

Nov 12, 2021

# Running the Titan\_ONT Workflow on Terra.bio

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

Francis J Ambrosio<sup>1</sup>, Jill V Hagey<sup>2</sup>, Kevin Libuit<sup>1</sup>,  
Technical Outreach and Assistance for States Team<sup>2</sup>

<sup>1</sup>Theiagen Genomics; <sup>2</sup>Centers for Disease Control and Prevention

1



[dx.doi.org/10.17504/protocols.io.bzyip7ue](https://dx.doi.org/10.17504/protocols.io.bzyip7ue)

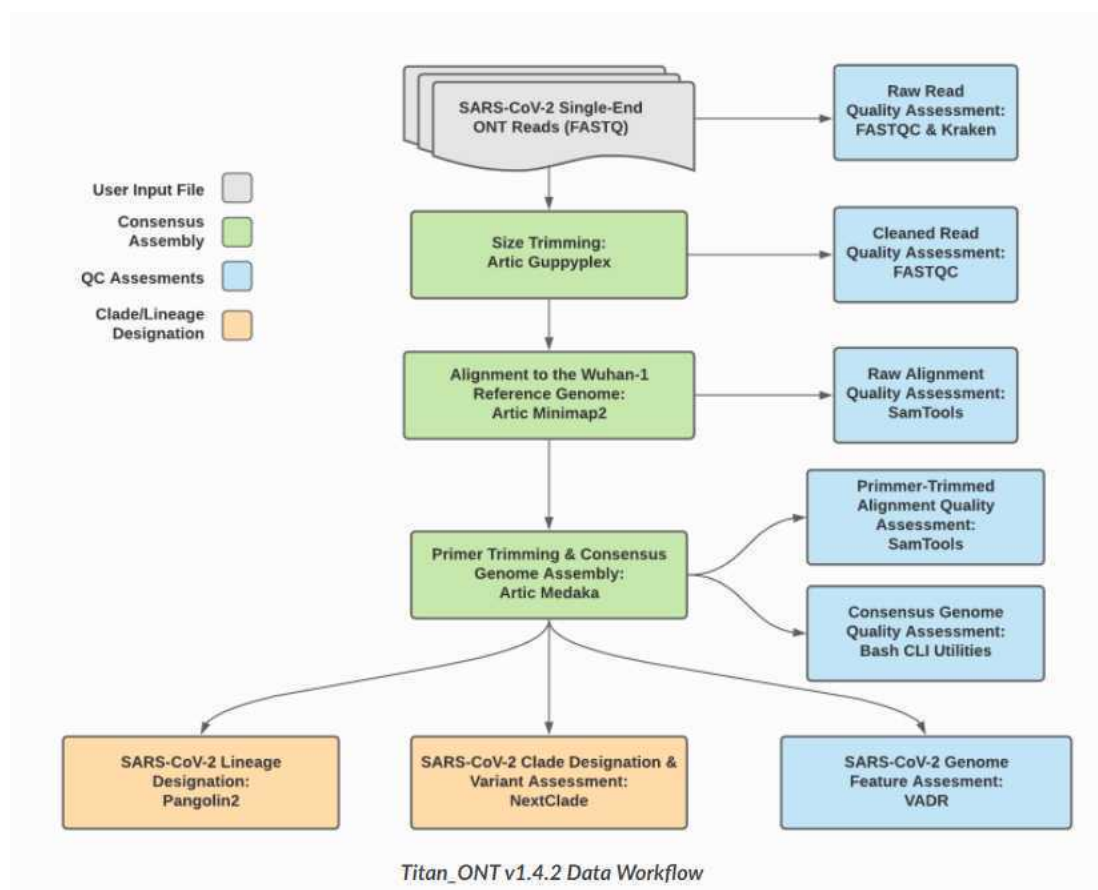
Theiagen

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legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](https://protocols.io) is not peer reviewed and may not have undergone a formal approval of any kind.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://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\_ONT workflow is a part of the Public Health Viral Genomics Titan series for SARS-CoV-2 genomic characterization. Titan\_ONT was written specifically to process basecalled and demultiplexed Oxford Nanopore Technology (ONT) read data. Input reads are assumed to be the product of sequencing ARTIC V3 tiled PCR-amplicons designed for the SARS-CoV-2 genome. Upon initiating a Titan\_ONT 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\_ONT data workflow diagram below.

