**Title:** Investigating species boundaries for the multidrug-resistant Enterobacter cloacae complex (ECC).

**Student:** Susan Noonan

**Supervisor team:** Prof Ben Howden, Dr Jake Lacey and Dr Kristy Horan.

**Project Summary:** The Enterobacter cloacae complex (ECC) includes opportunistic, nosocomial pathogens of six Enterobacter species capable of producing a wide variety of infections. Broad-spectrum antibiotic resistance, including the emergence of resistance to last resort carbapenems, has led to increased interest in this group of organisms. This project will utilise microbial genomics, bioinformatic and statistical analysis to quantify the species boundaries within the ECC to better assist in accurate identification for public health surveillance.

**Section A: Literature review on species boundaries**

* How are the current species defined?
* How are the subspecies defined?
* Is it mostly phenotypic data or genomic data that is used for Enterobacter classification?
* Are there any genes used to defined species/subspecies?
* What typing schemes are used? Eg. MLST (multi-locus sequencing typing) is definitely one, are there any other classification schemes we can use to help differentiate isolates?
* Are there any special virulence factors (genes that help with disease) that we should be aware of?
* What are the most commonly reported/concerning antimicrobial resistance genes found in human infections of Enterobacter.

**Yeh et all (2022)** – review article that outlines all the MDR genes reported from hormachei isolates from around the world.

* Extended-spectrum β-lactamase (ESBL) genes in Enterobacter hormaechei
* genes encoding β-lactamases, including ESBLs, AmpC and carbapenemases
* ESBLs, including SHV, CTX-M and SFO
* Carbapenemase-producing ECC (CP-ECC)
* AmpC β-lactamases belong to Ambler class C. They confer re- sistance to expanded-spectrum β-lactams and cephamycins but are not susceptible to clavulanic acid
* Colistin (polymyxin E), Polymyxins are one of the last- line therapeutic options for treating serious infections caused by CRE and have been extensively used in the animal breeding industry for treating infections caused by Enterobacterales
* Several case reports have revealed tigecycline resistance in E. hormaechei.

**Section B: Clustering to evaluate species boundaries of public Enterobacter species**

* Download closed genome representatives of all Enterobacter species.
* Perform Average nucleotide identity (ANI) using mash to cluster isolates into genomic groups.
  1. (Note: according to research into Klebsiella which is a similar species in terms of AMR and genome structure usually clusters of isolates within a 0.02 pairwise distance is regarded as same species and within 0.04 is a poorer match but closely related species).
* Examine closer the within species groups – eg. Enterobacter hormacheii sub-species, what thresholds separate into established subspecies?
  1. Is there really 7 sub-species as determine by genomics? Or are some not really a thing?
  2. Same thing for the cloacae subspecies. Are they real or not.
* How does the genomic clustering align with names/uploaded species names?
  1. I suspect most of the sub-species should be ok but the more generic names, eg. Enterobacter cloacae might be uploaded incorrectly into the wrong species groups but that’s only a guess.

**Section C: Comparison of typing schemes with defined species**

* Extend the dataset to include all genome from refseq not just the close genomes
* Run MLST on all isolates
* Determine species designation for all isolates using same method as Section B.
* Is each MLST unique to a species/subspecies? Are any MLST types shared across species? Can we use MLST as an indicator for what species/subspecies an isolate is?
* Is there any genes from Section A we can screen to add to this screen? To help defined these groups.

**Section D: Genomics epidemiology of defined species/subspecies**

* Using the results from Section B and C as a baseline we can then group isolates (whether its species or subspecies will depend on results) and run some tools to establish the what the population stricture is within each of these groups. This can tell us a lot about how the pathogen is evolving and how different groups are interacting and exchanging genes.
* We will examine structure using PopPUNK and compare that to all the typing scheme metrics
* We will use Panaroo to define the pan-genome within each group and then examine the antimicrobial resistance profiles using Abritamr.
* Depending on Section A (virulence factors) this will be where we then screen for those.
* Overall question to ask:
  1. are there any trends within this data based on Virulence and AMR profiles?
  2. What does the population structure look like? Do all the species/subspecies appear similar or does each species have a different structure?
  3. How does this compare to other Enterobacteriaceae families eg. E.coli, Kelbsiella etc.