**CODING TEAM WORKFLOW BY FAIZAN**

| **STEP** | **INPUT** | **TOOL** | **USES** | **OUTPUT** |
| --- | --- | --- | --- | --- |
| **1** | Raw reads (fastq) | Fastqc | Check quality of reads | Summary html report |
| **2** | Raw reads(fastq) | Fastp | Improve raw reads quality | Trim reads(fastq) |
| **3** | Trim reads(fastq) | Spades | trimmed reads into longer contiguous sequences (contigs) | Assembled sequences(contigs.fasta) |
| **4** | Assembled sequences(contigs.fasta) | SeqSero2 | predict the serotype | predicted serotype(result.txt) |
| **5** | Assembled sequences(contigs.fasta) | PlasmidFinder | uses a database of known plasmid sequences to identify matches | predicted plamidtyping  (result.txt) |
| **6** | Assembled sequences(contigs.fasta) | AMRFinderPlus | compares contigs against a database of known resistance genes | predicted amrgenes  (result.txt) |
| **7** | Trim reads(fastq) + reference genome (fasta) | Snippy | variant calling and core genome alignment | variant calls (vcf) + consensus sequence based on the reference and the called variants(fasta) + BAM file of reads aligned to the reference genome (BAM) + Provides details on the process and results(Summary Stats) |
| **8** | BAM file of reads aligned to the reference genome (BAM) | Snippy-core | combine the outputs of multiple Snippy runs to create a core genome alignment. | A single multiple sequence alignment (MSA) FASTA file of the core genome across all included samples. (fasta) |
| **9** | FASTA file of the core genome across all included samples. (fasta) | RAxML-NG | building phylogenetic trees using maximum likelihood estimation. | The tree is often in Newick format.(.tree) + logfiles |
| **10** | Combine file contain identifier ,serotype,plasmidtype,amrtypes(csv) + treedata(.tree) | ggplot(R) | Visulization of plot | Final figure |

TASK 1

1. download raw reads

2. quality control to rawreads using fastqc

3. trimming using fastp

4. quality control to trimmed reads

5. genome assembly using spdes

6. insilico serotyping using seqsero2 à TSV

7. plasmidtyping using abricate plasmidfinder à CSV

8. AMR gene detection using amrfinderplus à CSV

**Faizan**

INTIAL 20 RAW READS

**Ayo**

MID 15 RAW READS

**Mariam**

LAST 15 RAW READS

TASK2

50 RAW READS DOWNLOAD > FASTQC > FASTP > SNIPPY > SNIPPYCORE >RAXML-NG > TREEFILE

TASK3

Plotting the phylogenetic tree [plot function], heatmap [Heatmap function in ComplexHeatmap package] and the summary data “isolation country, isolation date, and sample source” [rowAnnotations functions in ComplexHeatmap Function] using Rstudio.