



0



Yuan Li¹, Jill V Hagey¹, Michael Weigand¹, Technical Outreach and Assistance for States Team¹

¹Centers for Disease Control and Prevention

Terra Troubleshooting



protocol

TOAST_public
Tech. support email: toast@cdc.gov

Technical Outreach and Assistance for States Team Centers for Disease Control and Prevention

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only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind.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.

This protocol contains step-by-step instructions and tips for troubleshooting issues you may encounter while using Terra. If your problem cannot be addressed with this protocol, please email **TOAST@cdc.gov** for assistance and to add your question to help others in the future.

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

Yuan Li, Jill V Hagey, Michael Weigand, Technical Outreach and Assistance for States Team 2021. Terra Troubleshooting. **protocols.io**

https://protocols.io/view/terra-troubleshooting-bvtpn6mn

Terra, Illumina, Sequencing, MiSeq, iSeq, MiniSeq, NextSeq, NovaSeq, Paired-end, Next Generation Sequencing, NGS, SARS-CoV-2, COVID, Pangoin, Short-Read, Coronavirus, Genomics, Genetics, Virology, Molecular Biology, Troubleshooting

_____ protocol,

Jun 14, 2021

Jan 21, 2022

Jan 20, 2022 Technical Outreach and Assistance for States Team Centers for Disease Control and Prevention

50767

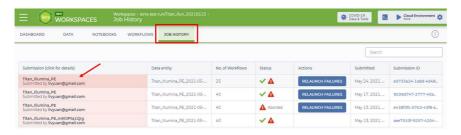
:



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.

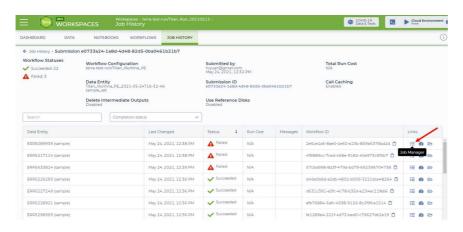
The workflow status is "Failed" for a sample, why?

1 To start troubleshooting the workflow (Titan in this case) for a failed sample, click on the 'JOB HISTORY' panel in the workspace page. It should bring you to the following page:



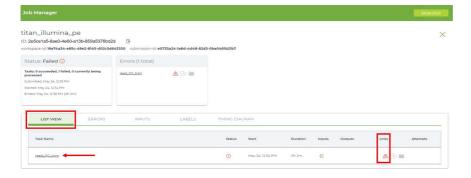
Job history Page

Click on the job submission that contains the failed sample under the 'Submission' tile, and it will take you to a new page that shows a table. In the table find the row that contains the failed sample (e.g. ERR5089939 as shown below), click on the "Job Manager" tile in the last column of the row, and it will take you to a new page.



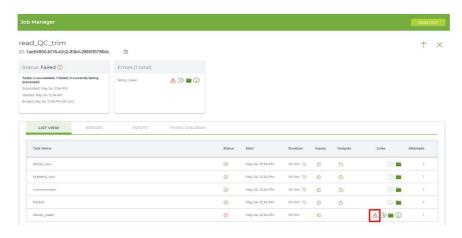
Workflow status of job

You might be asked to authentic your account at this step. Just click on your account and wait for the page to load. On this new page, click on the "LIST VIEW" tab and it will show the task(s) that failed (e.g. "read_QC_trim" as shown below), which is indicate by a red exclamation mark in a triangle under the "Links" tab.



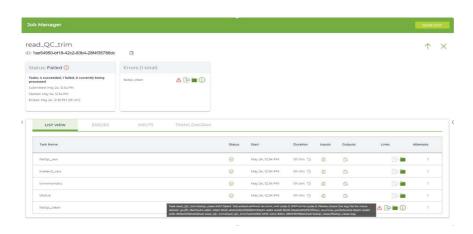
List View tab of the Job Manager page

Click on the task that failed (e.g. "read_QC_trim" as shown below) and it will show the step(s) within the task that failed (e.g. fastqc_clean as shown below), which is indicated by a red exclamation mark in a triangle under the "Links" tab.



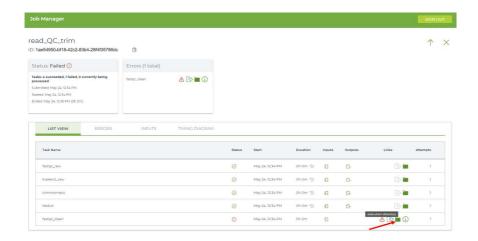
List of tasks that are part of the read_qc_trim job.

