README for Final T502 RNA-Seq Paper

I pulled directly from your repository to this one just in case I ever wanted to repeat any of this process. All of the code I actually used is here:

**For the first script we ran the test repository this was what I entered into Carbonate:**

mkdir git-repos1

cd git-repos

git clone https://github.com/rtraborn/T502\_Workflow\_tutorial.git

$ ls scripts/

init git

cd T502\_Workflow\_tutorial

cd scripts

example\_batch\_script.sh

nano

^R

Hit enter

Open the file

Adjust it

qsub example\_batch\_script.sh

qstat

shows the jobs

to delete the job

qdel 95991.si

This is the actual script that we ran:

|  |
| --- |
| #!/bin/bash |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| #PBS -N Example\_job\_script\_T502 |
|  |

|  |
| --- |
| #PBS -k o |
|  |

|  |
| --- |
| #PBS -q debug |
|  |

|  |
| --- |
| #PBS -l nodes=1:ppn=16,vmem=16gb |
|  |

|  |
| --- |
| #PBS -l walltime=1:00:00 |
|  |

|  |
| --- |
| #PBS -m abe |
|  |

|  |
| --- |
| #PBS -M rtraborn@indiana.edu |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| module load sra-toolkit |
|  |

|  |
| --- |
| module load fastqc |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| WD=<path to desired working directory> |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| cd $WD |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| echo "Starting job" |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| echo "Downloading the files" |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| fastq-dump SRR2078285 |
|  |

|  |
| --- |
| fastq-dump SRR2078286 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| echo "Running fastqc on the files" |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| fastqc SRR2078285.fastq |
|  |

fastqc SRR2078286.fastq

**This created the actual fastqc files:**

<https://github.com/jshinkl/T502_RNAseq.git>

mkdir git-repos2

cd git-repos2

git clone <https://github.com/jshinkl/T502_RNAseq.git>

$ ls

git init

cd T502\_RNAseq

cd scripts

fastqc\_batch.sh

nano

^R

Hit enter

Open the file

Adjust it

qsub fastqc\_batch.sh

qstat

shows the jobs

to delete the job

qdel 95991.si

job today is 108440.s1

qstat –u jshinkl

this shows all of your jobs

put it in /N/u/jshinkl/Carbonate/git-repos2

then make a symbolic link………

change it to your directory

Then from the readme

|  |
| --- |
| #### Sources of P. pacificus genome files #### |
|  |

|  |
| --- |
| # To download these files, please type 'source 0README' from this directory |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## The Hybrid2 genome assembly |
|  |

|  |
| --- |
| mkdir fasta |
|  |

|  |
| --- |
| cd fasta |
|  |

|  |
| --- |
| wget http://pristionchus.org/download/pacificus\_Hybrid2.gz |
|  |

|  |
| --- |
| gzip -d pacificus\_Hybrid2.gz |
|  |

|  |
| --- |
| ln -s pacificus\_Hybrid2 pacificus\_Hybrid2.fa |
|  |

|  |
| --- |
| cd .. |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## Hybrid2 gene annotation files |
|  |

|  |
| --- |
| mkdir annotation |
|  |

|  |
| --- |
| cd annotation |
|  |

|  |
| --- |
| wget http://pristionchus.org/download/pacificus\_Hybrid2\_annotations.tgz |
|  |

|  |
| --- |
| tar xvf pacificus\_Hybrid2\_annotations.tgz |
|  |

cd ..