Additional technical documentation for the Titan\_ONT workflow is available at: [https://public-health-viral-genomics-theiagen.readthedocs.io/en/latest/titan\\_workflows.html#titan-workflows-for-genomic-characterization](https://public-health-viral-genomics-theiagen.readthedocs.io/en/latest/titan_workflows.html#titan-workflows-for-genomic-characterization)



Titan workflow for use with Oxford Nanopore sequencing read data

### Required input data for Titan\_ONT:

Basecalled and demultiplexed ONT read data files (single FASTQ file per sample)

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

*Titan\_ONT has not been written to process FAST5 files*

### 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\_ONT.png

DOI

[dx.doi.org/10.17504/protocols.io.bzyip7ue](https://dx.doi.org/10.17504/protocols.io.bzyip7ue)

Francis J Ambrosio, Jill V Hagey, Kevin Libuit, Technical Outreach and Assistance for States Team 2021. Running the Titan\_ONT Workflow on Terra.bio. **protocols.io**  
<https://dx.doi.org/10.17504/protocols.io.bzyip7ue>



## Titan ONT SARS-CoV-2 Strain Characterization Workflow for the Terra Platform, Technical Outreach and Assistance for States Team

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

\_\_\_\_\_ protocol ,

Nov 10, 2021

Nov 12, 2021

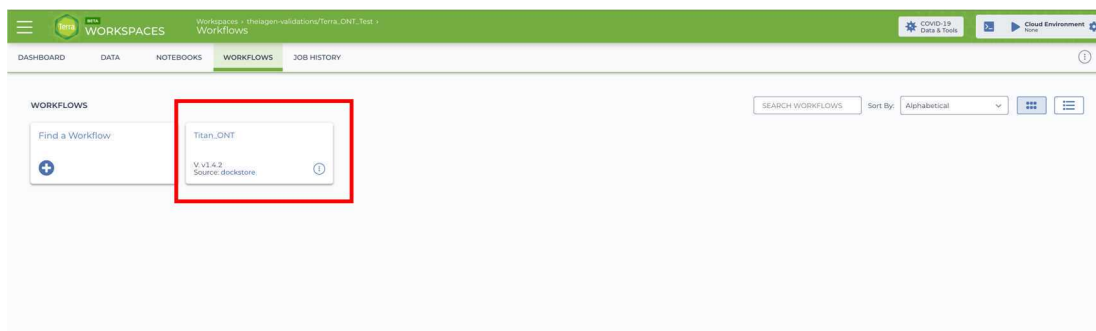
55018

:

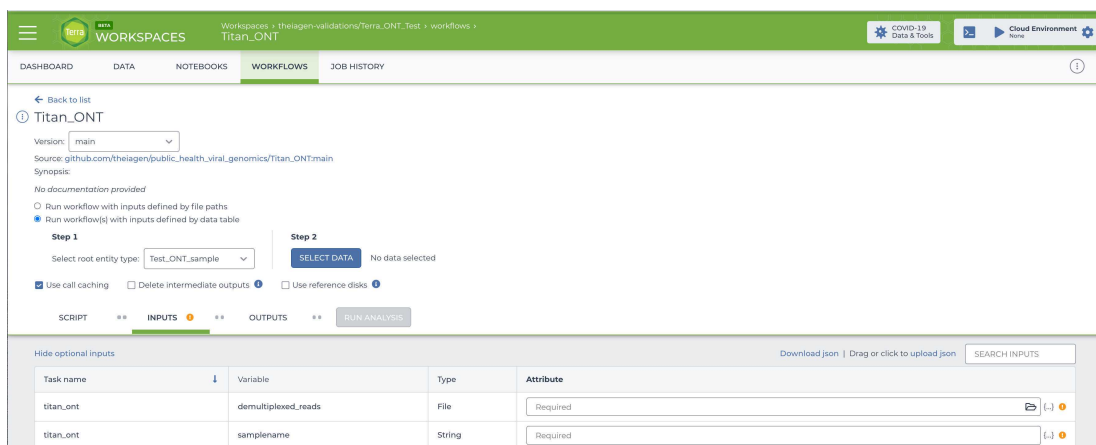
The opinions expressed here do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. The protocol content here is under development and is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](https://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://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\_ONT Workflow