Move the mouse over the warning sign (red exclamation mark in a triangle) and you will see the error message. In this case the error message reads "Task read_QC_trim.fastqc_clean:NA:1 failed. Job exited without an error, exit code 0. PAPI error code 9. Please check the log file for more details: gs://fc-16e74a34-e85c-48e2-8145-d02c5d643350/e0733a24-1a8d-4d48-82d5-0ba0461b21b7/titan_illumina_pe/2e5ce1a5-8ae0-4e60-a13b-859a5376bd2d/call-read_QC_trim/read_QC_trim/1ae54950-bf18-42c2-83b4-28f4f35786dc/call-fastqc_clean.log."



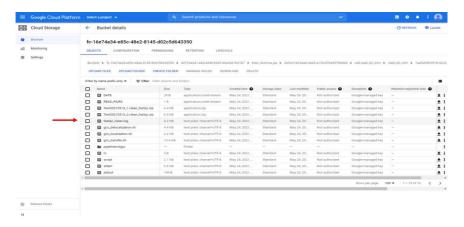
Error message for failed task

To find the log file shown in the error message ("fastqc_clean.log"), click on the file folder icon ("execution directory") to the right of the warning sign, and this will bring you to a new Cloud Storage page.



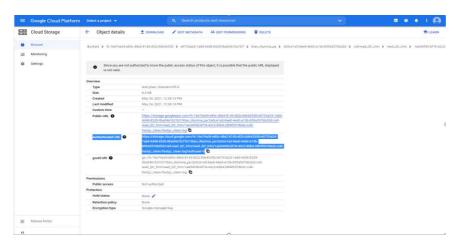
Execution directory button for accessing log files

On the cloud storage page, click on the "fastqc_clean.log" file and it will bring you to a new page.



Google Cloud storage page with all files generated by workflow.

On the "fastqc_clean.log page click on the "Authenticated URL" and it will show you the content of the "fastqc_clean.log" file in a new page.



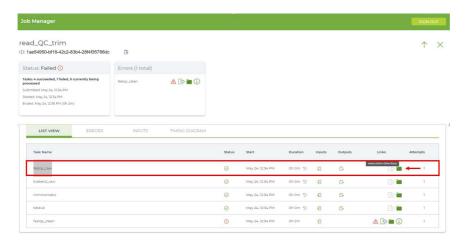
"fastqc_clean.log page

In the highlighted log, it shows "Failed to process file Test20210513_1.clean.fastq.gz", which indicates something might be wrong with the input fastq files.



Error log file.

Now we know were to look for the problem so we will go have a look at these files. To examine the quality of the fastq files, navigate to the "fastqc_raw" execution directory through the Job Manager page as before.



Job Manager page

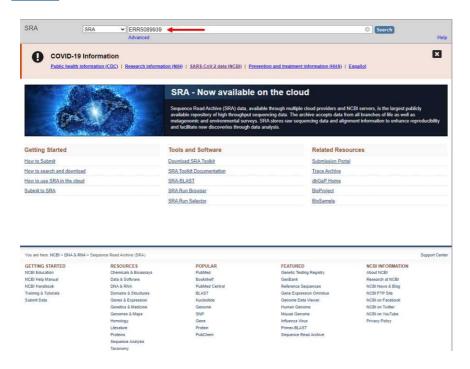
Click on the "fastqc_raw.log" and then the "Authenticated URL" to show the log file content as we did before. In the log file below we can see two things. First, fastQC completed successfully and second it shows "Uneven pairs: R1=686801, R2=684454", suggesting the reads in the fastq files are not properly paired.



fastqc_raw.log output

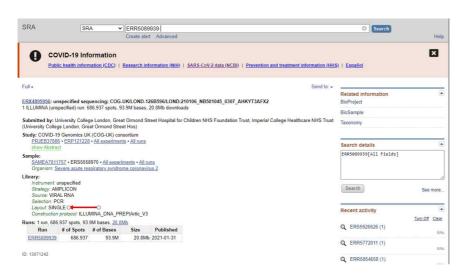
Because the raw reads file (ERR5089939_1.fastq.gz and ERR5089939_2.fastq.gz) are download from NCBI, one can

search the accession number to get additional information. Navigate to the <u>SEQUENCE READ ARCHIVE (SRA)</u> database.



SRA home page search

Search for the SRA number ERR5089939.



SRA record for ERR5089939

Examining the SRA record, we see that **the library has a SINGLE layout** and therefore did not generate pair-end sequencing reads. This is the reason why fastq files were detected as not properly paired and failed the workflow. **The fastq files of this sample should be analyzed by another workflow that processes single end illumine sequencing data.**

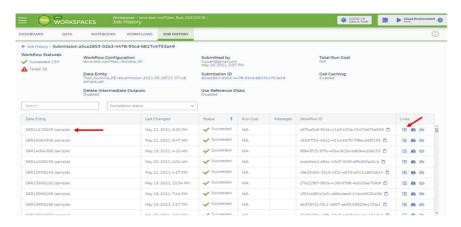
The workflow status is "Succeeded" for a sample, but it takes long time (>10 hours), why?

