

2019 GRC:

# Soil-persistent *E. coli*, mobile elements, and the pan-mobilome

..and other updates

---

Nick Waters

May 8, 2019

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

## Background

---

# Project Overview



- ⊖ *E. coli* persists 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 AMR/virulence factors are harboured by these strains?
- ❑ What can we infer about adaptation?

## *E. coli* Pan-genome

---



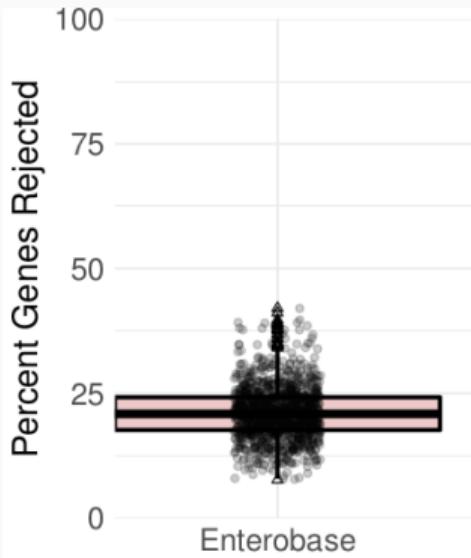
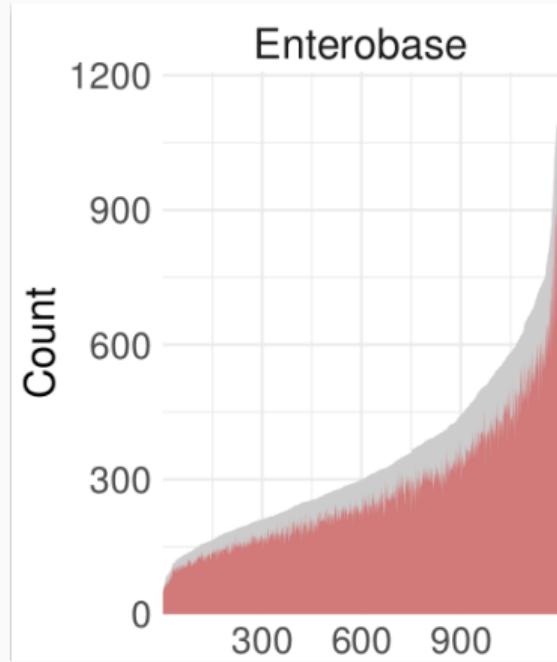
Assembly ⇒ Annotation ⇒ presence/absence matrix  
riboSeed ⇒ Prokka ⇒ Roary

- ✓ 149 soil strains
- ✓ 1193 clinical<sup>1</sup> strains
- ✓ 2937 core genes (>95% of genomes)
- ✓ 44,567 total genes
- ✓ 90% blastp identity threshold

---

<sup>1</sup>representatives from each sequence type on Enterobase

# Annotation QC with annofilt



<https://github.com/nickp60/annofilt>





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

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

---

# Estimates from the literature



## Estimates

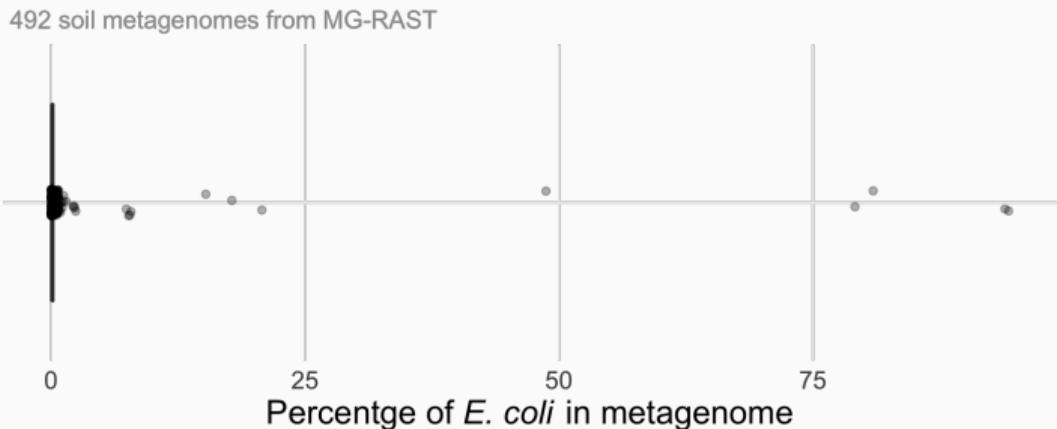
ꝝ ≈ 72k generations

ꝝ ≈ 10k generations

## No viable coliforms after:

ꝝ 22 days

ꝝ 16 weeks



# Hypothesis



Stressed and outnumbered?



