

MAPPING

There are 2 software available in the server for mapping: **Bowtie**, **BWA**.

Samtools are used to convert files and extract information.

Please, use a single mapping software for the practical.

We ask you to map reads in paired-end either with BWA or with Bowtie and generate a consensus assembly (with samtools).

For more information about the parameters please refer to the help message or web manual of each program.

Web manuals:

Bowtie: <http://bowtie-bio.sourceforge.net/manual.shtml>

BWA: <http://bio-bwa.sourceforge.net/bwa.shtml>

SAMtools: <http://samtools.sourceforge.net/samtools.shtml>

You can use raw or filtered data, but filtered is preferred.

BOWTIE (single-end mapping) (30-40 min):

```
cd data
mkdir bowtie
cd bowtie
```

build the index of the reference genome

```
bowtie-build -f ../CP001844.fa refgenome
```

map the reads onto the reference save a sam file

```
bowtie -q --phred64-qual -n 3 --un ref_un -S -p 8 refgenome  
../s_1_1_sequence.txt.filtered,../s_1_2_sequence.txt.filtered 1> refgenome.sam 2> bowtie.log
```

for your info: you can also align in paired-end with bowtie (please don't do it today)

```
bowtie -q --phred64-qual -v 2 -I 300 -X 600 --un ref_un -S -p 8 refgenome -1  
../s_1_1_sequence.txt.filtered -2 ../s_1_2_sequence.txt.filtered 1> refgenome.sam 2> bowtie.log
```

or BWA (paired-end mapping) (30-40 min):

```
cd data  
mkdir bwa  
cd bwa  
#build the index of the reference genome  
bwa index -a bwtsw ../CP001844.fa  
#generate alignments in the suffix array coordinate  
bwa aln -t 4 -I ../CP001844.fa ../s_1_1_sequence.txt.filtered > s_1.sai 2> bwa1.log &  
bwa aln -t 4 -I ../CP001844.fa ../s_1_2_sequence.txt.filtered > s_2.sai 2> bwa2.log &  
# map the reads onto the reference save a sam file  
bwa sampe -a 600 ../CP001844.fa s_1.sai s_2.sai ../s_1_1_sequence.txt.filtered  
../s_1_2_sequence.txt.filtered > refgenome.sam
```

BOTH

convert sam to bam

```
samtools view -S -T ../CP001844.fa -b refgenome.sam > refgenome.unsorted.bam
```

sort the bam file (creates refgenome.bam)

```
samtools sort refgenome.unsorted.bam refgenome
```

index the bam file

```
samtools index refgenome.bam
```

view with the terminal (? for help, q for quit)

```
samtools faidx ../CP001844.fa
```

```
samtools tview refgenome.bam ../CP001844.fa
```

extract the consensus

```
samtools mpileup -uf ../CP001844.fa refgenome.bam | bcftools view -cg - | vcfutils.pl vcf2fq >  
targetgenome.fq
```

convert to fasta

```
seqret targetgenome.fq -filter > targetgenome.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} }' <output.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:

Questions:

1. How many reads were mapped onto the reference?
2. How many reads were not mapped onto the reference?
3. What is the length of the new reconstructed genome? (Use `infoseq targetgenome.fa` and look at the sum)

4. Why is it the same length as the reference genome?

VISUALIZATION

Use **IGV** program for visualizing your mapping by typing in the shell:

```
igv_linux-64.sh
```

Create a new genome with "File->Import Genome...".

Give a name (e.g., refgenome), select the reference in fasta format (CP001844.fa) for the Sequence File, and the BED description file (CP001844.bed) for the Gene File. Click Save.

Load your mapping BAM file (File->Load from File...).

View different regions of the genome, when you zoom deeply the reads are loaded.

Question:

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

2. Compare with known *S.aureus* genomes in their SCCmec region (http://www.sccmec.org/Pages/SCC_ClassificationEN.html). Is it the same type as the reference genome? Could you explain why?

Last modified: Monday, 21 March 2011, 11:29 AM

You are currently using guest access ([Login](#))

NGS_BO