

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

---

Nick Waters

April 10, 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

---

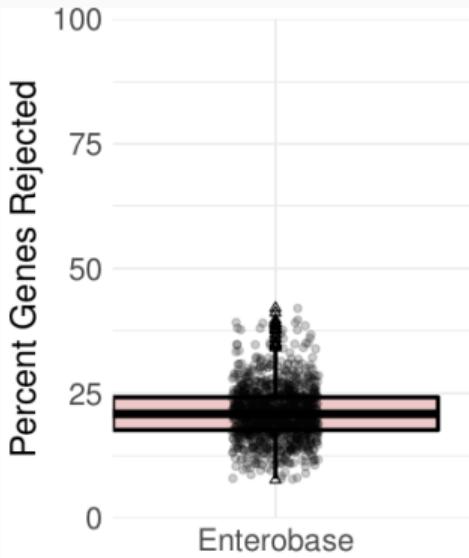
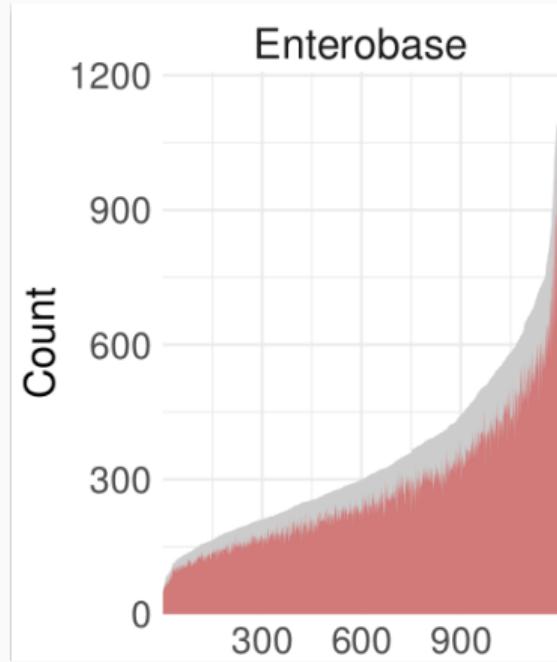
Assembly ⇒ Annotation ⇒ presence/absence matrix  
riboSeed ⇒ Prokka ⇒ Roary

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

---

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

# Annotation Correction with annofilt performance



<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

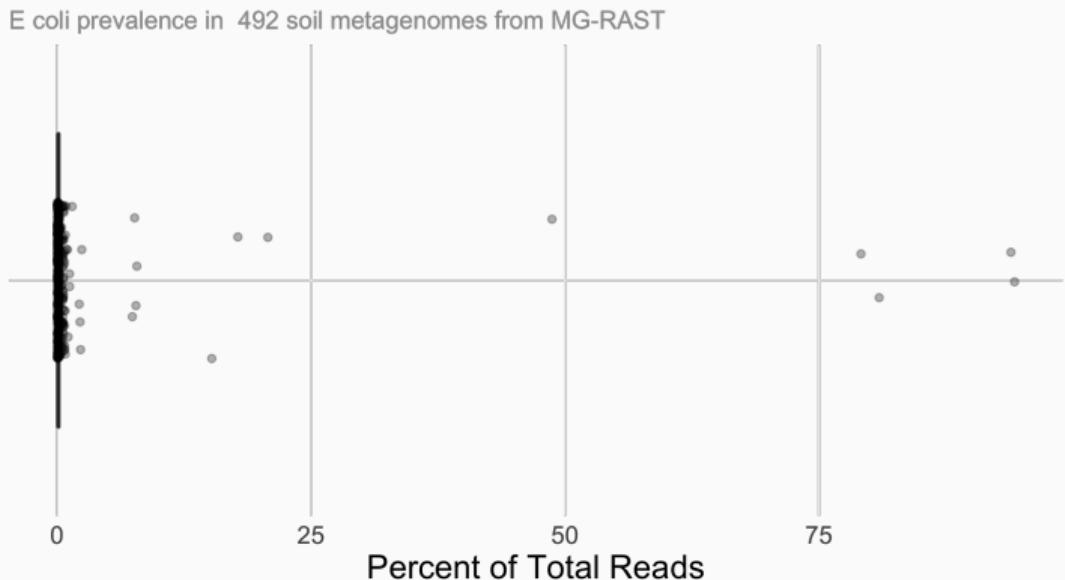


## Estimates

- ≈ 72k generations
- ≈ 10k generations

No viable coliforms after:

- 22 days
- 16 weeks



# Hypothesis



Stressed and outnumbered?



