Exercise 5

Ewa Szczurek
MIM UW

November 3, 2015

Exercise 1. Read mapping with Bowtie.

- Download Bowtie and extract to a fresh directory. Change to that directory.
- Bowtie comes with a pre-built index of the E. coli genome, and a set of 1,000 35-bp reads simulated from that genome. To align those reads, issue

```
./bowtie e_coli reads/e_coli_1000.fq
```

- The first argument is the basename of the index for the genome to be searched. The second is the name of a FASTQ file containing the reads.
- Output contains lines, one per each read.
- Next, issue

```
./bowtie -t e_coli reads/e_coli_1000.fq e_coli.map
```

- Now, the alignments are written to e_coli.map (the final argument) rather than to the screen. Also, the -t option instructs Bowtie to print timing statistics.
- We can also ask Bowtie to produce a SAM file as output

```
./bowtie -S e_coli reads/e_coli_1000.fq e_coli.sam
```

Exercise 2. Installing a pre-built index of Bowtie

- Download the pre-built S. cerevisiae genome package
- unzip the archive into the Bowtie 'indexes' subdirectory
- To test that the index is properly installed, issue

```
./bowtie -c s_cerevisiae ATTGTAGTTCGAGTAAGTAATGTGGGTTTG
```

- The -c argument instructs Bowtie to obtain read sequences directly from the command line rather than from a file.
- This command should print a single alignment and then exit

Exercise 3. Building a new index. The pre-built E. coli index included with Bowtie is built from the sequence for strain 536, known to cause urinary tract infections. We will create a new index from the sequence of E. coli strain O157:H7, a strain known to cause food poisoning.

- Download the sequence file
- move it to the Bowtie install directory and issue

```
./bowtie-build NC_002127.fna e_coli_O157_H7
```

- The command should print several lines of status messages.
- When the command has completed, note that the current directory contains four new files named e_coli_O157_H7.1.ebwt, e_coli_O157_H7.2.ebwt, e_coli_O157_H7.rev.1.ebwt, and e_coli_O157_H7.rev.2.ebwt. These files constitute the index.
- Move these files to the indexes subdirectory to install it.
- To test that the index is properly installed, issue this command:

```
./bowtie -c e_coli_O157_H7 GCGTGAGCTATGAGAAAGCGCCACGCTTCC
```

- If the index is installed properly, this command should print a single alignment and then exit.

Exercise 4. Using SAMtools

- Download the samtools software
- Extract

```
tar -jxvf samtools-1.2.tar.bz2
cd samtools-1.2/
make
./samtools
```

We will use SAMtools to find SNPs in a set of simulated reads included with Bowtie. The reads cover the first 10,000 bases of the pre-built E. coli genome and contain 10 SNPs throughout. First, we run 'bowtie' to align the reads, being sure to specify the -S option.

- Change to the bowtie directory and run

```
./bowtie -S e_coli reads/e_coli_10000snp.fq ec_snp.sam
```

- Next, we convert the SAM file to BAM in preparation for sorting.

```
PATH_TO/samtools-1.2/samtools view -bS -o ec_snp.bam ec_snp.sam
```

- Next, we sort the BAM file, in preparation for SNP calling:

```
PATH_TO/samtools-1.2/samtools sort ec_snp.bam ec_snp.sorted
```

- We now have a sorted BAM file called ec_snp.sorted.bam. Sorted BAM is a useful format because the alignments are both compressed, which is convenient for long-term storage, and sorted, which is convenient for variant discovery.
- SAMtools uses own indexing of bam files

```
PATH_TO/samtools-1.2/samtools index ec_snp.sorted.bam
```

- And visualizes alignments

```
PATH_TO/samtools-1.2/samtools tview ec_snp.sorted.bam genomes/NC_008253.fna
```

Exercise 5. Using the Galaxy server for mapping Fission yeast reads with Bowtie. To access the Galaxy server from the outside use the link

- To access the Galaxy server from the outside use the link `http://bioputer.mimuw.edu.pl:12345/root`
- Otherwise use `http://centromere:8080`
- login: your email
- password: your last name (first letter is large)
- We will analyse the fission yeast reads from the "groomed" test1.fastq file
- Choose on the left panel NGS: Mapping, and Map with Bowtie for Illumina
- Select a reference genome: fission yeast
- FASTQ file: your "groomed" test1.fastq file

Homework 1. No homework as of today!