Running RepeatMasker in SGE mode is much faster but involves more steps. Depending on what you're doing, it may be preferred.

No need to migrate to a qlogin for this exercise.

We will use the same files as in the original RepeatMasker tutorial for this example.

#---#

SETTING UP FOR THE RUN

#---#

This method was developed by friend of the lab, Robert Hubley, one of the maintainers of RepeatMasker as a favor to us.

1] Create a new folder for your SGE run

'cd /lustre/work/<your username>

'mkdir sge_test'

'cd sge_test'

You'll need a setup file for this work. Borrow this one from my home directory:

'cp /home/daray/generic.sge.daray.sh .'

```
#!/bin/bash
#$ -V
#$ -cwd
#$ -S /bin/bash
#$ -N <your job>.SGE
#$ -o $JOB_NAME.o$JOB_ID
#$ -e $JOB_NAME.e$JOB_ID
#$ -q omni
#$ -pe sm 1
#$ -P quanah

cd <path to this folder>
/lustre/work/daray/software/generateSGEClusterRun_Chewie_S.pl \
        -nolow \
```

```
        -twoBit <yourgenome.2bit> \
        -batch_count <integer> \
        -lib <your library>
```

#-nolow is optional, if you want to skip analyzing simple repeats
#-twoBit, note that the genome must be in twobit format. To convert, use
/lustre/daray/software/faToTwoBit.
#faToTwoBit <input .fa> <output.2bit>
#-batch_count, how many batches do you want to break your job into. High number,
faster… If there are processors available.
#-lib or -species, or nothing. -lib = your custom library, -species = use a species library
in RepeatMasker's files, nothing = use all of repbase
#Also note that there are several flavors of this setup script to run, depending on your
needs. All are in my /lustre/work/daray/software directory.

#change the permissions
chmod u+x qsub.sh doLift.sh

#When this script is done, 'sh qsub.sh'.
#When those are done, qsub doLift.sh'.

Convert your genome, if necessary and change this file as necessary for your data.

Save the file with a new file name.

#--#

RUNNING REPEATMASKER IN SGE MODE

#--#

Pretty simple. Just run the modified 'generic' script as a qsub and you will see a new
directory pop up as well as some new files.

New directory – RMPart

New files – doLift.sh, qsub.sh

qsub.sh is just a list of files to submit to the queue.

doLift.sh is a script to compile all of the individual batch runs into a final set of files similar to those output by a standard RepeatMasker run.

RMPart is the folder where all of the action happens. Each of the X number of batches is deposited here.

Once the setup script is run, you will need to submit all of the batch jobs:

'sh qsub.sh'

Once all of those are done, you will need to submit doLift as a qsub.

'qsub doLift.sh'

Most of the resulting files will be the same as the ones output by regular RepeatMasker runs but there are a couple that are new. They're pretty self-explanatory.