Stressed and outnumbered?

rapid > incremental



Stressed and outnumbered?

rapid > incremental

. ∴ adaptation via mobile elements

# Mobile Genetic Elements

---

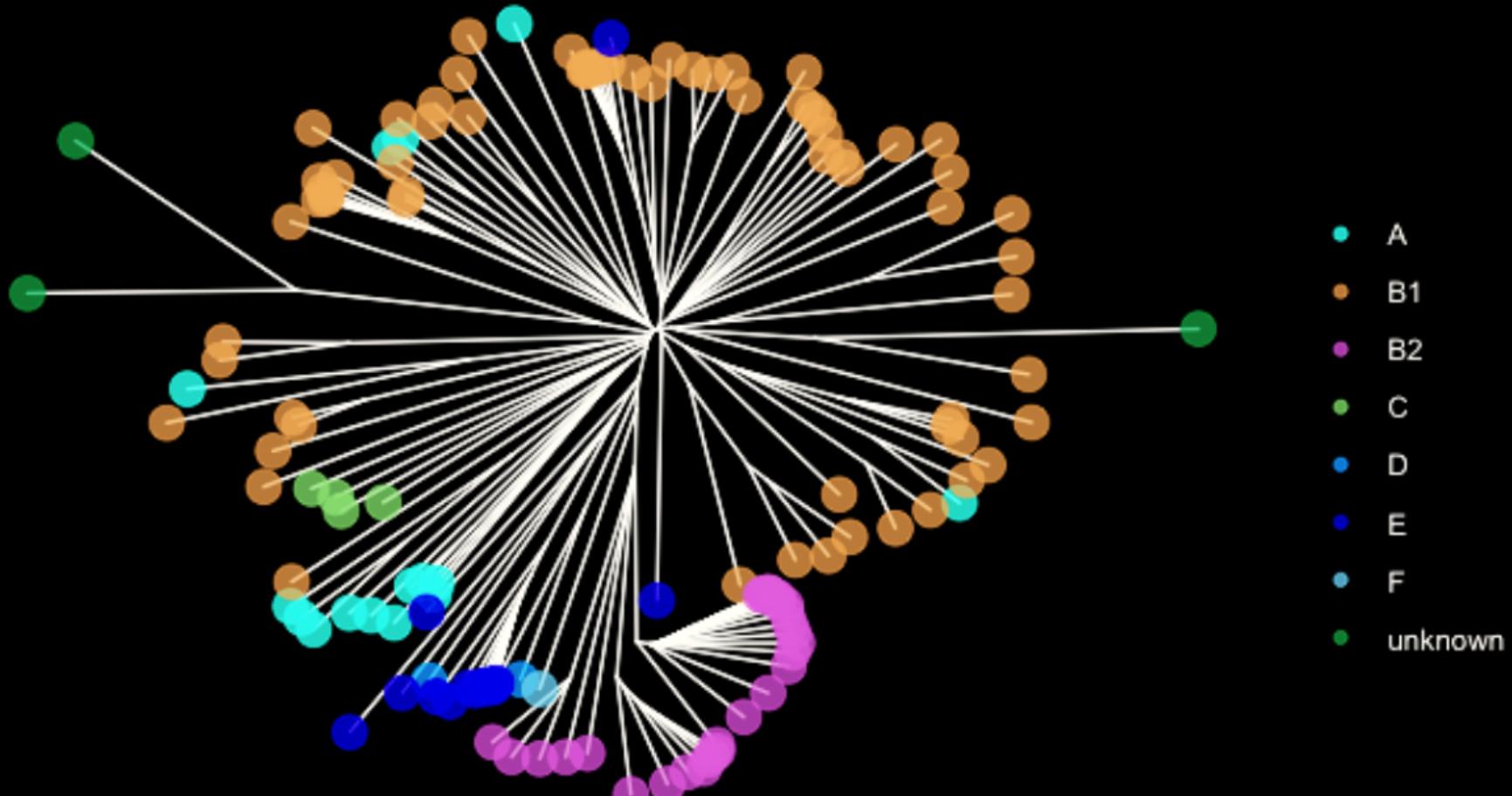
	Kb	Mobility	Detection
IS	.75-1.5	transposase	ITR and transposase
Non-composite Tn	<3	IS	IS with bonus genes
Composite Tn	<5kb	double IS	flanking ISs
Integron	<10	integrase	<i>attl</i> site, AMR genes
Genetic Island	>10	phage/plasmid	GC Skew, ORF phylogeny
Phage Inducible CI	5-15	phage	phage inhibitors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