Stressed and outnumbered?

rapid > incremental



Stressed and outnumbered?

rapid > incremental

. ∴ adaptation via mobile elements

# Mobile Genetic Elements

---

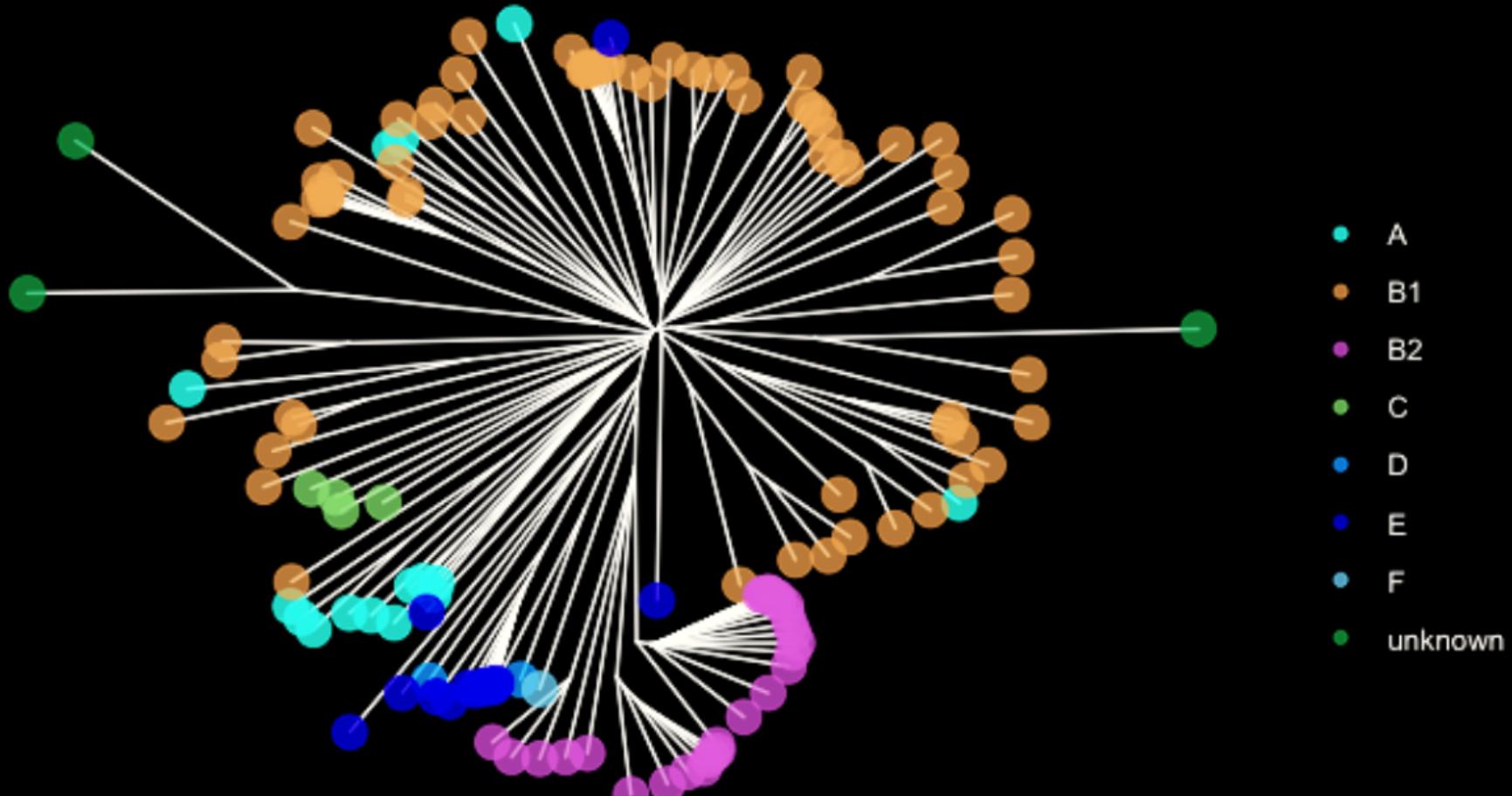
	Kb	Mobility	Detection
IS	.75-1.5	transposase	ITR and transposase
Non-composite Tn	<3	IS	IS with bonus genes
Composite Tn	<5kb	double IS	flanking ISs
Integron	<10	integrase	<i>attl</i> site, AMR genes
Genetic Island	>10	phage/plasmid	GC Skew, ORF phylogeny
Phage Inducible CI	5-15	phage	phage inhibitors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



