

μ bioinfo Virtual Lab Talk:

Soil-persistent *E. coli* and Mobile elements

Nicholas Waters

April 1, 2019

National University of Ireland, Galway, Ireland
The James Hutton Institute, Dundee, Scotland

Background



- *E. coli* has been found to persist stably in the soil
- Isolates were cultured from lysimeter leachate
- Strains were sequenced, resulting in 149 soil-persistent *E. coli* genomes



• What types of *E. coli* are able to persist in soil?



- What types of *E. coli* are able to persist in soil?
- What virulence factors are harboured by these strains?



- What types of *E. coli* are able to persist in soil?
- What virulence factors are harboured by these strains?
- What can we infer about adaptation?



- What types of *E. coli* are able to persist in soil?
- What virulence factors are harboured by these strains?
- What can we infer about adaptation?
- Can we differentiate soil-persistent *E. coli* from recent contamination?

E. coli Pangenome



Assembly \Rightarrow Annotation \Rightarrow presense/absence matrix

Assembly \Rightarrow Annotation \Rightarrow presense/absence matrix
riboSeed \Rightarrow Prokka \Rightarrow Roary

Assembly \Rightarrow Annotation \Rightarrow presense/absence matrix
riboSeed \Rightarrow Prokka \Rightarrow Roary

	N	Core	total
Soil	149	2662	21,662
Enterobase	1193	1822	79,288
All	1342	1806	83,868

Detecting differential presence/absence



- ⌘ No gene(s) correlated with source
- ⌘ *E. coli* diversity adds noise

What does 10 years look like to *E. coli* ?

BoE Calculations for Doubling Time



$\rho \approx 72\text{k generations}$

$\rho \approx 10\text{k generations}$

Other estimates



No viable coliforms after:

- o 22 days
- o 32 days
- o 8 weeks
- o 16 weeks



E. coli approximately .092% prevalence in soil metagenomes

Hypothesis



Stressed and outnumbered?

Hypothesis



Stressed and outnumbered?

rapid > incremental

Hypothesis



Stressed and outnumbered?

rapid > incremental

∴ adaptation via mobile elements

Mobile Genetic Elements

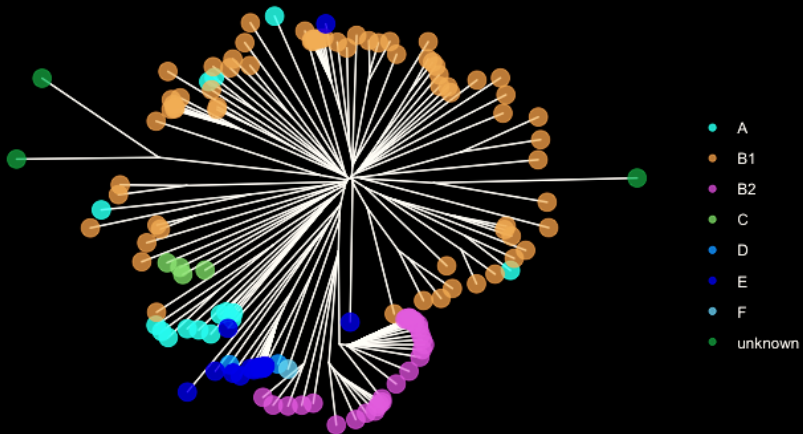


	Kb	Mobility	Detection
IS	.75-1.5	transposase	ITR and transposase
Non-composite Tn	<3	IS	IS with bonus genes
Composite Tn	<5kb	double IS	flanking ISs
Integron	<10	integrase	<i>attI</i> site, AMR genes
Genetic Island	>10	phage/plasmid	GC Skew, ORF phylogeny
Phage Inducible CI	5-15	phage	phage inhibitors
Prophage	≈50	lysogeny	integrase, tail, capsule
Plasmid	1-1Mb	various	run a gel, <i>oriT</i>

