Comparative Genomics Shows Site-Specificity of *Escherichia coli* in the Lower Gut of Humans

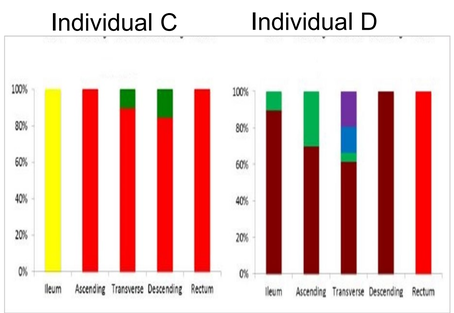
Rasel Barua [1, 2, 3], David M Gordon [2], Paul Pavli [1,3], Claire L O'Brien [1,3,4]

1. ANU Medical School, Australian National University, Canberra, ACT
2. Population Biology of Microorganisms Laboratory, Research School of Biology, Australian National University, Canberra, ACT
3. Inflammatory Bowel Disease Research Laboratory, Canberra Hospital, ACT
4. Faculty of Science and Technology, University of Canberra, Canberra, ACT

Introduction

Escherichia coli (*E. coli*) is a facultative anerobic bacterium of the lower gastrointestinal tract of vertebrates and humans (Denamur *et a*l., 2020). In some individuals, *E. coli* strains appear to be site-specific, for example, residing in the ileum but not the rectum [Figure- 1, individual C and D] (Gordon *et al*., 2015), yet site-specificity of commensal *E. coli* in the human gut has not been demonstrated. Site-specificity can be determined by looking at any differences in genomic characteristics of clone-pairs, i.e., two isolates collected from different gut regions of the same individual with similar multiple-locus variable-number of tandem repeat analysis [MLVA] gel electrophoresis profiles.

Figure- 1: Examples of individuals harboring strains in different gut locations.



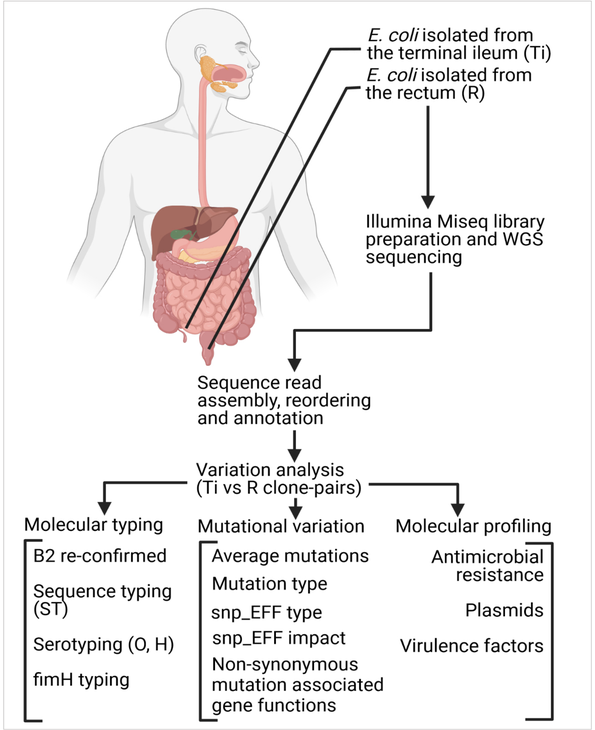
Aim

To determine whether or not clone-pairs of *E. coli* collected from different gut locations of the same individual contain different genomic elements, which may explain site-specificity in the gut.

Materials and methods

E. coli was cultured from the terminal ileum (Ti) and rectum (R) of 34 individuals. MLVA typing was used to identify clone-pairs at each gut site. Only B2 phylogroup strains of *E. coli* were used to limit genetic differences associated with phylogroup rather than niche specificity. Whole genome sequencing was performed on an Illumina Miseq platform. Bioinformatic tools available in PATRIC were used to identify mutations with a quality score of ≥90%. The Center for Genomic Epidemiology tools (genomicepidemiology.org) were used to identify sequence types (ST), serotypes, fimbriae, plasmids, virulence factors and antimicrobial resistance genes (Figure- 2).