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

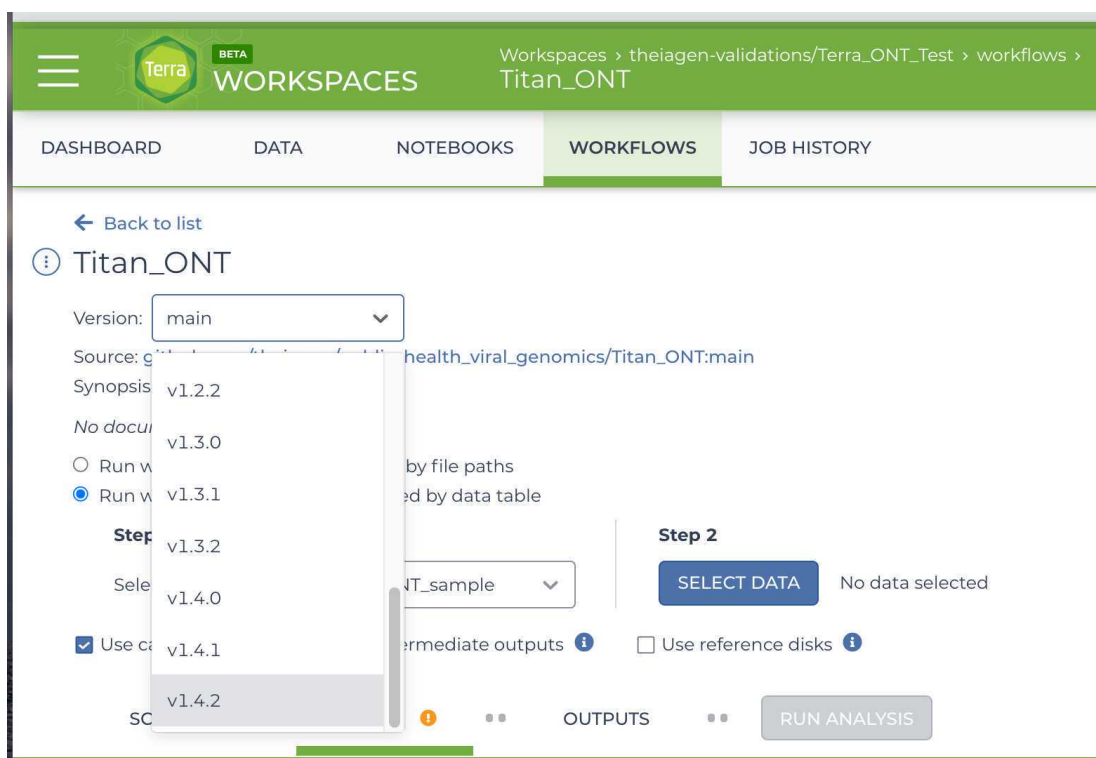


This will bring up the Titan\_ONT workflow page:



