Getting started on your project

Bonnie Hurwitz

Abstract

Today, we are going to get you set up for your class project. You will need a class directory for putting all of your files. Also, you will need to download the data for your project (see the list of SRR numbers in D2L).

Citation: Bonnie Hurwitz Getting started on your project. **protocols.io**
dx.doi.org/10.17504/protocols.io.jqrcmv6
Published: 05 Sep 2017

Protocol

Step 1.

Login to the HPC and move over to the ocelote cluster.

cmd COMMAND
```
ssh hpc
ocelote
```

Step 2.

Create a class directory. Choose a user name for storing your project files. All files for the class project will be written to this directory.

cmd COMMAND
```
mkdir /rsgrps/bh_class/username
```

NOTES
Bonnie Hurwitz 04 Nov 2016

change "username" to YOUR user name.

Change directories to get into your directory

Step 3.

Go into your user directory. Change "username" to YOUR user name.

cmd COMMAND
```
cd /rsgrps/bh_class/username
```

Create a directory for fastq files from the SRA

Step 4.

Create directories for downloading fastq project files. We will download the paired reads from the SRA.

cmd COMMAND
```
mkdir fastq
cd fastq
```
Create a file called "list"

Step 5.

Create a file called "list" with all of your SRR file names.

cmd COMMAND
```
nano list
```
Create a script for downloading the SRR files

Step 6.

Create a script for downloading the SRR files

cmd COMMAND
```
touch get-fastq.sh
chmod 755 get-fastq.sh
```
Add the following commands to the get-fastq.sh script

Step 7.

using nano edit the get-fastq.sh script by adding the commands. This script runs on a node on the HPC (we will execute in a step below). The script uses the sra-tookit to download files from Genbank (a public sequence repository). It downloads the files as paired ends if applicable.

cmd COMMAND
```
#!/bin/bash

#PBS -W group_list=bh_class
#PBS -q standard
#PBS -l select=9:ncpus=28:mem=27gb
#PBS -l pvmem=235gb
#PBS -l walltime=48:00:00
#PBS -l cput=48:00:00
#PBS -M username@email.arizona.edu
#PBS -m bea

module load sratools

echo “my job_id is: ${PBS_JOBID}”

FASTQ_DIR="/rsgrps/bh_class/username/fastq"
export $FASTQ_DIR
cd $FASTQ_DIR

for file in `cat list`; do
   fastq-dump --outdir $FASTQ_DIR --gzip --skip-technical --readids --dumpbase --split-
files --clip $file;
done
```
Run the script on the cluster to get all of the fastq paired-end files

Step 8.

cmd COMMAND
```
mkdir std-err std-out
qsub -o std-out -e std-err get-fastq.sh
```

EXPECTED RESULTS

you will see where the job runs:

```
$ qsub -o std-out -e std-err get-fastq.sh
815336.head1.cm.cluster
```

check the status of your job

Step 9.

You can check if you job is running using the qstat command.

cmd COMMAND

```
qstat
```

EXPECTED RESULTS

```
$ qstat | egrep hur
815336.head1 get-fastq.sh bhurwitz 0 Q oc_standard
```