File	this.clearlabs_fastq_gz
titan_clearlabs	primer_bed	File	workspace/V3_Artic_Primers
titan_clearlabs	samplename	String	this.clearlabs_specimen_id

Once your input form is complete, move on to the OUTPUTS form and select "Use Defaults". Terra will then populate the OUTPUTS form with all of the default outputs options generated by the workflow. If you forget to do this you won't have easily accessible results! **Save these changes by clicking the 'Save' button.**

Task name	Variable	Type	Attribute
titan_ont	aligned_bai	File	this.aligned_bai
titan_ont	aligned_bam	File	this.aligned_bam
titan_ont	amp_coverage	File	this.amp_coverage
titan_ont	artic_version	String	this.artic_version
titan_ont	assembly_fasta	File	this.assembly_fasta

Once your INPUTS and OUTPUTS forms are complete, click the 'Save' button on the top right-hand side of the page. The yellow caution icons should disappear and the Run Analysis option should be made available.

Titan_ONT

Version: v1.4.2

Source: github.com/theigen/public_health_viral_genomics/Titan_ONT

Synopsis: No documentation provided

☐ Run workflow with inputs defined by file paths

☒ Run workflow(s) with inputs defined by data table

Step 1

Select root entity type: Test_ONT_sample

☒ Use call caching ☐ Delete intermediate outputs ☐ Use reference disks

Step 2

SELECT DATA 2 selected Test_ONT_samples [will create a new Test_ONT_sample_set named 'Titan_ONT_2021-05-10T14-51-13']

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

Output files will be saved to: Files / submission unique ID / titan_ont / workflow unique ID

References to outputs will be written to: Tables / Test_ONT_sample

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

You are now ready to run the Titan_ClearLabs workflow! Click on the 'Run Analysis' button to the right of the 'Outputs' tab. A popup window should appear titled 'Confirm launch'. If the 'Run Analysis' button is greyed out, you need to save your recent changes by clicking the 'Save' button.

Titan_ONT

Version: v1.4.2

Source: github.com/theigen/public_health_viral_genomics/Titan_ONT

Synopsis: No documentation provided

☐ Run workflow with inputs defined by file paths

☒ Run workflow(s) with inputs defined by data table

Step 1

Select root entity type: Test_ONT_sample

☒ Use call caching ☐ Delete intermediate outputs ☐ Use reference disks

Step 2

SELECT DATA 2 selected Test_ONT_samples [will create a new Test_ONT_sample_set named 'Titan_ONT_2021-05-10T14-51-13']

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

Output files will be saved to: Files / submission unique ID / titan_ont / workflow unique ID

References to outputs will be written to: Tables / Test_ONT_sample

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

Confirm launch

This analysis will be run by Cromwell 61.

Output files will be saved as workspace data in:

multi-region: US

This will launch 2 analyses.

CANCEL **LAUNCH**

Clicking the 'Launch' button should bring you to the 'Job History' panel where each sample will be queued for the Titan_ClearLabs analysis. The status will change from queued to submitted to running.

WORKSPACES Workspaces > theigen-validations/Terra_ONT_Test > Job History

DASHBOARD **DATA** **NOTEBOOKS** **WORKFLOWS** **JOB HISTORY**

Job History > Submission 3a86c2df-0044-4541-9f73-5288338d9838

Workflow Statuses Submitted 2

Workflow Configuration theigen-validations/Titan_ONT

Data Entity Titan_ONT_2021-05-10T14-51-13 Test_ONT_sample_set

Delete Intermediate Outputs Disabled

Submitted by kevinlibuit@theigen.com May 10, 2021, 8:05 AM

Submission ID 3a86c2df-0044-4541-9f73-5288338d9838

Use Reference Disks Disabled

Total Run Cost N/A

Call Caching Enabled

Search Completion status

Data Entity	Last Changed	Status	Run Cost	Messages	Workflow ID	Links
F_1 [Test_ONT_sample]	May 10, 2021, 8:05 AM	Queued	N/A			
F_2 [Test_ONT_sample]	May 10, 2021, 8:05 AM	Queued	N/A			

Job history screen after launching Terra job

Any trouble at all please feel free to check out our video on Running the Titan_ClearLabs Workflow, or reach out to our support email: support@terrapublichealth.zendesk.com