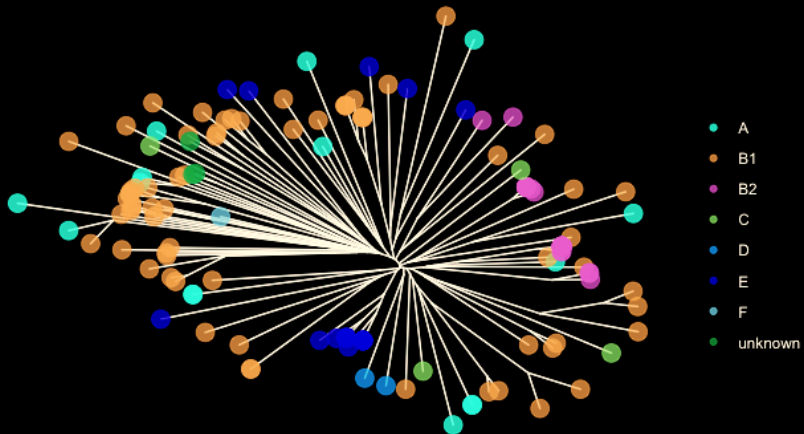
Horizontally Aquired Partial Pangenome of Inserted Elements

1. Run QC on genome assembly:
 - remove contigs $< 800\text{bp}$
2. Annotate whole genome (Prokka)
3. Filter out incomplete annotations with AnnoFilt
4. Detect mobile elements
 - Plasmids via mlplasmids
 - Prophages via ProphET
 - GIs via IslandPath-DIMOB
5. Extract and annotate mobile genome
6. Query both mobile and total genomes
 - detect AMR with resfinder
 - detect virulence genes with VFDB

Genomic NJ on Pangenome

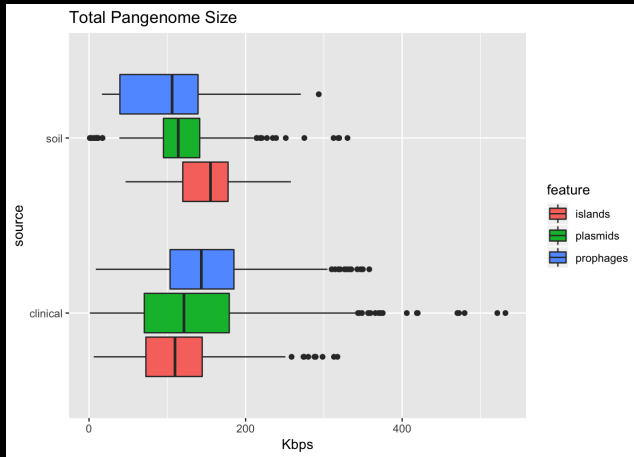


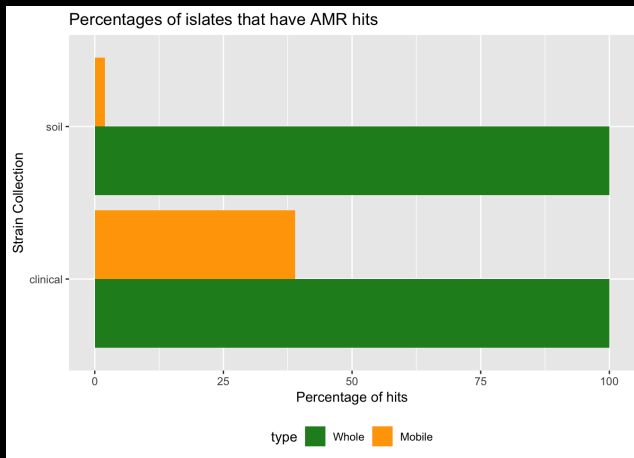
Genomic NJ on Pan-mobilome



Comparing the pan-mobilome

Overview





Enrichment Analysis (in progress)



- pyseer: kmers
- panX: diversity and visualization
- ML-based approaches

Enrichment Analysis (in progress)



- pyseer: kmers
- panX: diversity and visualization
- ML-based approaches

cold shock?

In Closing



What we know:

- soil *E. coli* is very diverse
- wide range of estimates exist for doubling time in soil
- diversity prevents robust statistical trait association
- MGE phylogeny does not reflect genome phylogeny
- AMR genes may be less frequently found in soil pan-mobilome

What we want to know:

- how the mobile pangenomes of soil vs enteric *E. coli* compare
- which *types* of phages, etc, are correlated by source
- which regions are interrupted by IS, Tns, etc

Acknowledgments



NUIG Microbiology

- ☐ Dr. Fiona Brennan
- ☐ Dr. Florence Abram
- ☐ Soil and Environmental Microbiology Research Group
- ☐ Functional Environmental Microbiology Group



James Hutton Institute, Dundee

- ☐ Dr. Leighton Pritchard
- ☐ Dr. Ashleigh Holmes