Genomic and phenotypic comparison of *Eschericia coli* isolates from host and environmental sources

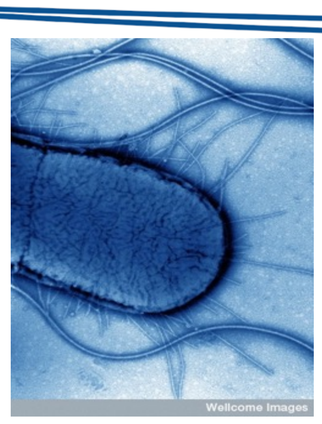
Jake Ireland, Guillaume Méric, Leonardos Mageiros, Samuel K. Sheppard

Medical Microbiology and Infectious Disease, College of Medicine, Swansea University, UK

Contact: 641415@swansea.ac.uk ; www.sheppardlab.com

Escherichia coli Facts

- A large percentage of *E. coli* has asymptomatic carriage in endotherms.
- *E. coli* can also be pathogenic and is responsible for 2 million deaths a year (28 per 100,000)
- *E. coli* are the predominant aerobic microbe in the human gut.
- *E. coli* is a facultative organisms that can survive in both aerobic and hypoxic environments.
- *E. coli* is expect to have two phenotypically distinct cell types with different molecular controls adapted to host and non-host environments.
- It is a necessity to understand the adaptations of commensal *E. coli* in order to understand how they may become pathogenic.



Electron micrograph of E.coli (1).

This project compares both host and non-host *E. coli*

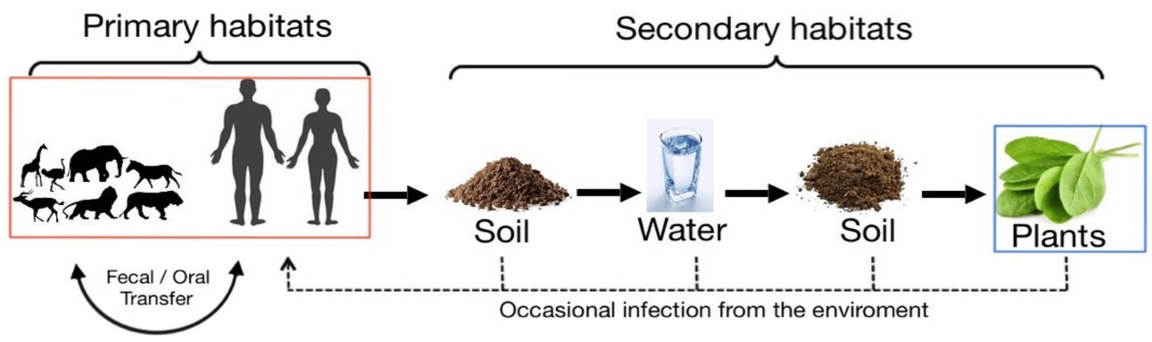


Figure 1. Ecological prevalence of *E. coli*. *E. coli* are found predominantly in the gut of endotherms where the temperature is uniformed at 37C. E.coli can be spread into secondary habitats through fecal transfer and can colonise on plants.

Host and non-host *E. coli* are found in all linages

- The phylogenetic tree of the *E. coli* strains shows that the host and Non-host genomes are inter-dispersed into different genetic linages.
- *E. coli* populations are largely clonal but there are combinations of enzymes that allow the strains to be clustered into lineages (3).
- *E. coli* has 6 lineages, 5 of which are represented here.
- The signature of ecological differences between the host and non-host environments, are not likely to be very profound.

