MG_HW5: Mapping reads with Bowtie2

James Thornton

Abstract

This protocol provides a procedure for mapping reads to co-assembled contigs using Bowtie2.

Citation: James Thornton MG_HW5: Mapping reads with Bowtie2. **protocols.io** dx.doi.org/10.17504/protocols.io.fyibpue
Published: 04 Oct 2016

Guidelines

Bowtie2 Documentation

Before start

If you haven't already, you should consolidate your fasta files into their own directory (seperate from the fastq files). From /rsgrps/bh_class/username :

$ mkdir fasta

Move your .fasta files from the fastq directory into the newly created fasta directory.

Protocol

Step 1.

Login to the HPC and move into Cluster(ICE).

cmd COMMAND

```
$ ssh hpc
$ ice
```

NOTES

James Thornton Jr 03 Oct 2016

Option 3 if you have menu enabled.

Step 2.

Move into your class directory.

cmd COMMAND
```
$ cd /rsgrps/bh_class/username
```
Use YOUR username

Step 3.

Make two directories, one named bt2_index and the other named bam.

cmd COMMAND
```
$ mkdir bt2_index
$ mkdir bam
```
Step 4.

Move into your assembly directory containing your contigs.

cmd COMMAND
```
$ cd assembly/megahit-out
```
NOTES
James Thornton Jr 03 Oct 2016

Note: Your contigs may be in assembly/ depending on which version of the assembly protocol you did.

Step 5.

The fasta headers need to be simplified to make downstream analysis easier. The Fastx Toolkit has a script that can do this quickly:

cmd COMMAND
```
$ module load fastx/0.0.14
$ fastx_renamer -n COUNT -i contigs.fa -o fixed-contigs.fa
```
Now the headers will be named 1 - total number of contigs. >1 >2 >3 .... >3700

NOTES
James Thornton Jr 03 Oct 2016

Megahit formats the fasta headers to include sequence information, such as length, seperated by spaces. The spaces are not compatible with some downstream analysis we will be doing so they must be renamed.

Step 6.

Now move to the bt2_index directory.

cmd COMMAND
```
$ cd /rsgrps/bh_class/username/bt2_index
```

Use YOUR username

Step 7.

Load Bowtie2 and create an index from the fixed-contigs.fa file.

cmd COMMAND

```
$ module load bowtie2/2.2.5
$ bowtie2-build -f /rsgrps/bh_class/username/assembly/fixed-contigs.fa contig_index
```

bowtie2-build creates an index which will allow for reads to be mapped against. -f option indicates that the config file is in fast format contig_index is the base name of the index files to be created

EXPECTED RESULTS

The result should be 6 index files:

NOTES

James Thornton Jr 03 Oct 2016

Important: Make sure you link to fixed-contigs.fa (step 5).

ANNOTATIONS

Bonnie Hurwitz 04 Oct 2016

the fixed contig file may also be in:

/rsgrps/bh_class/username/assembly/megahit-out/fixed-contigs.fa

Step 8.

Now move into the bam directory you created earlier and create std-err and std-out directories.

cmd COMMAND

```
$ cd /rsgrps/bh_class/username/bam
$ mkdir std-err
$ mkdir std-out
```

Step 9.

Copy the following into a new script called bt2_align.sh:

cmd COMMAND

```
#!/bin/bash

#PBS -W group_list=bh_class
```

```
#PBS -q windfall
#PBS -l jobtype=cluster_only
#PBS -l select=1:ncpus=4:mem=15gb
#PBS -l pvmem=14gb
#PBS -l place=pack:shared
#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}”

#----------EDIT THESE----------
UNMAPPED_DIR="/rsgrps/bh_class/username/unmapped"
BT2_INDEX="/rsgrps/bh_class/username/bt2_index/contig_index"
OUT_DIR="/rsgrps/bh_class/username/bam"
CONTIGS="/rsgrps/bh_class/username/assembly/fixed-contigs.fa"
#------------------------------

cd "$UNMAPPED_DIR"
module load bowtie2/2.2.5
module load samtools/1.3.1

for file in `cat list`; do
    # no human
    NH_R1=$UNMAPPED_DIR/$file".paired.1.fastq"
    NH_R2=$UNMAPPED_DIR/$file".paired.2.fastq"
    NH_S=$UNMAPPED_DIR/$file".singletons.fastq"

    bowtie2 -x $BT2_INDEX -1 $NH_R1 -2 $NH_R2 -U $NH_S -q --very-sensitive-local -p 4 -
S $OUT_DIR/$file.sam

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

done
```

Replace netid in the email and username with YOUR username in the "EDIT HERE" section. This script will map each of your fasta files against the index that was created from your contigs. The output will be SAM format but that is converted into BAM using samtools.

NOTES

James Thornton Jr 03 Oct 2016

Note: Your contigs may be in assembly/megahit-out or just assembly/ depending on which version of the assembly protocol you did. Make sure the CONTIGS variable points to the right place.

James Thornton Jr 03 Oct 2016

Important: Make sure you link to fixed-contigs.fa (step 5).

ANNOTATIONS

Bonnie Hurwitz 04 Oct 2016

Don't forget to make this file executable "chmod 755 bt2_align.sh"

Bonnie Hurwitz 04 Oct 2016

be sure to make the bt2_align.sh script executable.

"chmod 755 bt2_align.sh"

Step 10.

Submit bt2_align.sh:

cmd COMMAND
```
qsub -e std-err/ -o std-out/ bt2_align.sh
```
This script should be executed from /rsgrps/bh_class/username/bam

ANNOTATIONS
Amy Hudson 10 Oct 2016

does this error mean anything in particular:

827759.service0

when I check status, I receive:

executing qstat_local

Step 11.

Check the status of your job. The status of the job will go from a 'Q' to 'R' when it is running. Once complete the list will be empty. You should receive email notifications once the job begins running and is complete.

cmd COMMAND
```
$ qstat -u jamesthornton
```
Use YOUR netid

Step 12.

Once the job is complete check the contents of your bam directory. You should have a total of 16 files, 8 .sam and 8 .bam files.

Step 13.

Make a directory to store the .sam files in the bam directory. Move those 8 .sam files into that directory.

cmd COMMAND

```
$ mkdir sam
$ mv *.sam sam
```

Step 14.

Now move into your std-err directory.

cmd COMMAND

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

Step 15.

Cat the file found in this directory. This file contains alignment statistics. Report these statistics and the methods you took to obtain them in your google doc.

cmd COMMAND

```
$ cat 65465.service2.ER
```

NOTES

James Thornton Jr 03 Oct 2016

Look at the read counts to know which report belongs to which sample.

ANNOTATIONS

Trace Ayotte 06 Oct 2016

I followed the protocol exactly and when I went to cat the file, this is what I got.

[service0@/rsgrps/bh_class/traceayotte/bam/std-err]$ cat 827149.service0.ER
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)
[samfaipath] build FASTA index...
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)
Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)

Error: reads file does not look like a FASTA file
terminate called after throwing an instance of 'int'
(ERR): bowtie2-align died with signal 6 (ABRT)