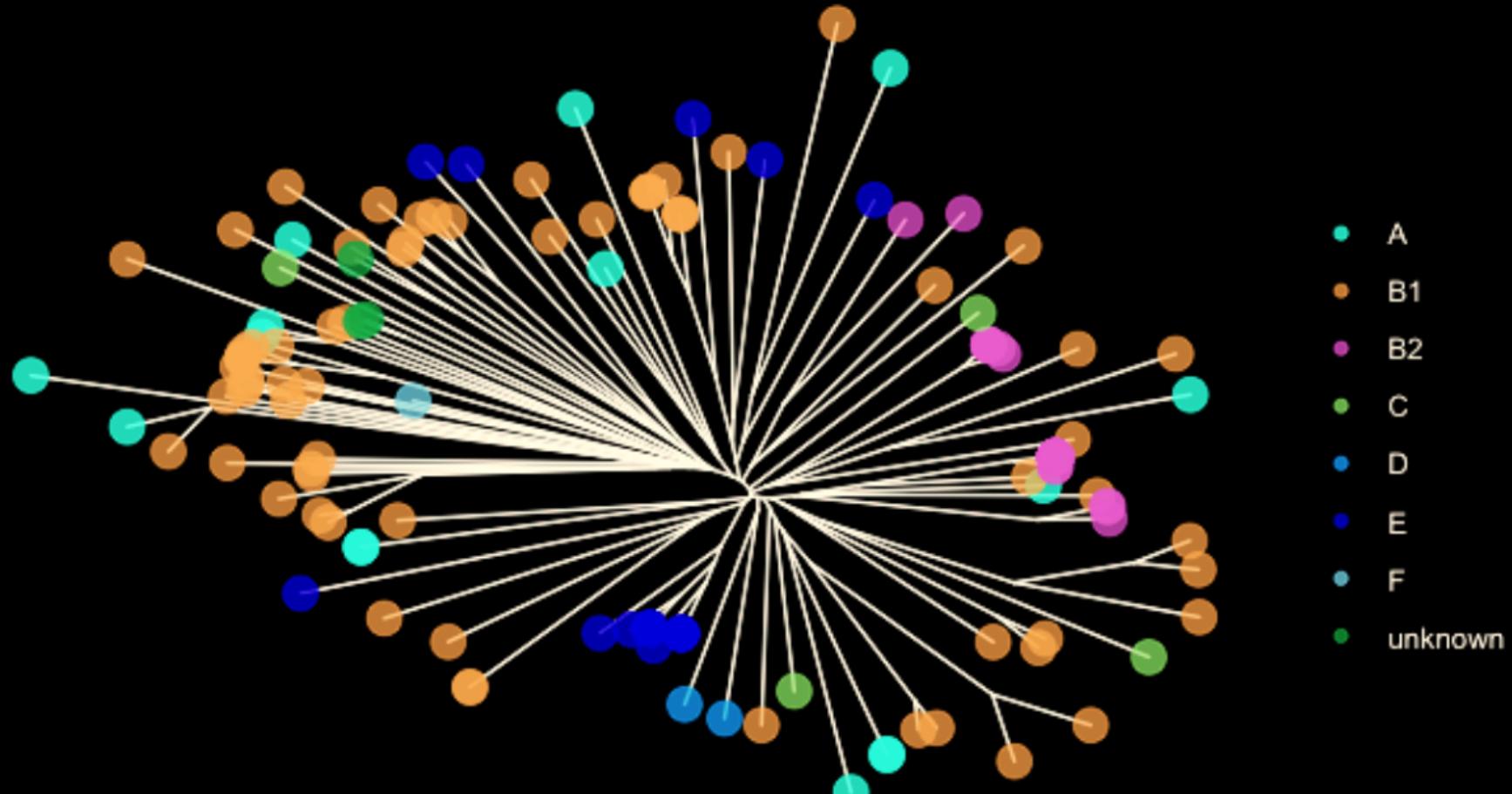
## Horizontally Acquired Partial Pan-genome of Inserted Elements