2 To troubleshoot the unusually long runtime for a succeeded sample, click on the 'JOB HISTORY' panel in the workspace page. It should bring you to the following page:



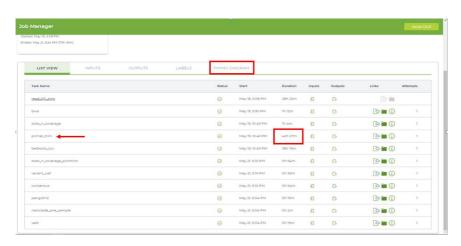
Job History page

Click on the job submission that contains has some sample failures (red triangle) that has been running for a long time under the 'Submission' tile, and it will take you to a new page that shows a table.



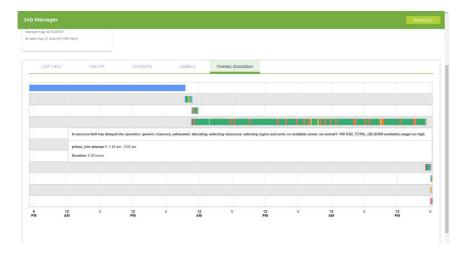
Workflow status of jobs

In the table, find the row that contains the failed sample (e.g. SRR14219269 as shown below), click on the "Job Manager" tile in the last column of the row, and it will take you to a new page. In this page we can see that the primer_trim task had a run duration of 42h27min. **To identify the possible reason, click on the "TIMING DIAGRAM" tab.**



List of tasks and their run times.

Under the "TIMING DIAGRAM" tab, move the mouse over the green/yellow/orange bars and you will see warning message like "The resource limit has delayed the operation: generic::resource_exhausted:allocating: select resources: selecting region and zone: no available zones: us-central1:100 SSD_TOTAL_GB (0/500 available) usage too high".

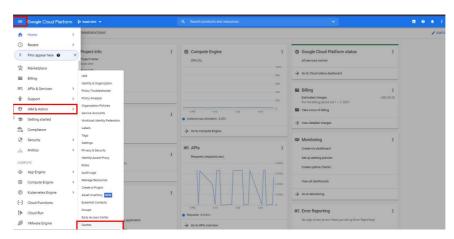


TIMING DIAGRAM tab

In the case here we are using the free google cloud credits and the default quotas, when you start a GCP [Google Could Platform] project are really, really low and not amendable to using Terra (8 CPUs and 500GB of memory). You need to request a quota increase form google and move a paid account to improve performance. You can make a quota increase request by sending a message to Terra support by click the "Contact Us" in the Terra UI, or by emailing support@terra.bio.

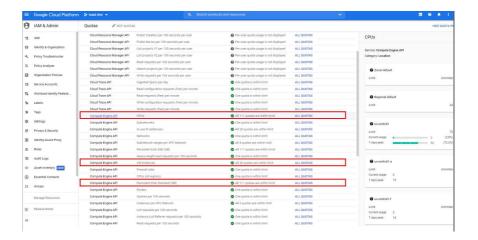
For more information see <u>Terra's support page on the topic</u>.

- 2.1 Please note that, according to the Google Could Platform (https://cloud.google.com/iam/), "free trial accounts for Google Cloud Platform have limited quota during their trial period. In order to increase your quota, please upgrade to a paid account by clicking "Upgrade my account" from the top of any page once logged in to Google Cloud Console."
- 2.2 If you have a paid account, you can check your quota limits if you are having issues with things taking too long. Log into the Google Cloud Platform and click the three dashes at the top left of the page. Then click "IAM & Admin" in the drop down menu. Another drop down menu will popup on the side and you can click on "Quotas".



Navigation to the quotas page.

On the next page, you will find all the quotas. While there is a lot here, only a few will commonly cause a problem.



Terra cares about the following Google Compute Engine API quotas:

- CPUs: how many CPUs you can use at once across all of your tasks
- Preemptible CPUs: the pool of CPUs that would only be used by preemptible instances. You
 can learn more about this quota here and about preemptible instances here.
- Persistent disk standard(GB): how much total disk non-SSD you can have attached at once to your task VMs
- Persistent disk SSD(GB): how much total SSD disk you can have attached at once to your task VMs
- Local SSD(GB): how much SSD is attached directly to the server running the task VMs. You can learn more here. Only applicable if you are using local SSD in your Task.