**SUPPLEMENTARY FILE**

Here, a detailed description of the WGS data pre-processing, software, parameter settings, and the sequence analysis of the *E. coli* ACM5 genome is given.

NOTE: As a recommendation and for simplicity, it is preferable to install and run the programs on a Linux-based or Mac operating system (OS). However, except for the genome assembler tool, they could be installed in Windows OS as well. Additionally, it is recommended to revise the manuals of the tools for further information.

1. **Quality assessment of raw sequencing data with FastQC1 (v.0.11.9)**
   1. Download and install the FastQC binary files according to the OS available (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
   2. Run the program either as an interactive graphical application by clicking on the **FastQC Application** icon or non-interactively (command line) by typing the following command in the directory containing the FASTQ files:

***$*** *fastqc sample\_xx\_R1.fastq.gz sample\_xx\_R2.fastq.gz*

Where:

sample\_xx\_R1.fastq.gz = raw reads (forward)

sample\_xx\_R2.fastq.gz = raw reads (reverse)

1. **Raw read pre-processing with Cutadapt2 (v.2.4)**
   1. Depending on the OS, follow the instructions for the Cutadapt tool installation (https://cutadapt.readthedocs.io/en/stable/index.html). Of note, the current version of Cutadapt program is 4.1.
   2. Once installed, run the following command in the directory containing the FASTQ files to remove adapter sequences, low-quality bases (<Q30), and short reads (<50 bp):

***$*** *cutadapt -q 30,30 -a adapter1\_seq -A adapter2\_seq -l 148 --trim-n -m 50 -o sample\_xx\_clean\_R1.fastq.gz -p sample\_xx\_clean\_R2.fastq.gz sample\_xx\_R1.fastq.gz sample\_xx\_R2.fastq.gz*

Where:

-q = 5’-end, 3’-end low-quality trimming

-a/-A = forward/reverse 3’ adapter sequence

--trim-n = remove ambiguous bases (N’s)

-l = shortening reads to a fixed length

-m = minimum read length

-o sample\_xx\_clean\_R1.fastq.gz = output processed reads (forward)

-p sample\_xx\_clean\_R2.fastq.gz = output processed reads (reverse)

sample\_xx\_R1.fastq.gz = input raw reads (forward)

sample\_xx\_R2.fastq.gz = input raw reads (reverse)

NOTE: The adapter sequence for WGS data trimming depends on the library kit employed. The adapter sequence recommended for the Illumina Nextera XT kit is CTGTCTCTTATACACATCT in both paired end reads.

1. ***E. coli* genome assembly using SPAdes3 (v.3.15.1) and Medusa4 web server**
   1. Depending on the OS, follow the instructions in the manual for the SPAdes assembler installation (https://cab.spbu.ru/software/spades/). Of note, the current version of SPAdes program is 3.15.4.
   2. Once installed, run the following command in the directory containing the quality-checked and trimmed sequencing data to perform contig-level assembly:

***$*** *spades.py --careful -k 21,33,55,77,99,107,117 -1 sample\_xx\_R1.fastq.gz -2 sample\_xx\_R2.fastq.gz -o output\_dir\_name*.

Where:

--careful = turn on the mismatch, short indels and error correction tool

-k = *k-mer* size to be used (comma-separated list)

-1 sample\_xx\_clean\_R1.fastq.gz = input processed reads (forward)

-2 sample\_xx\_clean\_R2.fastq.gz = input processed reads (reverse)

-o = output directory name for assembly results

* 1. If required, perform a scaffold-level assembly uploading the FASTA file containing the contig-level assembled genome and a closer reference genome to the Medusa server (http://combo.dbe.unifi.it/medusa). Use the complete genome of *E. coli* strain ESBL 15 (RefSeq accession GCF\_008370755.1) as reference for genome scaffolding of *E. coli* ACM5.