1. Check quality of genome assembly:
  - ✓ remove contigs < 800bp
2. Annotate whole genome (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

# Pan-genomee neighbor-joining tree



# Pan-mobilome neighbor-joining tree

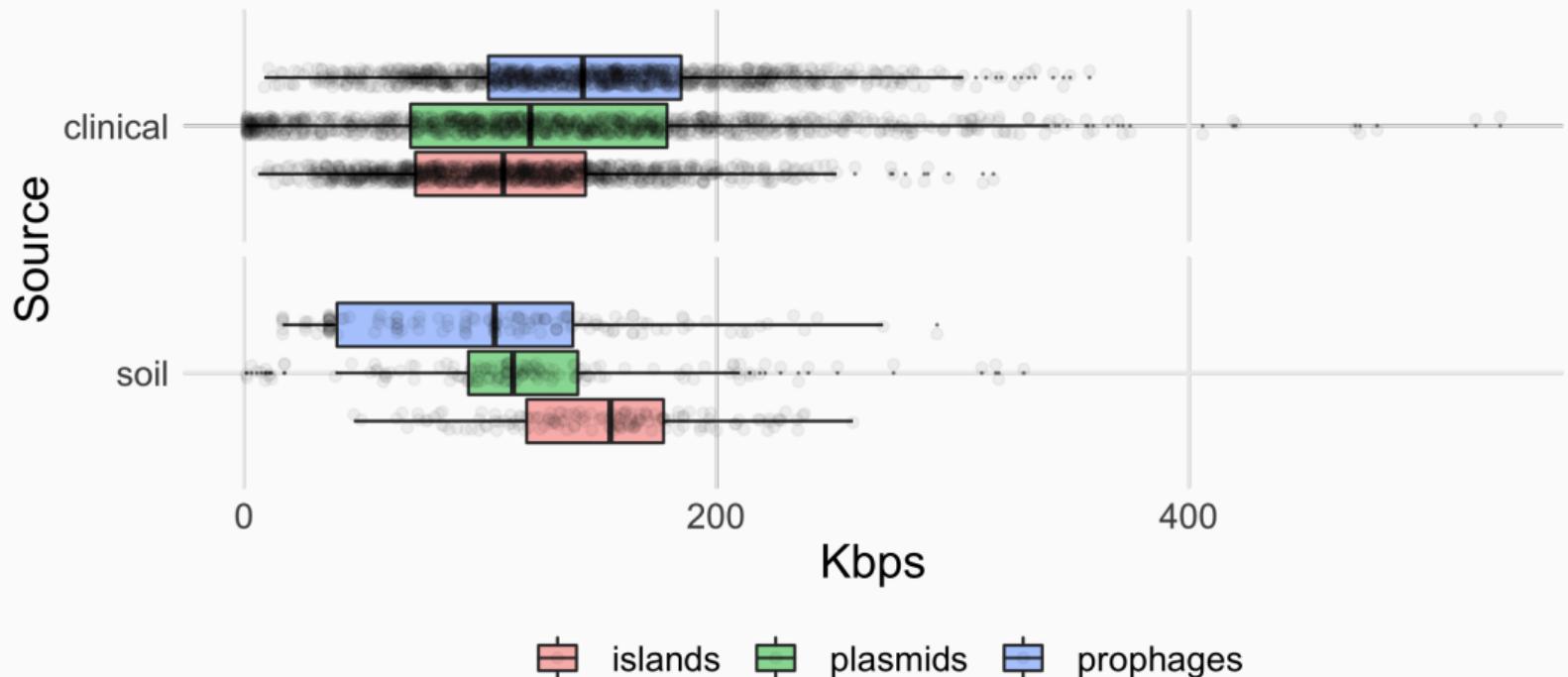


