107 Standardisation of Public Genomic Datasets of E. coli from Food-Animals: A First Step Towards Global Genomic Comparisons

João Pires[1;2], Jana Huisman[1], Sebastian Bonhoeffer[1], Thomas Van Boeckel[1;3]

[1]*Institute for Environmental Decisions, Department of Environmental Sciences, ETH Zurich, Zurich, Switzerland*, [2]*Institute for Integrative Biology, Department of Environmental Sciences, ETH Zurich, Zurich, Switzerland*, [3]*Center for Disease Dynamics, Economics & Policy, United States*

Background: The increased use of next-generation technologies in antibiotic-resistance epidemiology fueled the number of deposited genomes in public repositories. The heterogeneous nature of genomic data in these repositories alongside the lack of standardized metadata impairs accurate global epidemiological comparisons. This is especially problematic in data from food-animals, where isolation sources and hosts are very diverse. We aim to query *Escherichia coli* genomes of food-animal origin in public repositories to understand the global epidemiology of *E. coli* in this setting and identify biases in public datasets.

Materials/methods: A systematic search for *E. coli* genomes was performed in NCBI, PATRIC and EnteroBase databases. Metadata was standardized by assigning a generic host – poultry, cattle, swine and caprine/ovine -, an isolation location/country and date. Assemblies were downloaded and duplicates were removed. Assemblies were screened for acquired resistance genes (ARG) with ResFinder and population structure was analyzed using *in silico* MLST and Clermont typing.

Results: A total of 11'363 genomes had an identifiable food-animal host. Due to data accessibility, 4052 genomes from PATRIC and NCBI were analyzed. 176 were identified as non-*E. coli* by Clermont typing and were excluded. Most genomes were recovered from cattle (47.8%), followed by poultry (30.1%) and swine (17.3%). The bulk of genomes originate from USA (59.7%), and to lesser extent China (13.9%) and UK (11.6%). *E. coli* belonged mostly to B1 (43.6%) and A (28.8%) phylogenetic groups, which have been widely associated with *E. coli* from animals. The identified clonal complexes (CCs) are also consistent with previous food-animal reports: CC10 (10.8%), CC11 (6.9%) and CC155 (5.5%). 2363 genomes (61%) contained ARG, being the most frequent those conferring resistance to streptomycin (*aph(6)-Id*, 53.3%), sulphonamides (*sul2*, 52.6%), tetracyclines (*tet(A)*, 48.9%) and narrow-spectrum β-lactams (*bla*$_{TEM-1}$, 32.1%), which are ARG known to be present in integrons circulating among *E. coli*.

Conclusions: Public genomic repositories yielded information consistent with reports on antibiotic-resistant *E. coli* from food-animals. This highlights the reliability of using such data in global surveillance studies and detailed meta-analysis after metadata standardization. Given the uneven geospatial distributions among deposited genomes, analysis accounting for sampling bias will be crucial to fully exploit such datasets.