

DAY 1

Download reference files:

```
!wget
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF_000005845.2_ASM584v2/GCF_000005845.2_ASM584v2_genomic.fna.gz
```

reference sequence of the parental (unevolved, not resistant to antibiotics) E. coli strain

```
wget
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF_000005845.2_ASM584v2/GCF_000005845.2_ASM584v2_genomic.gff.gz
```

```
wget https://figshare.com/ndownloader/files/23769692
wget https://figshare.com/ndownloader/files/23769689
```

```
zcat 23769689 | head -20
```

Managing environments:

```
micromamba install -c bioconda trimmomatic fastqc bwa samtools igv varscan snpeff
```

```
zcat reads1.fa.gz | wc -l          \\ 1823504 /4 = 455876 reads
zcat reads2.fa.gz | wc -l          \\ 1823504 /4 = 455876 reads
```

```
fastqc -o . reads1.fa.gz reads2.fa.gz
```

Trimming reads by 20 quality:

<https://datacarpentry.org/wrangling-genomics/03-trimming.html>

```
gunzip reads1.fa.gz > rreads1.fastq
gunzip reads2.fa.gz > rreads2.fastq
trimmomatic PE -threads 4 -phred33 rreads1.fastq rreads2.fastq trimmed1.fastq
untrimmed1.fastq trimmed2.fastq untrimmed2.fastq LEADING:20 TRAILING:20
SLIDINGWINDOW:10:20 MINLEN:20
fastqc -o . trimmed1.fastq trimmed2.fastq
```

Trimming reads by 30 quality:

```
mkdir trimmed_30
trimmomatic PE -threads 4 -phred33 rreads1.fastq rreads2.fastq \
  trimmed_30/R1.trimmed.fastq trimmed_30/R1_un.trimmed.fastq \
  trimmed_30/R2.trimmed.fastq trimmed_30/R2_un.trimmed.fastq \
  LEADING:30 TRAILING:30 SLIDINGWINDOW:10:30 MINLEN:20
```

```
!mkdir fastqc_trimmed_30
!fastqc trimmed_30/* -o fastqc_trimmed_30
```

DAY 2

Aligning sequences to reference

Index the reference file

```
bwa index ref/GCF*fna.gz
```

Align your reads

```
bwa mem ref/GCF_000005845.2_ASM584v2_genomic.fna.gz trimmed/R1.trimmed.fastq  
trimmed/R2.trimmed.fastq > align/GCF_000005845.2_ASM584v2_genomic.sam
```

Compress SAM file

```
samtools view -S -b align/*.sam > align/GCF_000005845.2_ASM584v2_genomic.bam
```

```
samtools flagstat align/*.bam # 99.87 % of reads were mapped
```

Sort and index BAM file

```
samtools sort align/*.bam -o align/GCF_000005845.2_ASM584v2_genomic.sorted.bam  
samtools index align/*.sorted.bam
```

Variant calling

```
samtools mpileup -f ref/GCF_000005845.2_ASM584v2_genomic.fna  
align/GCF_000005845.2_ASM584v2_genomic.sorted.bam >  
var_call/GCF_000005845.2_ASM584v2_genomic.mpileup
```

```
varscan mpileup2snp var_call/*.mpileup --min-var-freq 0.4 --variants --output-vcf 1 >  
var_call/vs_results.vcf
```

Automatic SNP annotation

```
wget  
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF\_000005845.  
2\_ASM584v2/GCF\_000005845.2\_ASM584v2\_genomic.gbff.gz  
echo "k12.genome : ecoli_K12" > snpEff.config  
mkdir -p db/k12  
gunzip *gbff.gz  
ls  
cp *gbff db/k12/genes.gbff  
snpEff build -genbank -v k12
```

GENES FROM IGV

1. *ftsI*

Loc 93,043

Ref C

Alt G

544 aminoacid from Ala(hydrophobic) to Gly

Inhibited by beta-lactam antibiotics such as penicillin, moenomycin, macarbomycin, furazlocillin and piperacillin. Antibiotics inhibit the activity by binding to the catalytic serine
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC305773/>

FtsI (also called PBP3) of Escherichia coli is a transpeptidase required for synthesis of peptidoglycan in the division septum and is one of several proteins that localize to the septal ring.

Ampicillin acts as an irreversible inhibitor of the enzyme transpeptidase, which is needed by bacteria to make the cell wall.

<https://www.rcsb.org/structure/7ONN>

<https://pubmed.ncbi.nlm.nih.gov/34356681/>

2.acrB (for discussion)

569 amino acid Gln (polar uncharged) to Leu(hydrophobic)

Acroflavine resistance family

<https://pubmed.ncbi.nlm.nih.gov/15155734/>

<https://pubmed.ncbi.nlm.nih.gov/18781920/>

The E. coli AcrB protein is a transporter that is energized by proton-motive force and that shows the widest substrate specificity among all known multidrug pumps, ranging from most of the currently used antibiotics, disinfectants, dyes, and detergents to simple solvents

<https://www.rcsb.org/structure/2HQC>

Unimportant:

rsgA - s30 mitochondrial

EnvZ functions as a membrane-associated protein kinase that phosphorylates OmpR in response to environmental signals

glnH - glutamine transport

DAY 3

TASKS

- Figure out what mutations do
- Classify mechanism of resistance
- Make recommendations for alternative treatment

Github page - https://github.com/Aryunaa/bioPrac_p1