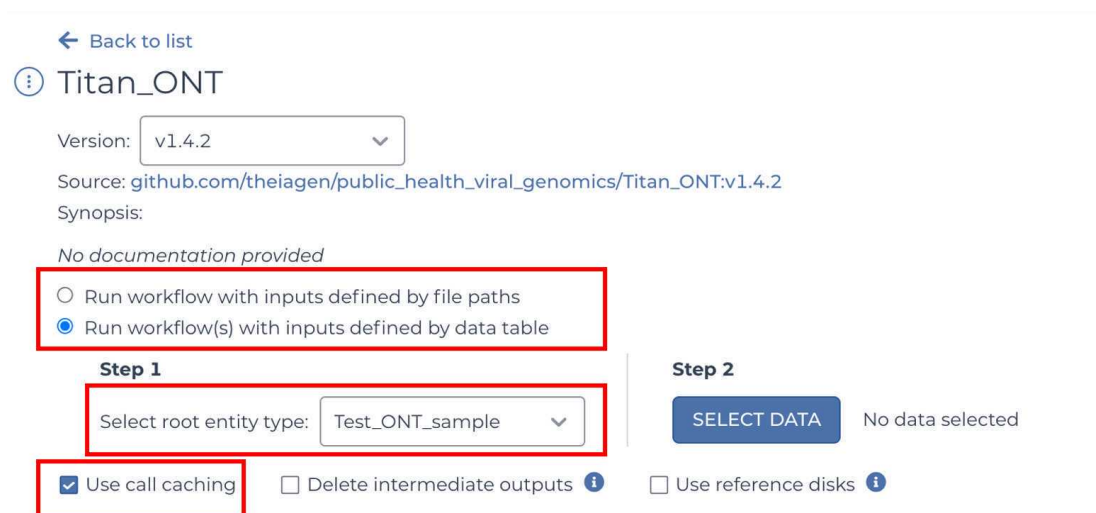
The Titan ONT assembly workflow page

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. Preferably the latest version.**



Ensure that "Run workflow(s) with inputs defined by data table" is selected. Select the root entity type for the data you wish to analyze. If you would like to check for previous runs using the same inputs and configuration, check the "Use call caching" box.

**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](https://terra.bio/docs/faq/#call-caching).



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](https://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 ⓘ

**Step 2**

No data selected

☐ Use reference disks ⓘ

SCRIPT    INPUTS    OUTPUTS   

Complete the INPUTS form with the appropriate attributes (if you have not yet uploaded any data to Terra.bio, please see our video on uploading ONT reads: <https://youtu.be/XKNFMj8ic1k>)

**The top two rows of inputs 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 'demultiplex\_reads' variable, we clicked on the 'Attribute' text box and wrote 'this.reads' to indicate that the 'demultiplex\_reads' we wish to analyze are under the 'reads' 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/Qrbz7uRtwmQ>). In the third row, the 'samplename' variable, we selected on the 'Attribute' text box and wrote 'this.ont\_specimen\_id' to indicate the 'samplename' of each sample we are analyzing can be found in the 'this.ont\_specimen\_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.

SCRIPT    INPUTS    OUTPUTS   

Hide optional inputs Download json | Drag or click to upload json

Task name ↓	Variable	Type	Attribute
titan_ont	demultiplex_reads	File	this.reads <input type="button" value="[-]"/>
titan_ont	primer_bed	File	workspace:midnight_primers <input type="button" value="[-]"/>
titan_ont	samplename	String	this.ont_specimen_id <input type="button" value="[-]"/>

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.

You are now ready to run the Titan\_ONT 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.

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

Terra

WORKSPACES

Workspaces > theigen-validations/Terra\_ONT\_Test > Job History

COVID-19 Data & Tools

Cloud Environment Home

DASHBOARD

DATA

NOTEBOOKS

WORKFLOWS

JOB HISTORY

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

Workflow Statuses

Submitted: 2

Workflow Configuration

theigen-validations/Titan\_ONT

Submitted by

kevin.libuit@theigen.com

May 10, 2021, 8:05 AM

Total Run Cost

N/A

Data Entity

Titan\_ONT\_2021-05-10T14-51-13

Test\_ONT\_sample\_set

Submission ID

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

Call Caching

Enabled

Delete Intermediate Outputs

Disabled

Use Reference Disks

Disabled

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