## Comparing the pan-mobilome

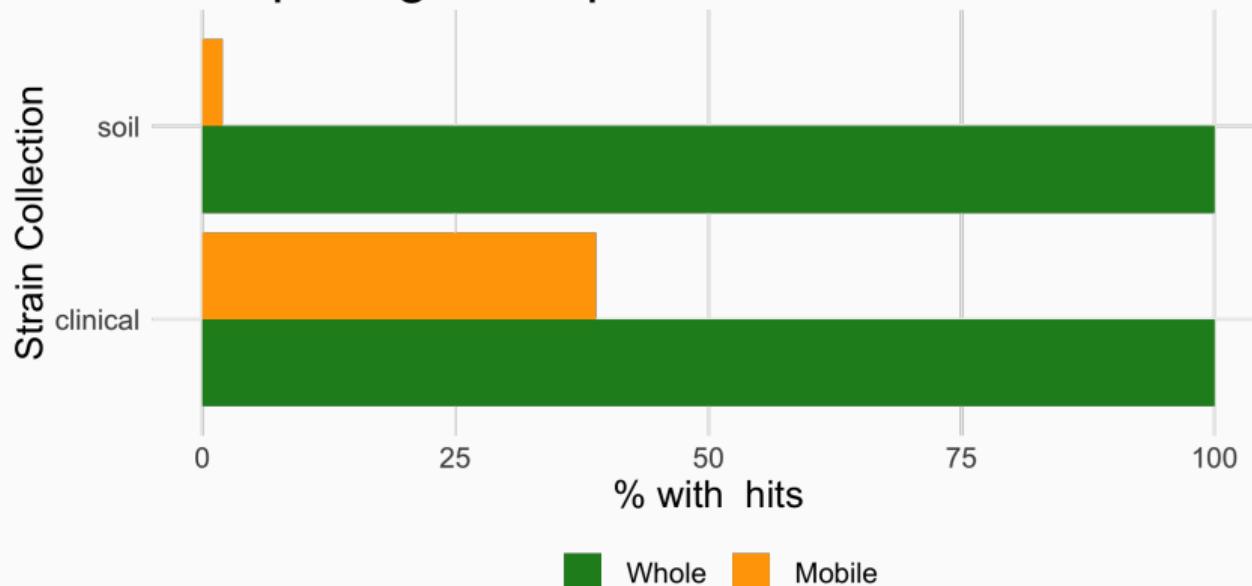
---

# Pan-mobilome size

Length of all mobile elements per genome



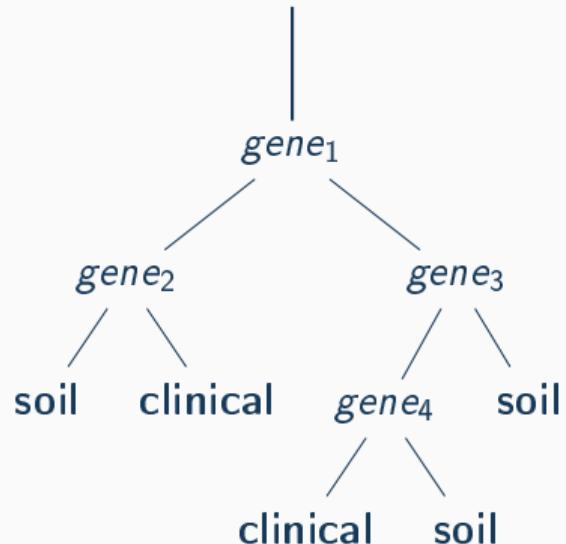
## Comparing AMR profiles between source



