WorldCOM Deliverable 1: Prevalence of ESBL subtypes in bacterial pathogens and a sequence database of selected alleles

Arnoud H. M. van Vliet[1,*], Marwa M. Hassan[1], Owen Higgins[2], Liam P. Burke[2], Louise O'Connor[2], Dearbhaile Morris[2], Rob M. La Ragione[1], Terry Smith[2]

1. University of Surrey, Guildford, UK
2. NUI Galway, Ireland

* Correspondence: a.vanvliet@surrey.ac.uk

This dataset is connected to Work Package 1, Task1 of the WorldCOM consortium grant within the One Health EJP group. The aim was to analyse publicly available sequences for antimicrobial resistance genes associated with *Salmonella*, *Campylobacter* and *E. coli*. For the initial phase of this work package, we have focused on ESBL-related AMR genes. As these genes are absent from *Campylobacter*, we have not included this bacterium in these analyses, and have used the important pathogens *Klebsiella* and *Acinetobacter*. All types and subtypes of Extended Spectrum β-Lactamases (ESBLs) and plasmid-mediated colistin resistance genes have been analysed for frequency among reported and extracted sequences. High frequency resistant genes subtypes have been highlighted for further sequence analysis to illustrate geographic distribution and geographic-specific single nucleotide polymorphisms (SNPs). The task is still ongoing.

Files in archive:

1) **ESBL_subtypes_ecoli_klebsiella_salmonella_acinetobacter_14may2020.xlsx**
Collated counts of ESBL beta-lactamase gene alleles and *mcr* gene alleles for *E. coli/Shigella*, *Salmonella*, *Klebsiella pneumoniae* and *Acinetobacter baumannii*, and selection of the most common alleles for each bacterial pathogen.

2) **ESBL_betalactamase_subtypes_ecoli_klebsiella_salmonella_acinetobacter_14may2020.xlsx**
Individual counts of ESBL beta-lactamase gene alleles and *mcr* gene alleles for *E. coli/Shigella*, *Salmonella*, *Klebsiella pneumoniae* and *Acinetobacter baumannii*, for selection and comparison of main alleles for further study.

3) **NCBI_pathogens_data_ecoli_10April2020.xlsx**

Extracted data for 105,165 *Escherichia coli* and *Shigella* spp. genomes from the NCBI Pathogens database, downloaded 10 April 2020, with specific ESBL beta-lactamase gene alleles and *mcr* gene alleles counted.

4) NCBI_pathogens_data_salmonella_10april2020.xlsx

Extracted data for 270,559 *Salmonella* spp. genomes from the NCBI Pathogens database, downloaded 10 April 2020, with specific ESBL beta-lactamase gene alleles and *mcr* gene alleles counted.

5) NCBI_pathogens_data_klebsiella_10april2020.xlsx

Extracted data for 24,799 *Klebsiella pneumoniae* genomes from the NCBI Pathogens database, downloaded 10 April 2020, with specific ESBL beta-lactamase gene alleles and *mcr* gene alleles counted.

6) NCBI_pathogens_data_acinetobacter_28april2020.xlsx

Extracted data for 8,039 *Acinetobacter* genomes from the NCBI Pathogens database, downloaded 28 April 2020, with specific ESBL beta-lactamase gene alleles and *mcr* gene alleles counted.

7) WorldCOM_ecoli_ESBL_NCBI_selection_10june2020.xlsx

Metadata of 5,199 *E. coli / Shigella* genomes filtered from File 1 for presence of *mcr1* alleles, *bla-*CTX-M-15, *bla-*NDM-5, *bla-*KPC-2 and *bla-*OXA-48.

8) WorldCOM_salmonella_ESBL_NCBI_selection_10june2020.xlsx

Metadata of 410 *Salmonella* genomes filtered from File 1 for presence of *mcr1* alleles, *bla-*CTX-M-15, *bla-*NDM-5, *bla-*KPC-2 and *bla-*OXA-48.

9) WorldCOM_klebsiella_ESBL_NCBI_selection_10june2020.xlsx

Metadata of 4,789 *Klebsiella pneumoniae* genomes filtered from File 1 for presence of *mcr1* alleles, *bla-*CTX-M-15, *bla-*NDM-5, *bla-*KPC-2 and *bla-*OXA-48.

10) WorldCOM_acinetobacter_ESBL_NCBI_selection_10june2020.xlsx

Metadata of 25 *Acinetobacter baumannii* genomes filtered from File 1 for presence of *mcr1* alleles, *bla-*CTX-M-15, *bla-*NDM-5, *bla-*KPC-2 and *bla-*OXA-48.

11) selected_ESBL_MCR_genes.fasta

FASTA-formatted reference sequences of ESBL and MCR-alleles: *mcr1.1-1.27* and *bla*-CTX-M-15, *bla*-NDM-5, *bla*-KPC-2, *bla*-OXA-48. Please note that the sequences are not ordered.

METHODS

Datasets were downloaded from the NCBI Pathogens database (https://www.ncbi.nlm.nih.gov/pathogens/isolates/#) as comma-separated value (CSV) files, and further used in Microsoft Excel. The field with antibiotic resistances were distributed over individual cells, and only *bla* (beta-lactamase) genes and *mcr* genes (colistin resistance) were counted, including the individual allele number. Data were manually transferred to new worksheets for further processing, with the aim of selecting the most common alleles in each of the 4 target organisms (*E. coli*, *Salmonella*, *Klebsiella pneumoniae* and *Acinetobacter baumannii*). The genomes positive for these alleles were then selected for further investigation to check for inclusion of geographic regions such as EU countries.

The AMRfinder database (https://github.com/ncbi/amr/wiki/AMRFinder-database) included with the Abricate software (https://github.com/tseemann/abricate) was used to extract specific *bla* and *mcr* alleles.