Methodology

1. Downloaded reference genome in FASTA format - assembled genome of *Escherichia coli.*

   Downloading the fasta file directly from the ENA database or use FTP on Linux/Ubuntu to download the fasta file on the desired directory. I preferred direct downloading using the ENA browser as the file was of small size.

   Link: https://www.ebi.ac.uk/ena/browser/view/GCA_000259695.1
   E coli strain- LCT-EC106 used here as reference produces alpha-hemolysin and adhesion that could be related to virulence.

2. Downloaded the reads (forward and reverse) - for using them against the reference genome. They should be in FASTQ format.
   I downloaded the Illumina WGS of *E coli* (NextSeq2000 model) from the ENA database. The reads provided have FTP link that can be used to download the reads.

   Using the below command we can download the reads (forward and reverse):

```
wget -nc
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR228/040/SRR22895640
/SRR22895640_1.fastq.gz
wget -nc
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR228/040/SRR22895640
/SRR22895640_2.fastq.gz
```

   Link: https://www.ebi.ac.uk/ena/browser/view/SRR22895640

3. Ran indexing of reference genome using **bowtie2** tool.
   This command is to create indexed files of the reference sequence. A total of 6 index files is created in the directory

```
bowtie2-build GCA_041870105.1.fasta ref_index
```

4. Used **bowtie2** to assemble the reads.

5. Executed **samtools** package to create the bam file and sort the file as well.

```
bowtie2 -x ref_index -1 SRR22895640_1.fastq -2
SRR22895640_2.fastq | samtools sort -o sorted_01.bam
```

N.B: Both the commands were run using pipe (|) between bowtie2 and samtools command syntax

6. The next step necessitates to index the sorted bam file which is a required step in variant calling process.

```
samtools index sorted_01.bam
```

7. Then ran **bcftools** for creating vcf file with all variants listed with reference to the reference genome. The command expects the indexed & sorted bam file to identify the variants and list them in a file that can be viewed using the terminal window.

```
bcftools mpileup -f GCA_041870105.1.fasta sorted_01.bam |
bcftools call -mv -o variants_01.vcf
```