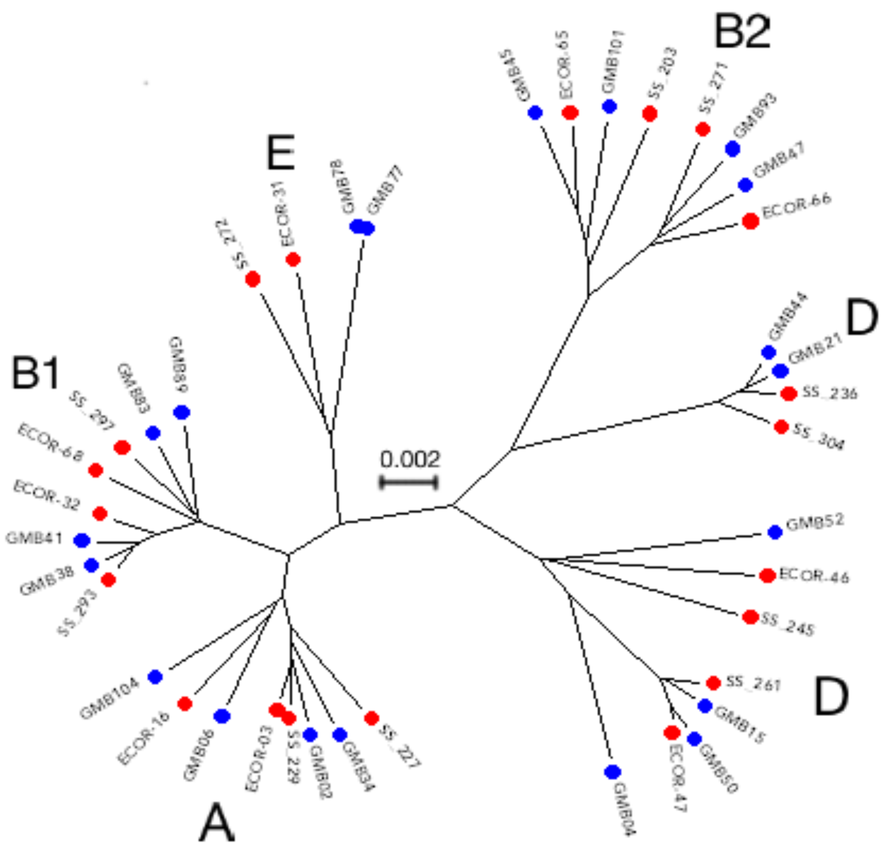


Figure 2. MEGA6 tree Phylogenetic tree of 40 E.coli. Maximum likelihood tree of 20 genomes belonging to *E. coli* host sources shown in red and 20 non-host sources shown in blue. The scale bar indicates the estimated number of substitutions per gene site. The tree is separated into 5 groups (A, B1, B2, D and E) that represent the main linages that the species is separated into.

E. coli growth shows similar growth responses in different environments

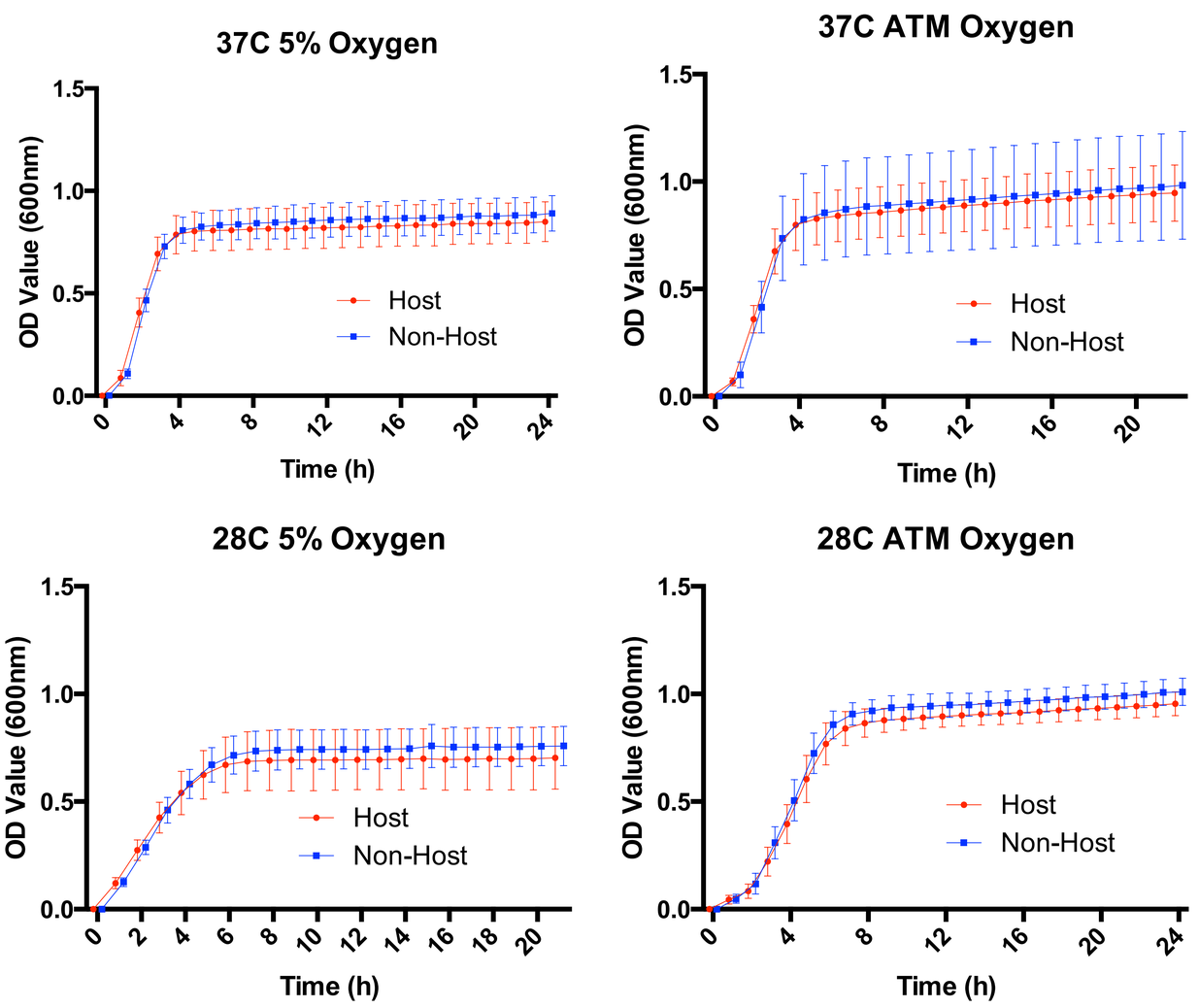


Figure 4. Growth comparisons between host and non-host. The 40 commensal strains are grown in four different environments. 1) At 37C and 25C at 5% oxygen 2) And at 37C and 28C in atmospheric conditions. It is observed that the difference between the two temperatures reaching Log phase is significant ($P<0.05$) but the difference between atmospheric and hypoxic is not significant.