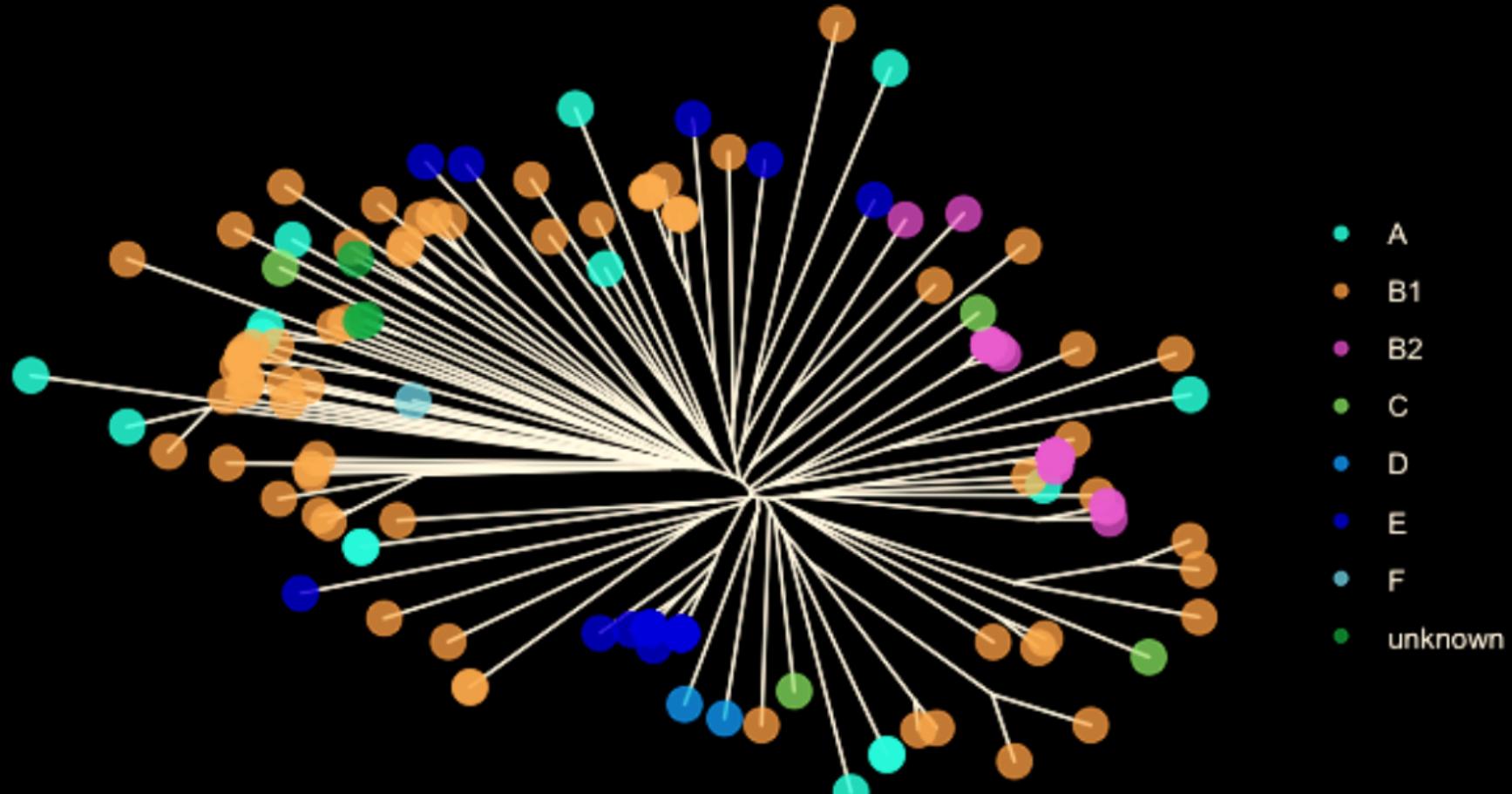
## Horizontally Acquired Partial Pan-genome of Inserted Elements

1. Check quality of genome assembly:
  - ✓ remove contigs < 800bp
2. Annotate whole genome with Prokka
3. Filter out incomplete annotations with AnnoFilt
4. Detect mobile elements
  - ✓ Plasmids via mlplasmids or mob-tools
  - ✓ Prophages via ProphET
  - ✓ GIs via IslandPath-DIMOB
5. Extract and annotate mobile genome
6. Query both mobile and total genomes with abricate
  - ✓ detect AMR with resfinder
  - ✓ detect virulence genes with VFDB