## Enrichment Analysis

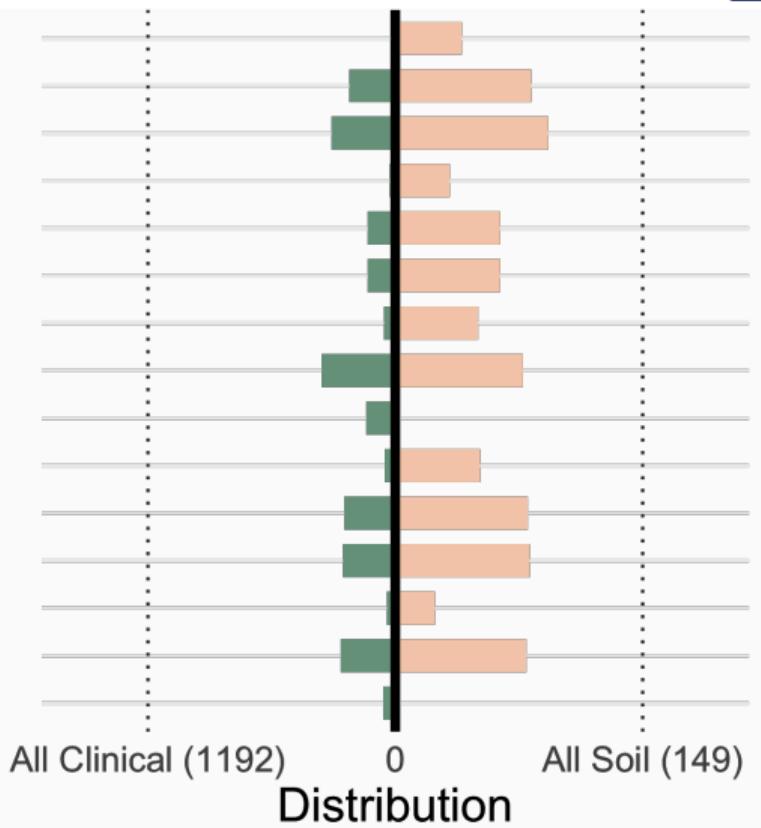
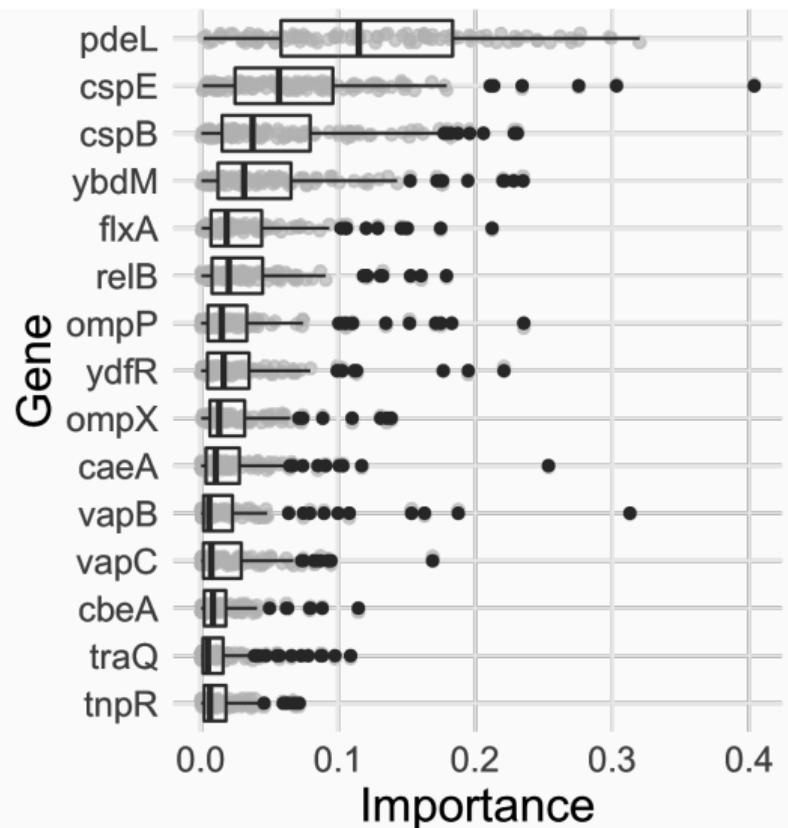
---

