



Nov 22, 2021

Uploading Data Files to Terra

Francis J Ambrosio¹¹Theiagen Genomics

1

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

Theiagen



Francis Ambrosio

Uploading data to Terra.bio is an essential step in the protocol for analyzing locally stored genomic sequencing data. The Terra.bio uploads page allows users to easily organize their data files using an associated metadata file via a browser-based graphical user interface. This protocol explains the process to prepare the data files and the associated metadata file for upload, and provides the link to the Terra.bio uploads page.

DOI

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

Francis J Ambrosio 2021. Uploading Data Files to Terra . **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.byxjpxkn>



protocol ,

Oct 11, 2021

Nov 22, 2021

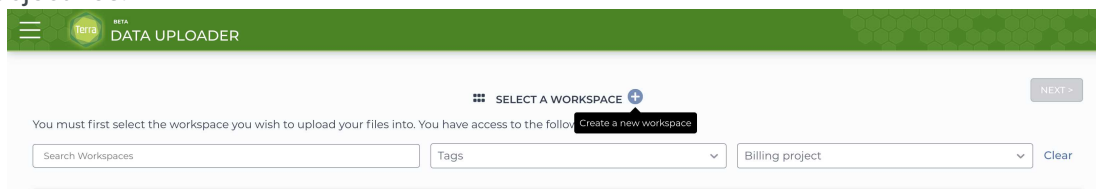
53963

Upload Data Files

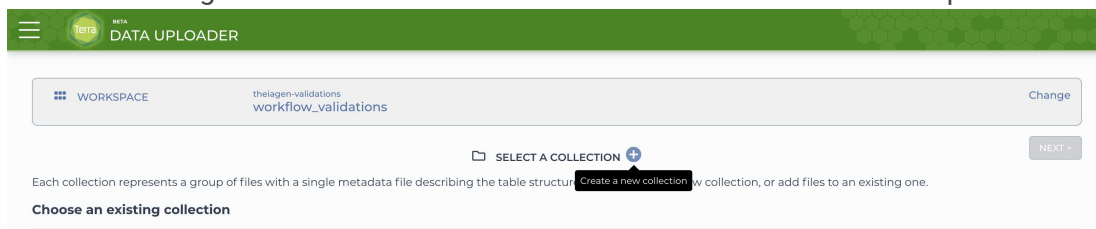
- 1 Navigate to the Terra.bio [uploads page](#).



- 2 Select an existing workspace or create the workspace where the data will be uploaded. If there are no workspaces created under the billing project associated with this account please reach out to support@theiagen.com and we will facilitate the creation of a workspace pre-loaded with workflows to perform the analyses needed to meet any variety of research objectives.



- 3 Select an existing collection or create the collection where the data will be uploaded.



- 4 Drag and drop files into the indicated region of the page (or use the file browser by selecting the plus button in the bottom right hand of the window).

Then click 'NEXT'

Upload Metadata File

- Using the sample names and file names of the files that have been uploaded to Terra, create a spreadsheet of this "metadata" to upload to Terra as well.

Cell A1 defines the root entity. This will be the name of the table where the data is stored in Terra. The root entity is defined using the "**entity:**" notation and must end with "**_id**".