# Pan-genome neighbor-joining tree



# Pan-mobilome neighbor-joining tree

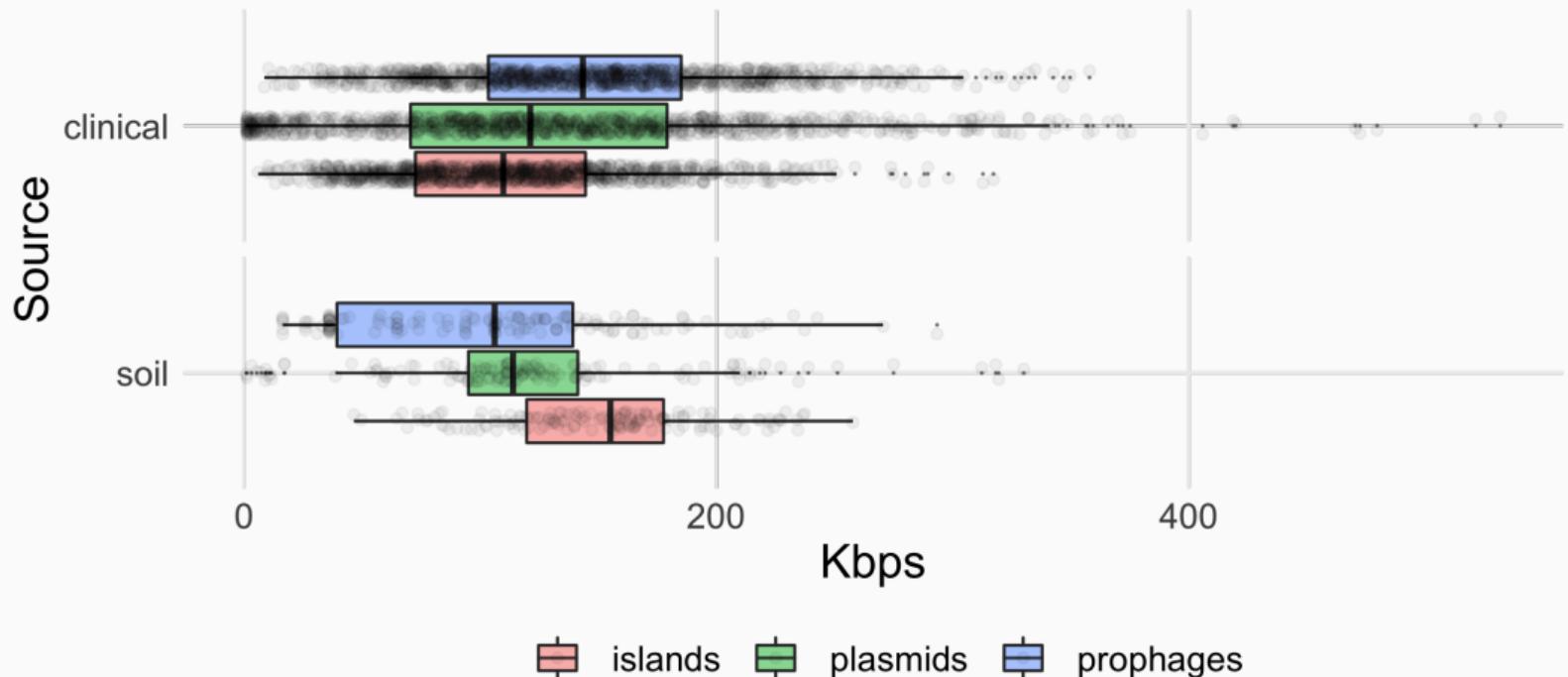


## Comparing the pan-mobilome

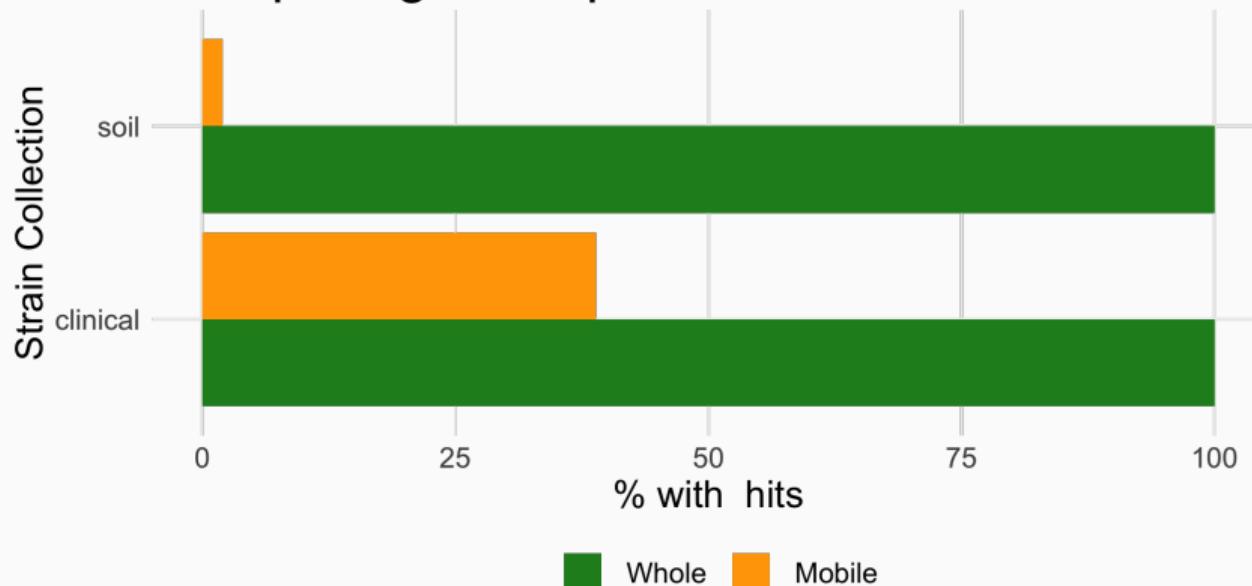
---

# Pan-mobilome size

Length of all mobile elements per genome



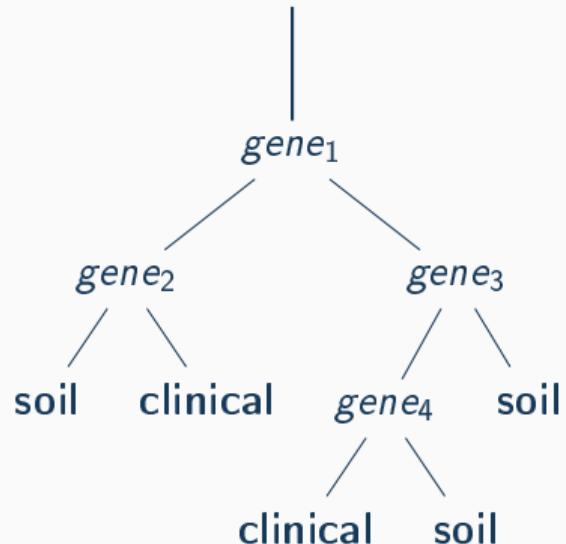
## Comparing AMR profiles between source



