

Microsoc Seminar Series 2018-09-20:

Soil-persistent *E. coli* and Mobile elements

Nicholas Waters

September 20, 2018

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

Background

Sources



- o <https://github.com/tseemann/prokka>
- o <https://github.com/nickp60/riboSeed>
- o <https://sanger-pathogens.github.io/Roary/>
- o <https://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518>
- o <https://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518>
- o https://www.researchgate.net/publication/283707425_The_Phage-Inducible_Chromosomal_Islands_A_Family_of_Highly_Evolved_Molecular_Parasites
- o <https://www.sciencedirect.com/science/article/pii/S0043135416302226>
- o http://www.bx.psu.edu/~ross/workmng/TranspositionCh9_files/
- o <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00762>

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



- *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

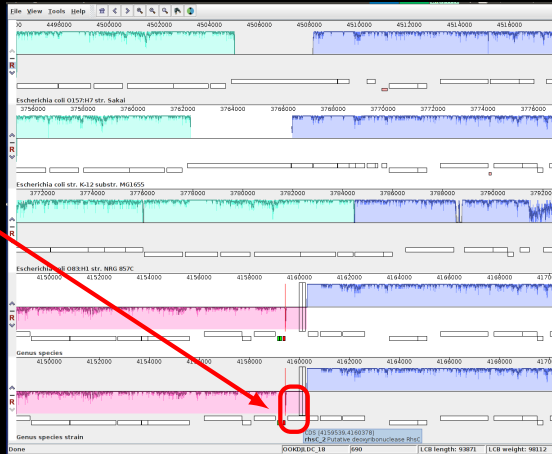
Assembly \Rightarrow Annotation \Rightarrow presense/absence matrix
riboSeed \Rightarrow Prokka \Rightarrow Roary

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

Annotation



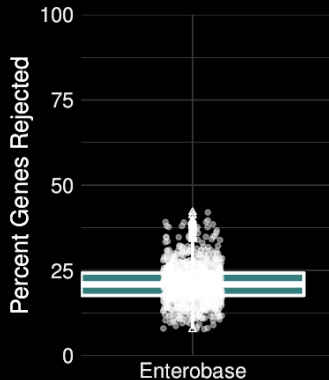
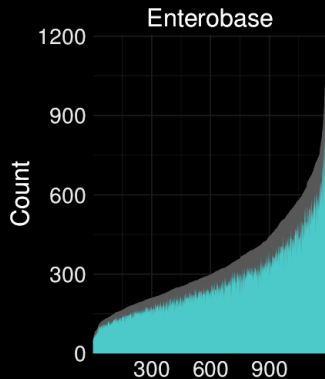
Partial



Annotation Correction with annofilt performance



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



Detecting differential presence/absence



- o No statistically significant gene associations with source
- o *E coli* diversity adds noise

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

BoE Calculations for Doubling Time



High estimate:

$$0.013865 * 60 * 24 * 365 * 10 \approx 72\text{k generations}$$

BoE Calculations for Doubling Time



Medium estimate:

$$(5.9\% / 2) * 365 * 10 \approx 10\text{k generations}$$

*(assuming generation time roughly equals half of turnover rate)

Other estimates



No viable coliforms after:

- 22 days
- 32 days
- 8 weeks
- 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

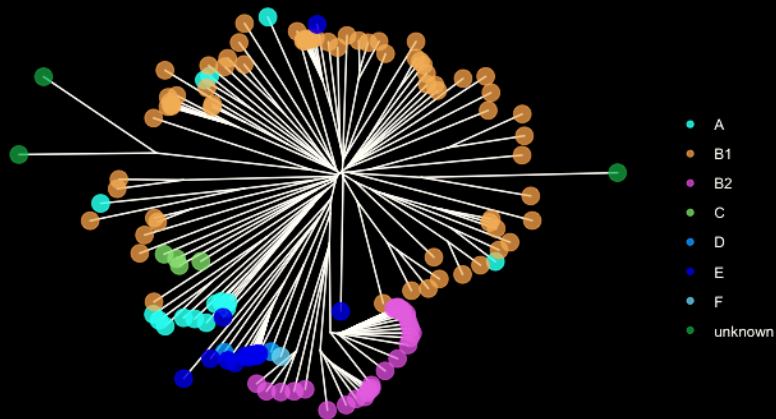
Mobile Genetic Elements

Overview



	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	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>

Genomic NJ on Pangenome

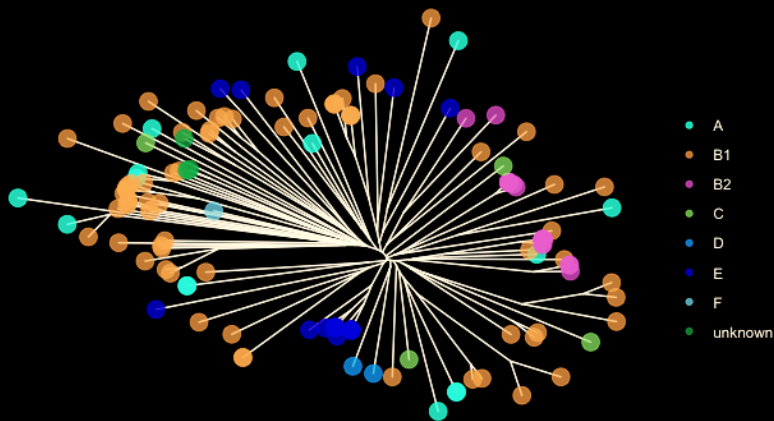


<https://github.com/nickp60/happie/>



Horizontally Aquired Partial Pangenome of Inserted Elements

Horizontally Acquired Partial Pangenome of Inserted Elements



In Closing

Summary



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

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
- whether genomic islands reflect phylogeny
- which regions are interrupted by IS, Tns, etc

Summary



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

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
- whether genomic islands reflect phylogeny
- which regions are interrupted by IS, Tns, etc

Questions?