- ***E. coli* has similar growth response in both hypoxic and atmospheric conditions.**
- ***E. coli* shows a preference for endotherm growth conditions (37C 5% Oxygen).**
- ***E. coli* Shows a reduced total volume in hypoxic conditions compared to atmospheric.**
- ***E. coli* being a facultative organism its not surprising that the growth rate is similar between hypoxic and atmospheric conditions.**

Host and non-host *E. coli* exhibit similar core and accessory genome sizes.

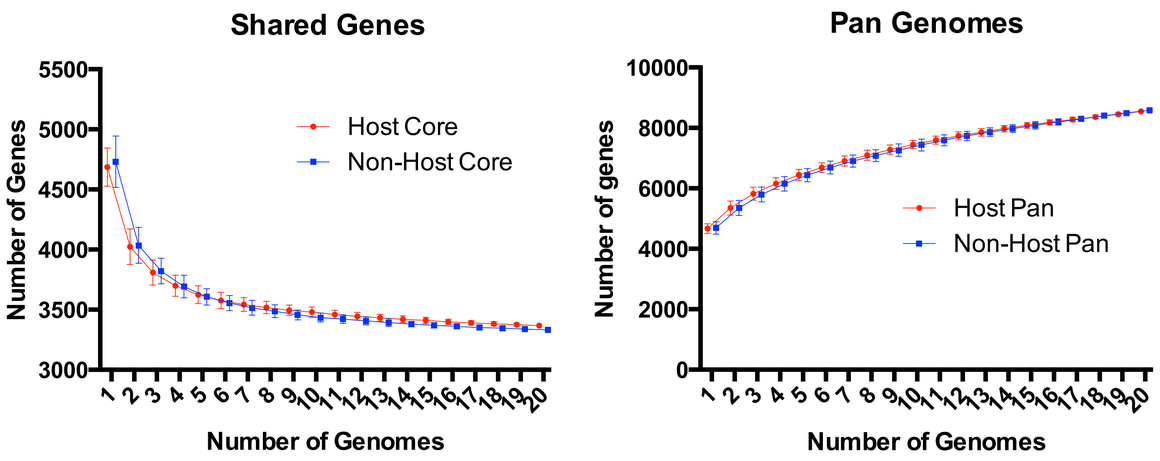


Figure 3. Accumulation curves. The number of shared genes represents the difference in gene volume between the host and non-host strains. The Pan genomes show how the number of genes increases as you include more strains into the reference pan genome list. Comparisons are made based on a matrices of gene presence/absence derived from the reference pan genome. The method of genome sampling was randomised and sampled 100 times to obtain the average number of genes for each sample (4).

- **The average size of the pan genome represents 47.3% of the super genome list.**
- **The average core genome represents 69.7% of the average whole genome size. This shows the core genome is largely conserved between strains.**

	Pan genome	Core Genome	Accessory Genome
All *E. coli*	4688	3266	1422
Host	4680	3368	1414
Non-Host	4696	3332	1430

Table 1. Genome size comparison The average number of genes in the core and pan genomes is different between host and non-host genomes. The accessory genome represents the number of genes not included in the core genome.

CONCLUSION

- It can be estimated that because the difference between the host and non-host strains is small, then the evolutionary adaptations are a fairly recent occurrence.
- The total growth of *E. coli* is lowest in 5% oxygen at 28C. This may show that the molecular systems involved in hypoxic growth are temperature dependent and haven't developed to be as efficient at temperatures lower than an endotherm environment.
- There may be a few genes that help provide fitness in the separate environments but they may not be linked to oxygen concentration or temperature. On the other hand the high level of exchange between hosts may mean that the genes to provide fitness in secondary habitats hasn't been isolated in non-host strains.
- There is no significant difference between core and pan genome sizes of the host and non-host strains. It is more likely that the commensal diversity is mutation related and dependent on horizontal transfer than it is on genome size alterations.

Ref 1. Gregory & Marshall (2003) Wellcome images sourced from: http://wellcomeimages.org.
Ref 2 Ahmed *et al* (2008) Nat Rev Microbiol 6(5): 10.1038/nrmicro1889
Ref 3 Tenaillion *et al* (2010) Nat Rev Microbiol 8(3): 10.1038/nrmicro2298.
Ref 4 Meric *et al* (2014) PlosOne, 9(3): e92798.