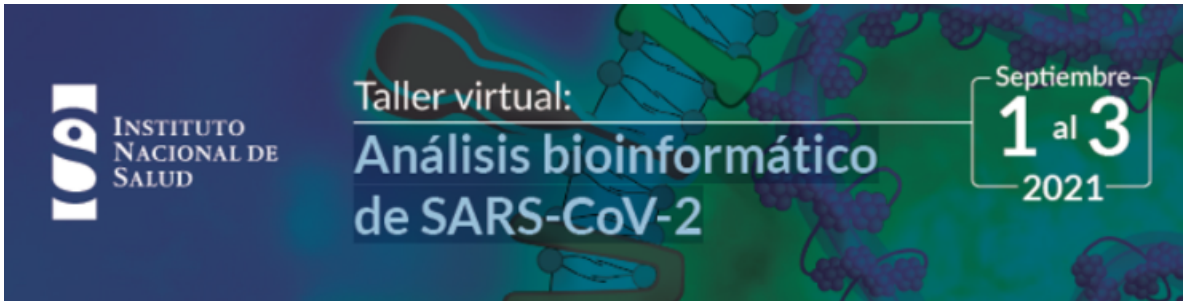
UNIVERSIDAD NACIONAL DE COLOMBIA
Sede Bogotá

Emiliano Barreto Hernández
Profesor Asociado PhD
Centro de Bioinformática
Instituto de Biotecnología
Email: ebarretoh@unal.edu.co

TALLER 3



Practical: Sequence your genome, remap and identify variants

Please note that these are example commands and that you might be forced to change them according to your files and paths.

`commands` are given in “Courier” font

The “\” sign should not be copied, it is here to show that the line continues on the next line...

#login to the main server
```
ssh userxxxx@168.176.61.249
```

#set path for samtools
```
echo export PATH=/data/samtools/bin:$PATH >>.profile
```

#exit session ssh
```
exit
```

#again login to the main server
```
ssh userxxxx@168.176.61.249
```

BWA (mapping)
```
mkdir bwa
cd bwa
```

#link the reference genome and the reads locally
```
ln -s /data/refgenome/refGenome.fasta refGenome.fasta
ln -s /data/filtered_reads filtered_reads
```

#index reference for bwa
```
bwa index refGenome.fasta
```

#map reads onto reference

Emiliano Barreto Hernández
Profesor Asociado PhD
Centro de Bioinformática
Instituto de Biotecnología
Email: ebarretoh@unal.edu.co

```
zcat filtered_reads/barcode09_filtered.fastq.gz| bwa mem -t 16 refGenome.fasta - > barcode09.sam
```

BOWTIE (mapping)

```
cd
mkdir bowtie
cd bowtie
```

#link the reference genome and the reads locally

```
ln -s /data/refgenome/refGenome.fasta refGenome.fasta
ln -s /data/filtered_reads filtered_reads
```

build the index of the reference genome

```
bowtie2-build -f refGenome.fasta refgenome
```

map the reads onto the reference save a sam file

```
zcat filtered_reads/barcode09_filtered.fastq.gz | bowtie2 -q -p 16 -x refgenome -U - -S barcode09.sam 2> bowtie.log
```

BOTH

#index reference for samtools

```
samtools faidx refGenome.fasta
```

#convert sam to bam

```
samtools view -b -@ 4 -t refGenome.fasta.fai barcode09.sam > barcode09_unsorted.bam
```

#sort bam

```
samtools sort -@ 4 -o barcode09_sorted.bam barcode09_unsorted.bam
```

#index bam

```
samtools index barcode09_sorted.bam
```

Extract the consensus

#call variants

```
bcftools mpileup -Ou -f refGenome.fasta barcode09_sorted.bam | bcftools call -mv -Oz -o calls.vcf.gz

bcftools index calls.vcf.gz
```

Emiliano Barreto Hernández
Profesor Asociado PhD
Centro de Bioinformática
Instituto de Biotecnología
Email: ebarretoh@unal.edu.co

#normalize indels
```
bcftools norm -f refGenome.fasta calls.vcf.gz -Ob -o calls.norm.bcf
```

#filter adjacent indels within 5bp
```
bcftools filter --IndelGap 5 calls.norm.bcf -Ob -o calls.norm.flt-indels.bcf
```

apply variants to create consensus sequence
```
cat refGenome.fasta| bcftools consensus calls.vcf.gz > consensus.fa
```

Once the mapping is complete, you can obtain some metrics about the alignment as follows:

BWA doesn't provide metrics of the alignment. So, use the next command for counting the number of unmapped reads from the SAM output.

```
awk '{ if ( $3~/\*/) {print} }' barcode09.sam | wc -l
```

Bowtie provides some metrics during the alignment process. Look at the entire log file (is very short).

```
more bowtie.log
```

Then, answer the following questions for both single / paired alignment:

1. How many reads were mapped onto the reference?
2. How many reads were not mapped onto the reference?

View with the terminal (? for help, q for quit)
```
samtools tview barcode09_sorted.bam refGenome.fasta
```

Is the reference uniformly covered? Do you observe any region of high coverage? Could you explain why?

Visualize with IGV or Tablet
#download IGV or use the java webstart application https://www.broadinstitute.org/software/igv/download