## Enrichment Analysis

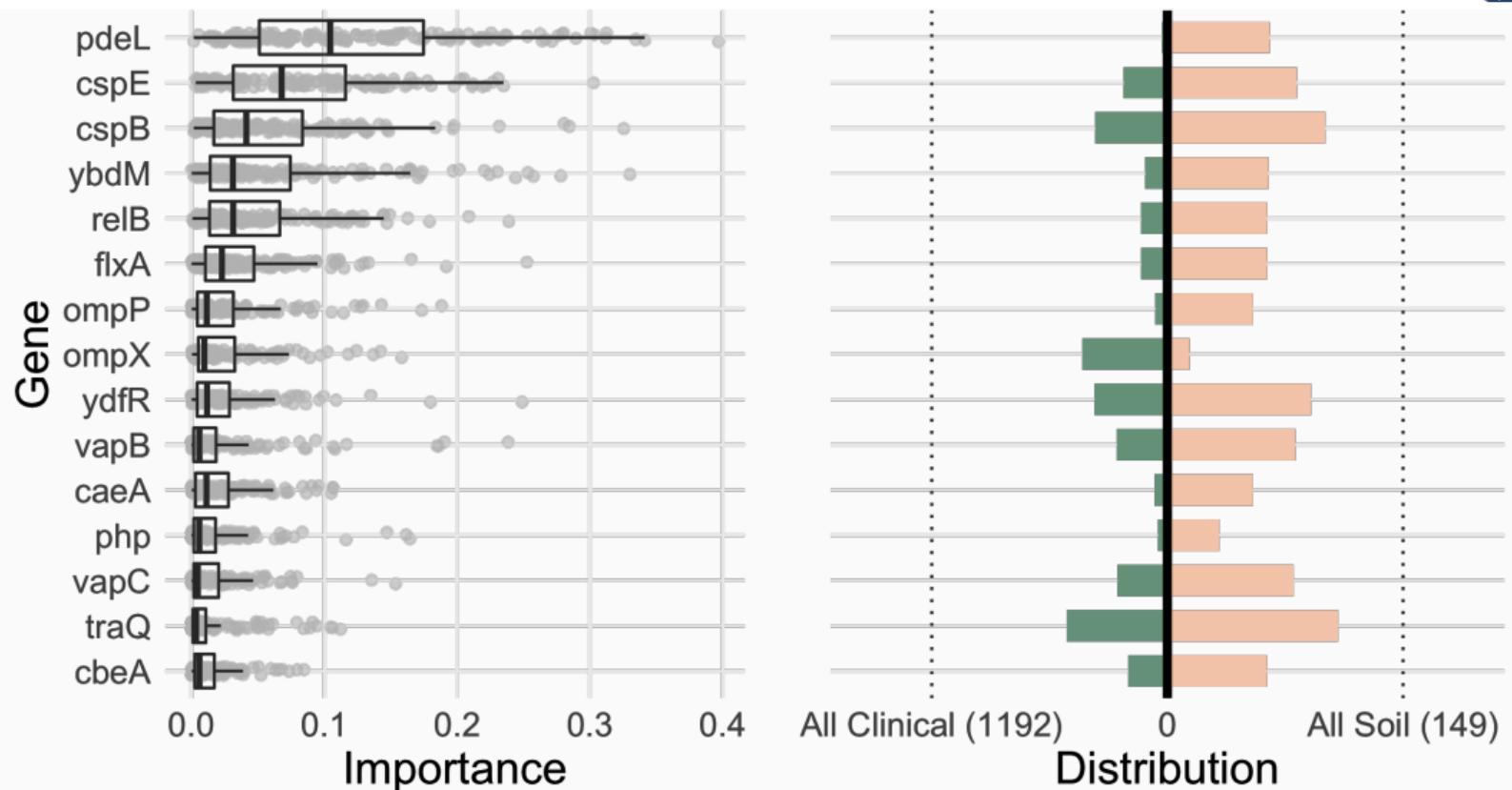
---

# Enrichment Analysis with