# Enrichment Analysis with Random Forests



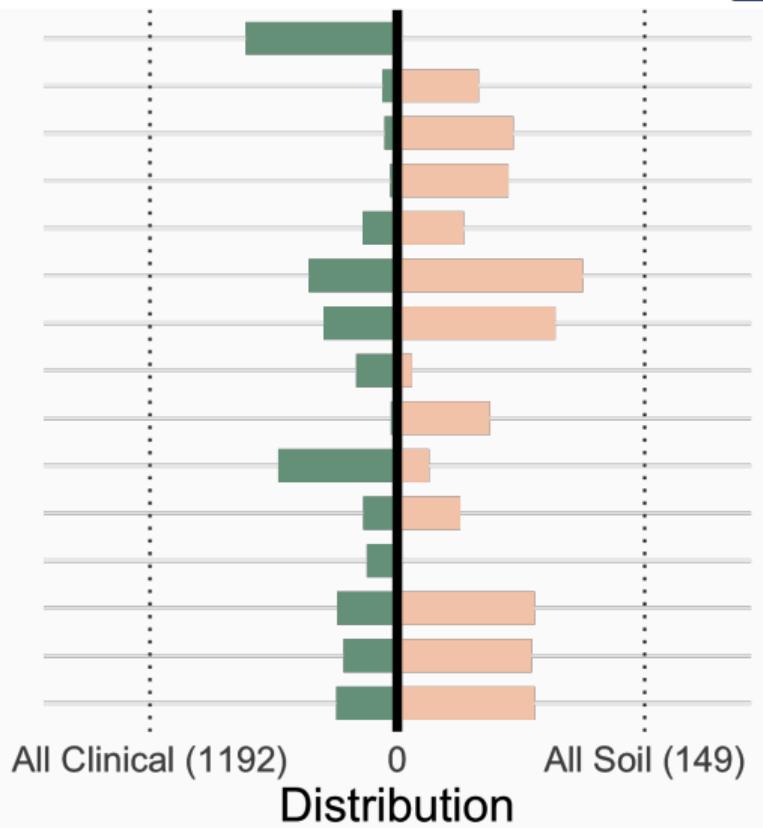
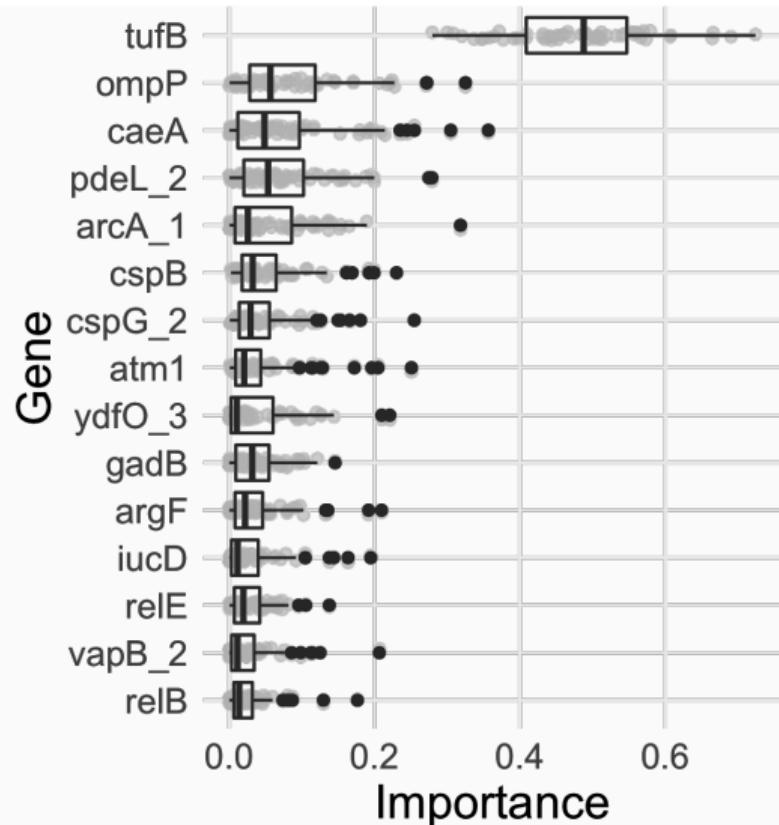
- ↪ Supervised machine learning
- ↪ Randomly generate decision trees
- ↪ Assess, repeat