The file we actual ran:

|  |
| --- |
| #!/bin/bash |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| #PBS -N PP\_RNAseq\_fastqc\_batch\_T502 |
|  |

|  |
| --- |
| #PBS -k o |
|  |

|  |
| --- |
| #PBS -l nodes=1:ppn=16,vmem=32gb |
|  |

|  |
| --- |
| #PBS -l walltime=1:00:00 |
|  |

|  |
| --- |
| #PBS -m abe |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| module load fastqc |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| fileDir=/N/dc2/scratch/rtraborn/T502\_fastqs/PP\_RNAseq |
|  |

|  |
| --- |
| ####### Before running the script, please enter path to desired output directory, below #### |
|  |

|  |
| --- |
| fqDir=<provide path to desired output directory> |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| cd $fqDir |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-1\_S5\_R1\_001.fastq.gz NHR40-1.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-1\_S5\_R2\_001.fastq.gz NHR40-1.R2.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-2\_S6\_R1\_001.fastq.gz NHR40-2.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-2\_S6\_R2\_001.fastq.gz NHR40-2.R2.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-3\_S7\_R1\_001.fastq.gz NHR40-3.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-3\_S7\_R2\_001.fastq.gz NHR40-3.R2.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-4\_S8\_R1\_001.fastq.gz NHR40-4.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-NHR40-4\_S8\_R2\_001.fastq.gz NHR40-4.R2.fastq.gz |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-1\_S1\_R1\_001.fastq.gz Seud1-1.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-1\_S1\_R2\_001.fastq.gz Seud1-1.R2.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-2\_S2\_R1\_001.fastq.gz Seud1-2.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-2\_S2\_R2\_001.fastq.gz Sedu1-2.R2.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-3\_S3\_R1\_001.fastq.gz Seud1-3.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-3\_S3\_R2\_001.fastq.gz Seud1-3.R2.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-4\_S4\_R1\_001.fastq.gz Seud1-4.R1.fastq.gz |
|  |

|  |
| --- |
| ln -s ${fileDir}/GSF1659-Seud1-4\_S4\_R2\_001.fastq.gz Seud1-4.R2.fastq.gz |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| echo "Starting job" |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| echo "Running fastqc on the RNA-seq files" |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| for fq in \*.fastq.gz; do |
|  |

|  |
| --- |
| echo "Starting fastqc for $fq" |
|  |

|  |
| --- |
| fastqc $fq |
|  |

|  |
| --- |
| done |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| echo "Fastqc job is complete" |
|  |

|  |
| --- |
|  |
|  |

exit

**This is exactly what I put in Carbonate for the alignment:**

mkdir git-repos4

cd git-repos4

git clone https://github.com/rtraborn/T502\_RNAseq.git

ls

git init

cd into

cd T502\_RNAseq

source 0README

ls

cd into scripts

cd T502\_RNAseq

cd scripts

and then

nano align\_STAR.sh (just type nano and go to this file)

make this change in your file:

WD=/N/u/jshinkl/Carbonate/T502\_RNAseq/

This is exactly what I changed:

WD=/N/u/jshinkl/Carbonate/git-repos4/T502\_RNAseq/

#assuming we're in scripts/

qsub align\_STAR.sh

job number is 110955.s1

Log back into Carbonate periodically to check on the status of your job as follows:

qstat -u yourID

in the alignments folder on carbonate in the T502\_RNAseq folder there are now 8 BAM files

Once your job is complete, you should have 8 BAM files in the alignments/ directory. You can look at them using a module called samtools.

cd alignments

module load samtools

samtools view NHR40-2.R1.bam | less

## that should open a window in your terminal that allows you to explore the alignments using your cursor.

## exit by typing the letter 'q' at any time.

We can use SAMtools

To check these in carbonate

ln –s path and your file

ln –s /N/u/jshinkl/Carbonate/git-repos4/T502\_RNAseq/fastqs/Seud1-1.R1.fastq.gz.STAR.Aligned.sortedByCoord.out.bam. test1.bam

ls test1.bam

samtools view test1.bam

**To create the plots on my own in R studio I followed exactly what was in the** [**de\_analysis\_Prist\_hybrid2.R**](https://github.com/jshinkl/Final_T502_RNAseq/blob/master/scripts/de_analysis_Prist_hybrid2.R) **and** [**de\_launch.sh**](https://github.com/jshinkl/Final_T502_RNAseq/blob/master/scripts/de_launch.sh) **in this repo.**