Random Forests

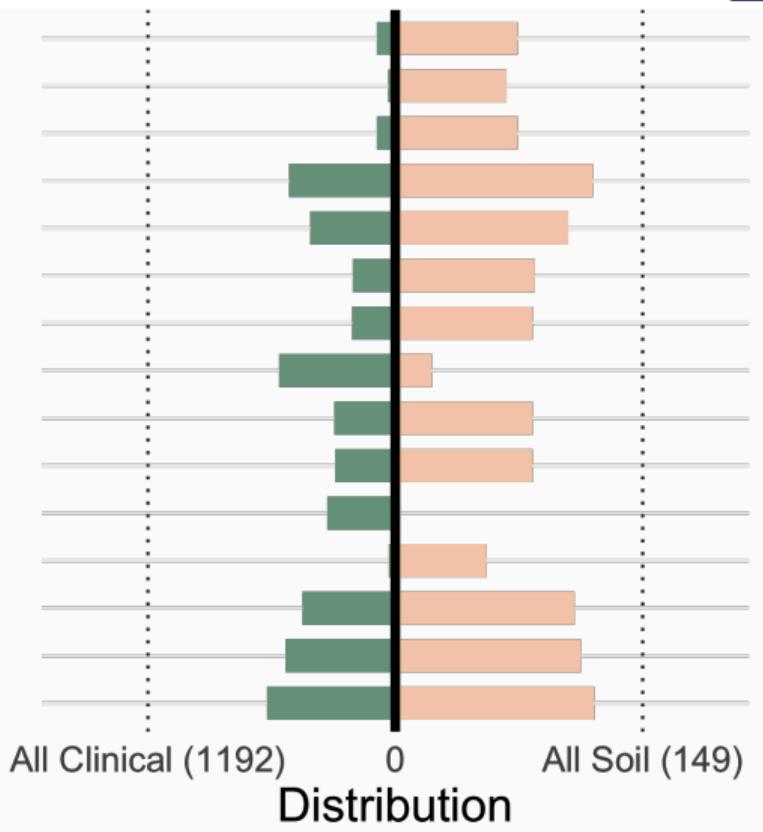
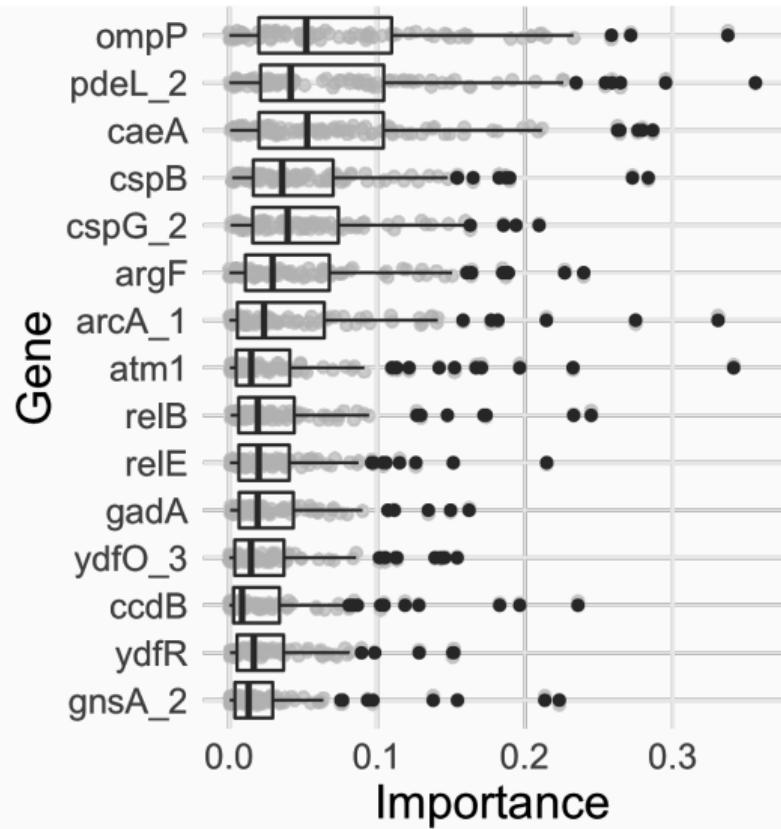


