Read Recruitment Using Bowtie2

James Thornton Jr

Abstract

This protocol details how to perform read recruitement to the contigs generated from the previous protocol using Bowtie2.

Protocol

Step 1.

Log into the HPC.

cmd COMMAND
```
$ ssh hpc
$ ocelote
```

PAIRED END CLEAN UP

Step 2.

For paired end only:

Move into your fasta directory. Ensure only 1 pair from each file is in the fasta directory. Keep the pair that is "_1.fasta". Move all "_2.fasta" files into the fastq directory.

cmd COMMAND
```
$ cd /rsgrps/bh_class/username/assembly/fasta
$ mv *_2.fasta ../../fastq
```

❶ NOTES

James Thornton Jr 17 Oct 2017

In the previous protocol we ended up only using 1 pair for assembly to simplify the process. In this protocol we only want to map the one pair that we used during assembly. Make sure the fasta directory only contains the _1.fasta of the pair.

Step 3.

Move into the assembly/megahit-out directory created from the previous protocol.

cmd COMMAND

```
$ cd /rsgrps/bh_class/username/assembly/megahit-out
```
Step 4.

Simpifly the fasta headers of your final.contigs.fa file using fasta_renamer from the Fastx toolkit.

cmd COMMAND
```
$ module load fastx
$ fastx_renamer -n COUNT -i final.contigs.fa -o fixed-contigs.fa
```
Step 5.

Move into your project directory. Then create a 'read_recruit' directory. Move into that directory.

cmd COMMAND
```
$ cd /rsgrps/bh_class/username
$ mkdir read_recruit
$ cd !$
```
Step 6.

Create a bam and bowtie2 index directory.

cmd COMMAND
```
$ mkdir bam bt2_index
```
Step 7.

Move into the contig indexing directory. And create the contig index.

cmd COMMAND
```
$ cd bt2_index
$ module load bowtie2
$ bowtie2-build -f /rsgrps/bh_class/username/assembly/megahit-out/fixed-
contigs.fa contig_index
```
➕ NOTES
James Thornton Jr 17 Oct 2017

This step could take awhile depending on the size of your contig file.

Step 8.

Move into your bam directory found at /rsgrps/bh_class/username/read_recruit/bam

cmd COMMAND
```
$ cd ../bam
```
Step 9.

Make directories for standard error and standard out.

cmd COMMAND
```
$ mkdir std-err std-out
```
Step 10.

Copy the following script named bt2_align.sh. Edit the username variables found in the script.

```
#!/bin/bash

#PBS -W group_list=bh_class
#PBS -q windfall
#PBS -l select=1:ncpus=4:mem=15gb
#PBS -l pvmem=14gb
#PBS -l walltime=24:00:00
#PBS -l cput=24:00:00
#PBS -M netid@email.arizona.edu
#PBS -m bea

echo "my job_id is: ${PBS_JOBID}"

#####change here #######
FASTA_DIR="/rsgrps/bh_class/username/fasta"
BT2_INDEX="/rsgrps/bh_class/username/read_recruit/bt2_index/contig_index"
OUT_DIR="/rsgrps/bh_class/username/read_recruit/bam"
CONTIGS="/rsgrps/bh_class/username/assembly/megahit-out/fixed-contigs.fa"
######################

cd $FASTA_DIR
export FASTA_LIST="$FASTA_DIR/fasta-list"
ls *fasta > $FASTA_LIST
echo "Samples to be processed:" $(cat $FASTA_LIST)

module load bowtie2
module load samtools

while read FASTA; do

    FASTA_N=$(basename $FASTA | cut -d '.' -f 1)

    bowtie2 -x $BT2_INDEX -U $FASTA -f --maxins 800 --fr --very-sensitive-local -p 4 -
S $OUT_DIR/$FASTA_N.sam

    cd $OUT_DIR
    echo "Converting $FASTA_N.sam using reference $CONTIGS"
    samtools view -@ 16 -bT $CONTIGS $FASTA_N.sam > $FASTA_N.temp
    echo "Sorting $FASTA_N"
    samtools sort -@ 16 $FASTA_N.temp > $FASTA_N.bam
    echo "Removing $FASTA_N.temp"
    rm $FASTA_N.temp
    cd $FASTA_DIR

done < $FASTA_LIST
```

⊕ NOTES

James Thornton Jr 17 Oct 2017

Many of you actually had your fasta files located in:

/rsgrps/bh_class/username/assembly/fasta

Which is fine, just make sure the script points to the correct location.

Step 11.

Submit the job.

```
$ sub -e std-err/ -o std-out/ bt2_align.sh
```

Step 12.

Upon job completion navigate to your bam std-err directory.

```
$ cd /rsgrps/bh_class/username/read_recruit/bam/std-err
```

Step 13.

Cat the standard error file to view the alignment rate for each file.

```
$ cat 881767.head1.cm.cluster.ER
```
Your file name will differ