Introduction to read mapping

Frank Aylward

Abstract

This is an example of a simple read mapping workflow. It is designed to be performed via the command line on an Ubuntu 16.06 OS.

After completing this tutorial you should:

1) Have a practical understanding of how read mapping analyses are performed in the command line.

2) Understand the basics of how to process SAM and BAM files.

3) Be able to calculate depth of coverage of a contig/scaffold/chromosome in a query read dataset.

Citation: Frank Aylward Introduction to read mapping. **protocols.io**
dx.doi.org/10.17504/protocols.io.pgwdjxe
Published: 17 Apr 2018

Protocol

Ensure the appropriate tools are installed

Step 1.

First ensure that the following tools are installed and are in your PATH:

Bowtie 2 version 2.2.6

samtools Version: 0.1.19-96b5f2294a

bedtools v2.25.0

The exact versions may not be critical, but those used in designing this tutorial are listed for completeness.

First we need to get a reference genome. In this case we will be working with a bacterium called

Marinimicrobial UBA

Step 2.

First we need to get a reference genome to map reads against. We'll download the genome of Marinimicrobia UBA2153 here.

Download using the unix command wget:

wget ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/328/885/GCA_002328885.1_ASM232888v1/GCA_002328885.1_ASM232888v1_genomic.fna.gz

And since the fasta file is compressed we will use the Unix tool gunzip to decompress it:
gunzip GCA_002328885.1_ASM232888v1_genomic.fna.gz

Build bowtie2 reference

Step 3.

To use the read mapping tool bowtie2 we will need to first index the fasta file. This can be done using the command bowtie2-build. It will create several index files with different suffixes and the prefix that we give in the command after the fasta file (in this case UBA2153).

bowtie2-build GCA_002328885.1_ASM232888v1_genomic.fna UBA2153

Get the reads for mapping

Step 4.

Now we need to get the reads that we will use for mapping.

fastq-dump -X 1000000 --gzip --split-3 SRR5322088

Map the reads with bowtie2

Step 5.

bowtie2 -1 SRR5322088_1.fastq -2 SRR5322088_2.fastq -x UBA2153 -S mapping_output.SAM

Now process the SAM file created by bowtie2 with samtools

Step 6.

samtools view -bS -F 4 mapping_output.SAM > mapping_output.bam
samtools sort mapping_output.bam mapping_output.sort
samtools index mapping_output.sort.bam

```
samtools idxstats mapping_output.sort.bam
```