- ↪ Supervised machine learning
- ↪ Randomly generate decision trees
- ↪ Assess, repeat

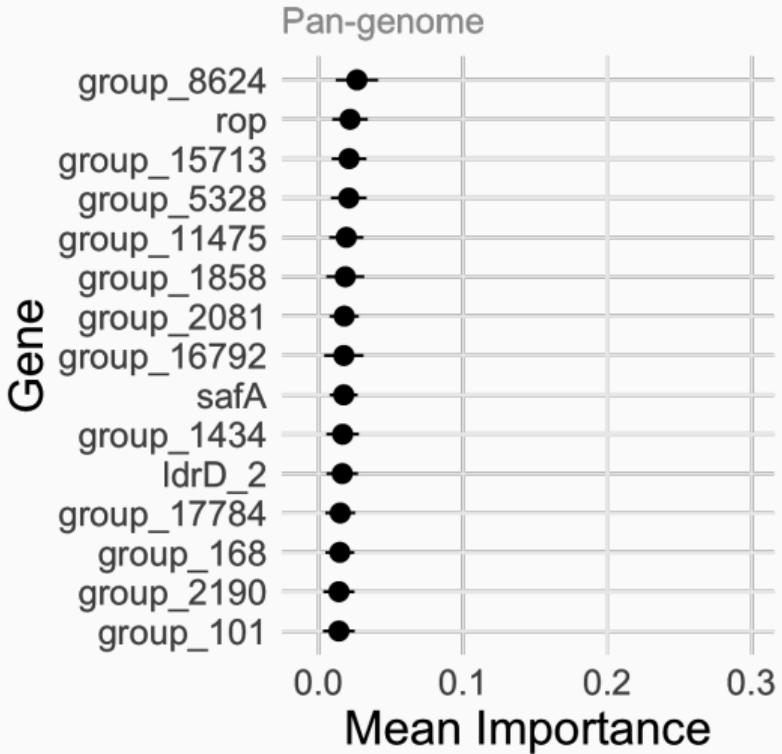
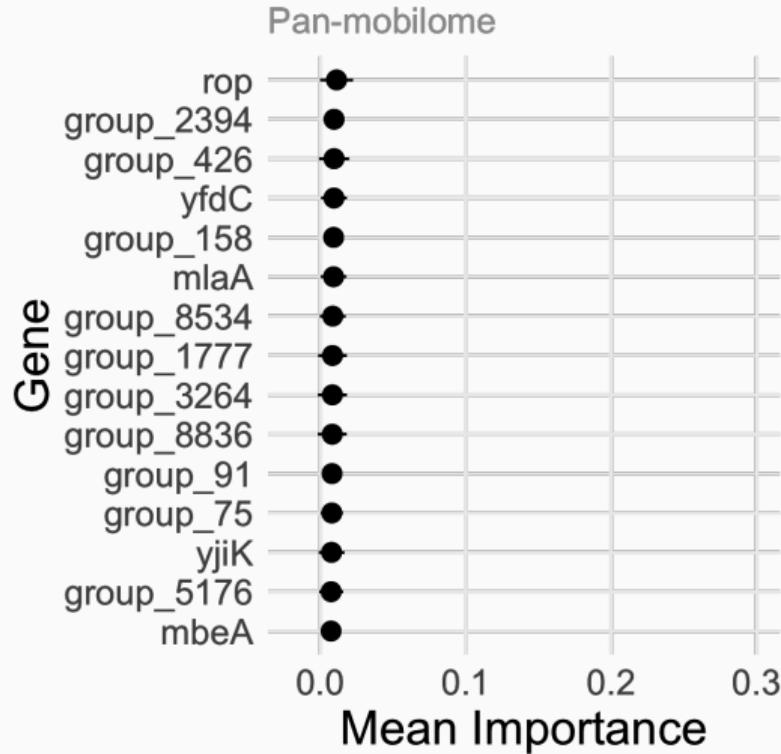
# Important Genes in Pan-mobilome



