Introduction to Bioinformatics using the eBioKit Platform

Commission for Biotechnology & Genetic Engineering, National Center for Research, December 1 - 6, 2014

Mapping with bowtie

Andreas Gisel, IITA – Ibadan, Nigeria

Reads: reads1_red.fq
reads2_red.fq
Ref ref1.fa
ref2.fa

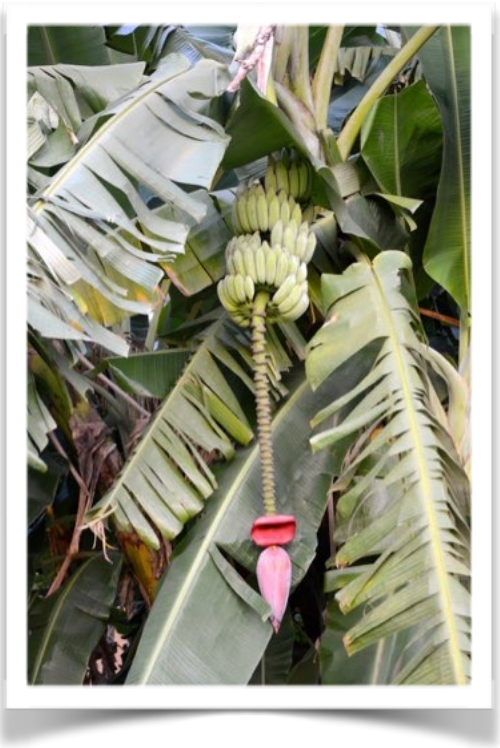
1. in your home create a directory 'results' (mkdir results)
2. copy the data above into the directory 'results (cp path results)
3. enter into the directory results (cd results)
4. check the content of the current directory (ls -l)
5. control the content of the reads file (more reads1.fq; Ctrl C)

Now we can start with the mapping:
index

1. create the index of the reference sequence ref1.fa
2. cmd: bowtie-build -f ref1.fa ref
3. control the index (ls -l); you should have 6 files with the extension 'ebwt'
4. alternative test: ls l grep -c ebwt (6)

single-read mapping
we will test different mapping options

1. we map with the -v option and print always in SAM file format
   a) cmd: bowtie -v 0 -S -p 2 ref -q reads1.fq > mapping_s.sam
   b) cmd: bowtie -v 1 -S -p 2 ref -q reads1.fq > mapping_s.sam
   c) cmd: bowtie -a --best -v 2 -S -p 2 ref -q reads1.fq > mapping_s.sam
2. we map with the –n option and print always in SAM file format



a) cmd: bowtie -n 1 -l 20 -S -p 2 ref -q reads1.fq > mapping_s.sam
b) cmd: bowtie -a --best -n 1 -l 20 -S -p 2 ref -q reads1.fq --un unmapped.fq > mapping_s.sam
3. control whether unmapped is really a fastq file! (head -4 unmapped.fastq)

paired-end mapping

1.cmd: bowtie -n 1 -l 20 -I 100 -X 300 -S -p 2 ref -1 reads1.fq -2 reads2.fq > mapped_p.sam
2.compare the two SAM filed - single reads and paired-end reads (head -10 mapping_s.sam; head -10 mapping_p.sam)
3. Check in the paired-end SAM file the insert size.(more mapped_p.sam; Crtl C)
4. Download mapped_p.sam and ref1.fa file
5. Visualize the data in tablet