# Important Genes in Pan-mobilome



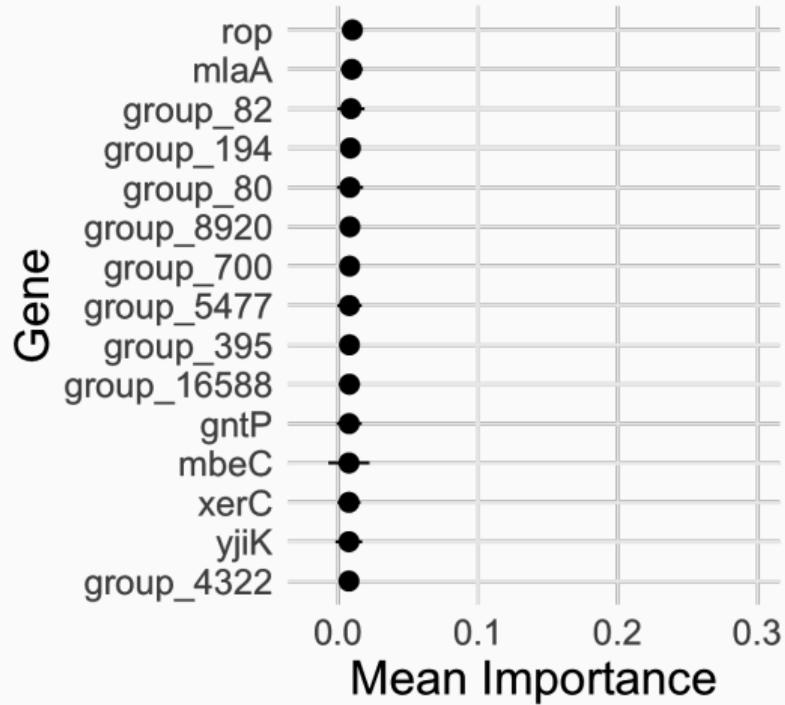


# Important Genes in Pan-genome

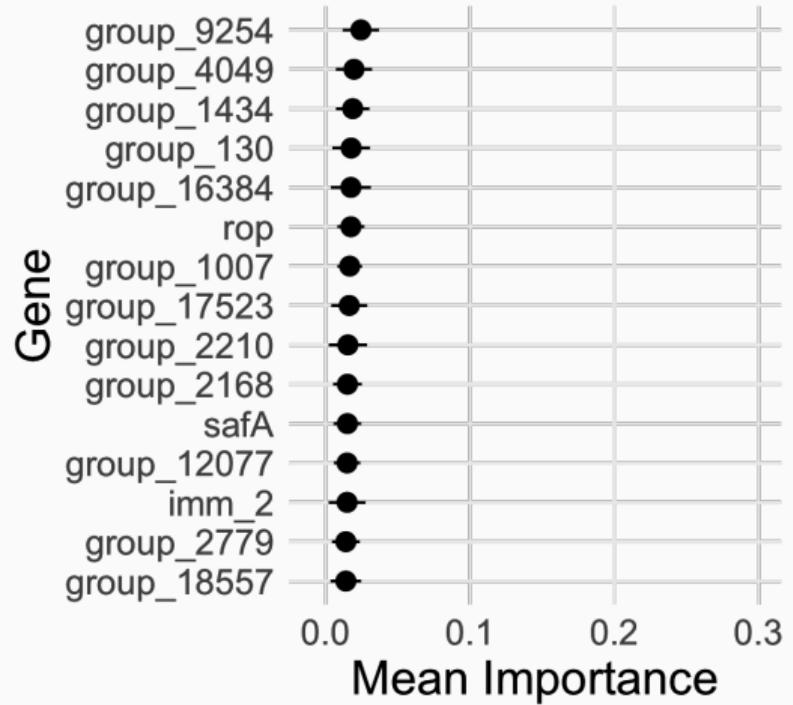


# Shuffled Classes

Pan-mobilome



Pan-genome





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

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

What we want to know:

- ❑ what inferences can we make about adaptation?
- ❑ which *types* of phages, plasmids, etc, are correlated by source?

# Acknowledgments



## NUIG Microbiology

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



## James Hutton Institute, Dundee

- Dr. Leighton Pritchard
- Dr. Ashleigh Holmes
- Information and Computational Sciences Group

# Questions?



- ⌚ [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

@nickwotters

n.waters4@nuigalway.ie