Read Recruitment Using Bowtie2 Version 6

James Thornton Jr

Abstract

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

Citation: James Thornton Jr Read Recruitment Using Bowtie2. **protocols.io**
dx.doi.org/10.17504/protocols.io.kevcte6
Published: 24 Oct 2017

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
```

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.

Append your final.contigs.fa file with your partners final.contigs.fa file.

cmd COMMAND

```
$ cat final.contigs.fa /rsgrps/bh_class/partnerusrname/assembly/megahit-
out/final.contigs.fa > combined-contigs.fa
```

The first argument to cat is YOUR fixed-contigs.fa file while the second is the path to your partners final.contigs.fa

NOTES

James Thornton Jr 20 Oct 2017

IMPORTANT: This step requires that your partner has also done the previous step. May require some coordination...

Step 5.

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

cmd COMMAND

```
$ module load fastx
$ fastx_renamer -n COUNT -I combined-contigs.fa -o final-contigs.fa
```

NOTES

James Thornton Jr 23 Oct 2017

Important: If you get 'This looks like a multi-line FASTA file', try this alternative method to rename sequence headers:

awk '/^>/{print ">" ++i; next}{print}' < contigs_in.fa > contigs_out.fa

Step 6.

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 7.

Create a bam and bowtie2 index directory.

cmd COMMAND

```
$ mkdir bam bt2_index
```

Step 8.

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/final-
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 9.

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

cmd COMMAND

```
$ cd ../bam
```

Step 10.

Make directories for standard error and standard out.

cmd COMMAND

```
$ mkdir std-err std-out
```

Step 11.

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

cmd COMMAND

```
#!/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/final-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.

James Thornton Jr 24 Oct 2017

Important: CONTIGS variable is the contigs that you combined with your partner.

Step 12.

Submit the job.

cmd COMMAND

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

Step 13.

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

cmd COMMAND

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

Step 14.

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

cmd COMMAND

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