1. ***In-silico* typing of the *E. coli* genome using the Center for Genome Epidemiology (CGE; https://www.genomicepidemiology.org/services/) and ClermonTyping5 web platform**
   1. Serotype prediction
      1. Upload the FASTA file containing the assembled genome to the SeroTypeFinder6 tool (v.2.0). Set threshold settings: 85% nucleotide identity (%ID)/60% coverage. Click the **Upload** button to submit the genome sequence for analysis.
   2. Multi-locus sequence typing (MLST) determination
      1. Upload the FASTA file containing the assembled genome to the MLST7 tool (v.2.0). Select the ***Escherichia coli #1*** database option for analysis. Click the **Upload** button to submit the genome sequence to analysis.
   3. Virulence-associated gene annotation
      1. Upload the FASTA file containing the assembled genome to the VirulenceFinder8 tool (v2.0). Select the ***Escherichia coli*** database option for analysis. Set threshold settings: 90% ID/60% coverage. Click the **Upload** button to submit the genome sequence for analysis.
   4. Acquired antimicrobial resistance determinants annotation
      1. Upload the FASTA file containing the assembled genome to the ResFinder9 tool (v4.1). Tick the **A*cquired antimicrobial resistance gene*** option and select the ***Escherichia coli*** database option for analysis.
      2. Set threshold settings: 90% ID/60% coverage. Click the **Upload** button to submit the genome sequence for analysis.
   5. Plasmid replicon identification
      1. Upload the FASTA file containing the assembled genome to the PlasmidFinder10 tool (v2.1). Select the ***Enterobacteriales*** database option for analysis.
      2. Set threshold settings: 90% ID/60% coverage. Click the **Upload** button to submit the genome sequence for analysis.
   6. *E. coli* phylogroup determination
      1. Upload the FASTA file containing the assembled genome to the ClermonTyping web platform (http://clermontyping.iame-research.center/index.php). Run the analysis under the default setting parameters. Click the **Run** button to submit the genome sequence for analysis.
2. **Draft genome illustration with DNAPlotter11 tool (v.18.2)**

NOTE: DNAPlotter is a tool implemented by the Artemis software to render circular and linear genome representations. Depending on the OS, follow the instructions for the Artemis software installation(http://sanger-pathogens.github.io/Artemis/Artemis/).

* 1. Login to the RAST server (https://rast.nmpdr.org/). On the main page, go to the Annotation Progress column from the current *E. coli* genome annotation job and click on **View details**.
  2. From the drop-down menu in **Available downloads for this job**, select the GenBank merged type file (for use in, for example, Artemis). Click on the **Download** button.
  3. Execute DNAPlotter. Select the **Read in sequence file** option in the new displayed window. Navigate to the directory containing the file downloaded from the RAST server and load the *XXX.xxxx.merged.gbk* file (where *XXX.xxxx* is the job ID assigned by RAST).
  4. A default plot with some gene annotations will be shown in a new window. Add the GC plots to the circular genome representation. On the graph tab, tick the **Draw** option in the GC plot and GC Skew menu. Customize the colors of the GC plot and GC Skew in the options menu.
  5. Select the **Track Manager** menu on the options tab and customize the tracks 1 to 4. Delete the fifth track.
     1. Track 1: Click on the **Colour** button. Click on the color box next to **Apply Colour to All** to select the blue color (0, 0, 255 RGB code) for CDSs on the forward DNA strand. Click on **Apply Colour to All**.
     2. Track 2: Repeat step 5.5.1, now selecting the red color for CDSs on the reverse DNA strand (255, 0, 0 RGB code).
     3. Track 3: Select the **tRNA** key from the drop-down menu on the key column. Tick the **Not** box on the qualifier column. Repeat step 5.5.1, selecting the green color for tRNAs on both DNA strands (0, 255, 0 RGB code).
     4. Track 4: Select the **rRNA** key from the drop-down menu on the key column. Tick the **Not** box on the qualifier column. Repeat step 5.5.1, selecting the black color for rRNAs on both DNA strands (0, 0, 0 RGB code).
  6. On the options tab, select the **Features** menu to add annotations on the circle representation manually. Enter the start and stop position of feature on the genome. Click on the color box next to the stop legend and assign colors for each type of feature (e.g., antimicrobial resistance determinant [255, 0 ,0 RGB], virulence gene [255, 0, 255 RGB], and prophage region [0, 255, 0 RGB]).

**REFERENCES:**

1. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).

2. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17** (1), 10 (2011).

3. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19** (5), 455–477 (2012).

4. Bosi, E. et al. MeDuSa: a multi-draft based scaffolder. *Bioinformatics*. **31** (15), 2443–2451 (2015).

5. Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E., Clermont, O. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. *Microbial Genomics*. **4** (7), e000192 (2018).

6. Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., Scheutz, F. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. *Journal of Clinical Microbiology*. **53** (8), 2410–2426 (2015).

7. Larsen, M. V. et al. Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of Clinical Microbiology*. **50** (4), 1355–1361 (2012).

8. Joensen, K. G. et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *Journal of Clinical Microbiology*. **52** (5), 1501–1510 (2014).

9. Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *The Journal of Antimicrobial Chemotherapy*. **75** (12), 3491–3500 (2020).

10. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy*. **58** (7), 3895–3903 (2014).

11. Carver, T., Thomson, N., Bleasby, A., Berriman, M., Parkhill, J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*. **25** (1), 119–120 (2009).