# Important Genes in Pan-genome



# Shuffled Classes





- ø paralog analysis
- ø study enriched kmers
- ø annotation curation
- ø

In Closing

---



## What we know:

- ↪ soil *E. coli* is very diverse
- ↪ wide range of estimates exist for doubling time in soil
- ↪ MGE phylogeny does not reflect genome phylogeny
- ↪ AMR genes may be less frequently found in soil pan-mobilome
- ↪ mobilized cold-shock genes may aid soil-persistence

## What we want to know:

- ↪ what inferences can we make about adaptation?
- ↪ which *types* of phages, plasmids, etc, are correlated by source?
- ↪ what role does cyclic di-GMP and phosphodiesterases play?

## Publications

---



## Published/Preprinted:

- ø "Easily phlyotyping *E. coli* via the EzClermont web app and command-line tool" by **Nicholas R. Waters**, Florence Abram, Fiona Brennan, Ashleigh Holmes, and Leighton Pritchard, *BiorXiv* <https://www.biorxiv.org/content/10.1101/317610v1>
- ø "Toward Assessing Farm-Based Anaerobic Digestate Public Health Risks: Comparative Investigation With Slurry, Effect of Pasteurization Treatments, and Use of Miniature Bioreactors as Proxies for Pathogen Spiking Trials", by Stephen Nolan, **Nicholas R Waters**, Fiona Brennan, Agathe Auer, Owen Fenton, Karl Richards, Declan J Bolton, Leighton Pritchard, Vincent O'Flaherty, Florence Abram, *Front. Sustain. Food Syst.*, 20 July 2018 <https://doi.org/10.3389/fsufs.2018.00041>

## Submitted:

- ø "The potential for polyphosphate metabolism in Archaea and anaerobic polyphosphate formation in *Methanosarcina mazei*" by Fabiana Paula, Jason Chin, Anna Schnürer, Bettina Muller, Panagiotis Manesiotis, **Nicholas R Waters**, Katrina Macintosh, John P Quinn, Jasmine Connolly, Florence Abram, John McGrath, and Vincent O'Flaherty

## In preparation:

- ø "Chordomics: a visualisation tool for linking function to phylogeny in microbiomes" by Kevin McDonnel, **Nicholas R Waters**, Enda Howley, Florence Abram. To be submitted to [Bioinformatics](#).



- ø Lightning talk, Microbiomes Underpinning Agriculture, Cork, Oct 2018
- ø Presentation,  $\mu$  bioinfo slack group jornal club, March 2019
- ø Presentation in "Microbial Pangenomes" session, MS Annual Meeting, Belfast, April 2019

- ⌚ [github.com/nickp60/plentyofbugs](https://github.com/nickp60/plentyofbugs) – semi-automated reference strain identification
- ⌚ [ezclermont.org](https://ezclermont.org) – webapp for in silico clermont typing
- ⌚ [github.com/nickp60/happie](https://github.com/nickp60/happie) – pan-mobilome pipeline
- ⌚ [github.com/nickp60/riboSeed](https://github.com/nickp60/riboSeed) – better short-read assemblies
- ⌚ [github.com/nickp60/annofilt](https://github.com/nickp60/annofilt) – filter short-read assembly annotations

## Completion Plan Deadlines

---



Working Title: "A collection of soil-persistent *E. coli* isolates from Irish soils" Story:

- ❑ A brief review/history of env *E. coli*
- ❑ *E. coli*, soil, and the environment
- ❑ Background of Teagasc experiment
- ❑ *E. coli* from soil are very diverse

Results:

- ❑ strain isolation and QC
- ❑ assembly summary
- ❑ Phylogenetic Analysis
- ❑ MLST, Clermont, and SNP types



Working title: Adaptive strategies in *E. coli* soil persistence Story:

- ❑ Selection of a comparison dataset
- ❑ Assembly, annotation, and pangenome construction

Results:

- ❑ Mobile elements may reveal trends obscured by traditional methods
- ❑ Cold shock proteins are enriched in soil persistent strains
- ❑ metabolic signaling mediated by pdeL impacts persistence



Working title: Filtering annotations from contig ends with Annofilt, and the impact on pangenome analysis Story:

- ☞ problems with genome assembly and annotation
- ☞ impact of annotation quality on pangenome analysis

Results:

- ☞  $\approx 25\%$  contig end annotations are incomplete



Working title: Using happy to infer recent adaptive strategies: a pipeline for panmobilome analysis Story:

- ☞ Intraspecies diversity make GWAS difficult

Results:

- ☞ mobile elements may reveal trends obscured by other enrichment analyses

February 2019: Thesis submission

