Short reads alignment

Prepare your files

```
cd
mkdir phage
cp /homes/qib/shared/phage_reads.fastq phage/
```

Your reference

Our reference genome is the FASTA file we already used to make some examples: '~/course/db/phage.fa'. If you don't find it, you can grab a copy from '/homes/qib/shared/phage_assembly.fa';

Creating a short read

Extract a subsequence of your choice from "~/course/db/phage.fa" and save it in FASTA format, in a file called '~/phate/seq.fa'. You can create more than one, you can add mismatches, small insertions or deletions. Also, you can reverse complement it.

```
# Try:
revcompl AAAAAAGTGT
```

BWA alignment

Indexing the genome

This is a one-step procedure: when you download a new FASTA file to be used as reference for an alignment, you have to index it first.

```
bwa index ~/course/db/phage.fa
```

Alignment: first test!

Now we can align sequences.

```
bwa mem -t 4 ~/course/db/phage.fa ~/phage/seq.fa > ~/phage/seq.sam
```

Time to inspect your first SAM file!

Alignment: a dataset

```
bwa mem -t 4 ~/course/db/phage.fa ~/phage/phage_reads.fastq >
~/phage/reads.sam
```

Samtools primer

samtools is the swiss-army knife for manipulating SAM files. We will see only the minimal pipeline to convert a SAM file to its binary version (BAM), sorting it by coordinate an finally indexing it.

This is a mock workflow: try to adapt it:

```
# Convert SAM to BAM: two alternatives
samtools view -b -T {reference} {sam_file} > {bam_output}
samtools view -b -S {sam_file} > {bam_output}

# Sort a BAM file
samtools sort -o {sorted_bam} {unsorted_bam}

# Indexing a _sorted_ BAM file
samtools index {sorted_bam}
# see with an 'ls' that a new file has been created
```

Samtools examples

Flags

Those hard to remember flags: list them, “explain” one flag, create a flag

```
samtools flags
samtools flags 96
samtools flags PAIRED,PROPER_PAIR
```

Flag summary

A summary of the aligned reads:

```
samtools flagstat phage/aligned.bam
```

- Sam format hands-on, PDF