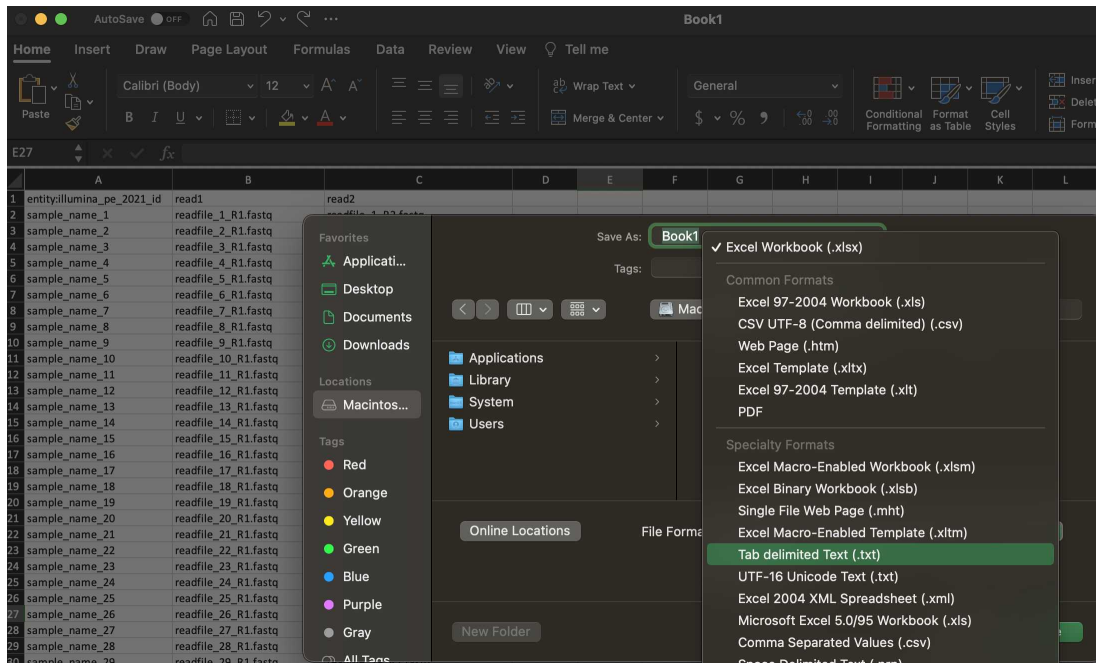
A1				entity:illumina_pe_2021_id
	A	B	C	
1	entity:illumina_pe_2021_id	read1	read2	
2	sample_name_1	readfile_1_R1.fastq	readfile_1_R2.fastq	
3	sample_name_2	readfile_2_R1.fastq	readfile_2_R2.fastq	
4	sample_name_3	readfile_3_R1.fastq	readfile_3_R2.fastq	
5	sample_name_4	readfile_4_R1.fastq	readfile_4_R2.fastq	
6	sample_name_5	readfile_5_R1.fastq	readfile_5_R2.fastq	
7	sample_name_6	readfile_6_R1.fastq	readfile_6_R2.fastq	
8	sample_name_7	readfile_7_R1.fastq	readfile_7_R2.fastq	
9	sample_name_8	readfile_8_R1.fastq	readfile_8_R2.fastq	
10	sample_name_9	readfile_9_R1.fastq	readfile_9_R2.fastq	
11	sample_name_10	readfile_10_R1.fastq	readfile_10_R2.fastq	

The file names in the spreadsheet should match the file names of the files that were uploaded in the previous section.

*NOTE: It is best practices to include the run_id of each sample in an addition column:


	A	B	C	D
1	entity:illumina_pe_2021_id	read1	read2	run_id
2	sample_name_1	readfile_1_R1	readfile_1_R2	miseq_00001
3	sample_name_2	readfile_2_R1	readfile_2_R2	miseq_00001
4	sample_name_3	readfile_3_R1	readfile_3_R2	miseq_00001
5	sample_name_4	readfile_4_R1	readfile_4_R2	miseq_00001
6	sample_name_5	readfile_5_R1	readfile_5_R2	miseq_00001
7	sample_name_6	readfile_6_R1	readfile_6_R2	miseq_00001
8	sample_name_7	readfile_7_R1	readfile_7_R2	miseq_00001
9	sample_name_8	readfile_8_R1	readfile_8_R2	miseq_00001
10	sample_name_9	readfile_9_R1	readfile_9_R2	miseq_00001
11	sample_name_10	readfile_10_R1	readfile_10_R2	miseq_00001


6 Save the file as a tab delimited text file:



7 Drag and drop the metadata file into the indicated region of the page (or use the file browser by selecting the plus button in the bottom right hand of the window).


8 Review the entries in the datatable and click 'CREATE TABLE' to complete the upload process


DATA UPLOADER


WORKSPACE


the1agen-validations
workflow_validations

Change


COLLECTION


titan_fasta

Change


DATA FILES

Includes **31** files


Change



METADATA TABLES


Created table **11umina_pe_2021**, added or modified **30** rows

Change

DONE!

 View the 11umina_pe_2021 table in the workspace

 Create a new table in the titan_fasta collection

 Start over with another workspace or collection

If video is your preferred medium of support check out our youtube video on uploading paired end sequencing data to Terra.bio:

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited