Title: Dynamics of mobile features in *Enterobacteriaceae*

Subtitle: Fine-scale surveillance of antimicrobial resistance genes, virulence genes and plasmids.

Supervisors

Zamin Iqbal, EMBL-EBI, computational microbial genomics (zi@ebi.ac.uk)
www.ebi.ac.uk/research/iqbal
Nicholas R Thomson (Wellcome Trust Sanger Institute) (nrt@sanger.ac.uk)
http://www.sanger.ac.uk/people/directory/thomson-nicholas-robert

In close collaboration with
Kathryn Holt, University of Melbourne (kholt@unimelb.edu.au)
www.holtlabnet

Abstract
The *Enterobacteriaceae* are a large family of gram-negative bacteria encompassing many species, including both commensals and opportunistic pathogens such as *E.coli*, *Salmonella*, *K. pneumoniae*, *Y. pestis*. Most of these commonly reside in the gut of animals, but also can be found in the environment. The spread of antimicrobial resistance genes amongst *Enterobacteriaceae* is a matter of global concern; carbapenemase-producing enterics are becoming more common, are generally resistant to almost all beta-lactams, and often to many other drug classes. Given gene flow between species in the gut is known to happen, as asymptomatic carriage of resistance genes in the gut becomes more common, there is greater risk of these occurring in bacteremia.

Resistance genes can spread not just via clonal expansion, but also by horizontal transfer via plasmid or transposons – in one recent outbreak in Virginia, the KPC gene was transferred across 12 genera by a transposon [1]. However the vectors that enable the transfer of these genes are themselves evolving – plasmid genomes undergo rearrangements and recombination, and transposons carry varying cargo and move within and between plasmids. Thus there is a hierarchy of mobile elements on mobile elements, and we are challenged to find the right level at which to perform analysis.
Although the bioinformatic methods for building phylogenies, placing samples in lineages, and tracing clonal outbreaks are well-developed, those for tracking the spread and evolution of horizontally transferred alleles are less mature, especially when potentially looking across genera and at structures which rearrange. **Zamin Iqbal**'s group at EMBL-EBI has been working on various computational problems that are designed to be of use here. First, methods to detect an allele in raw sequence data. Second, similarity metrics for plasmids that are robust to rearrangement. Third, indexing huge (>100,000 samples) multi-sample assembly graphs, allowing cross-sample comparisons. **Nick Thomson** has been studying many aspects of enterobacteriaceae, including plasmid structures, accelerated horizontal transfer in the inflamed gut [2], virulence and antimicrobial resistance, and has a long-standing

collaboration with **Kat Holt**. In particular they have studied the global diversity of *K. pneumoniae* and the occurrence of virulence and resistance therein.

Details:
This project seeks to

1. Develop whole-genome-assembly-free methodology for analysis of (resistance and virulence) gene and plasmid dynamics across lineage backgrounds based on novel genome sequence index data structures.
2. Build signatures of plasmid types (backbones) and of transposons
3. Track individual and co-occurring signatures across samples, lineages and time, focusing on *K. pneumoniae*, but adding any other samples containing related plasmids
4. Explore dynamics in smaller datasets with well-understood epidemiology, and then apply to global data.
5. Identify, obtain, culture, sequence (Illumina+Nanopore) and assemble any isolates with informative replicons, with a view to constructing a high quality catalog or repertoire of various mobile elements.

Items 1-4 would be primarily driven in the Iqbal/EMBL-EBI group, and Item 5 in the Thomson/WTSI lab.

Skills, Training and support provided:
From **Iqbal/EMBL-EBI** group: training in use of genome graph software, analysis of complex alleles, support from user-interface programmer and underlying genome graph developer, nanopore sequence analysis.
From **Thomson/Sanger** group: plasmid, transposon and *Enterobacteriaceae* biology, wet-lab techniques for culture and sequencing which preserve plasmids.
From **Holt/Melbourne** collaboration: expertise in genomic epidemiology, phylogenetics, *K. pneumoniae* biology.

Datasets:
1. 281 *Enterobacteriaceae* from a transposon-mediated hospital outbreak in Virginia, including both Illumina and high-quality PacBio assemblies [1]
2. All *Klebsiella* from the Alfred Hospital (Melbourne) during a 1 year period, plus carriage screening in the intensive care unit, (N=~ 600) [3]
3. All *Enterobacteriaceae* in the Short Read Archive.

References

1. Sheppard et al, “Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC”, AAC (2013)
2. Stecher et al, “Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae”, PNAS (2012)
3. Gorrie et al, “Gastrointestinal carriage is a major reservoir of K. pneumoniae infection in intensive care patients”, Biorxiv, 2016