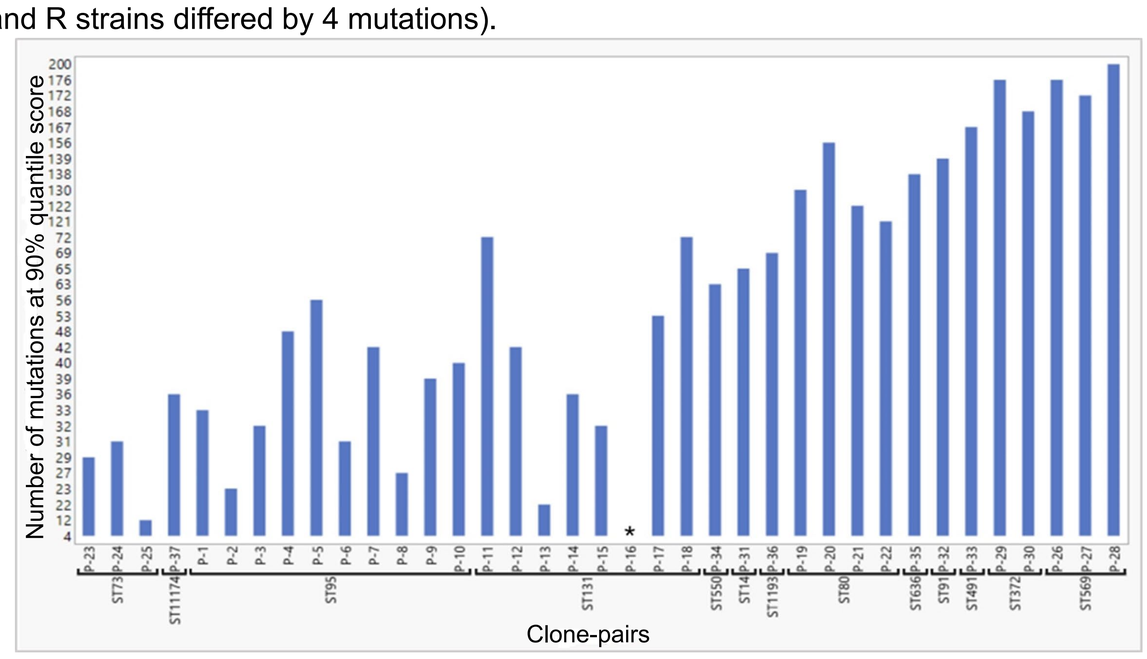
Figure- 2: Experimental work flow. Image adapted from "Digestive system (male)", by BioRender (2021). https://app.biorender.com/biorender-template.



Results

a) The sequence types, serotypes and fimbriae types profiles were similar in 97% clone-pairs; antimicrobial resistance gene profiles (range: 1-12) were similar in 97% clone-pairs; plasmid profiles (range: 1-9) were similar in 95% clone-pairs and virulence factors (range: 12-26) were similar in 100% clone-pairs.

Figure- 3: Mutational differences observed in clone-pairs (*- in the clone-pair P-16, the Ti and R strains differed by 4 mutations).



Results

b) On average each clone-pair varied by 78 mutations [range: 4-200; (Figure- 3)].

c) Clone-pairs belonging to human-associated STs (e.g., ST73, ST95, ST131) had significantly fewer mutational differences than non-human-associated STs (e.g., ST80, ST372, ST569) [Figure- 4]. Figure- 4 shows the average mutational variations across all STs (STs including single clone-pairs were excluded) was 73, with human-associated strains having less mutations than expected and non-human-associated STs having a higher number of mutations than expected.

Figure- 4: ST wise mutational differences (red tip: average mutational variations between clone-pairs for each ST; UDL: upper decision limit; LDL: lower decision limit).

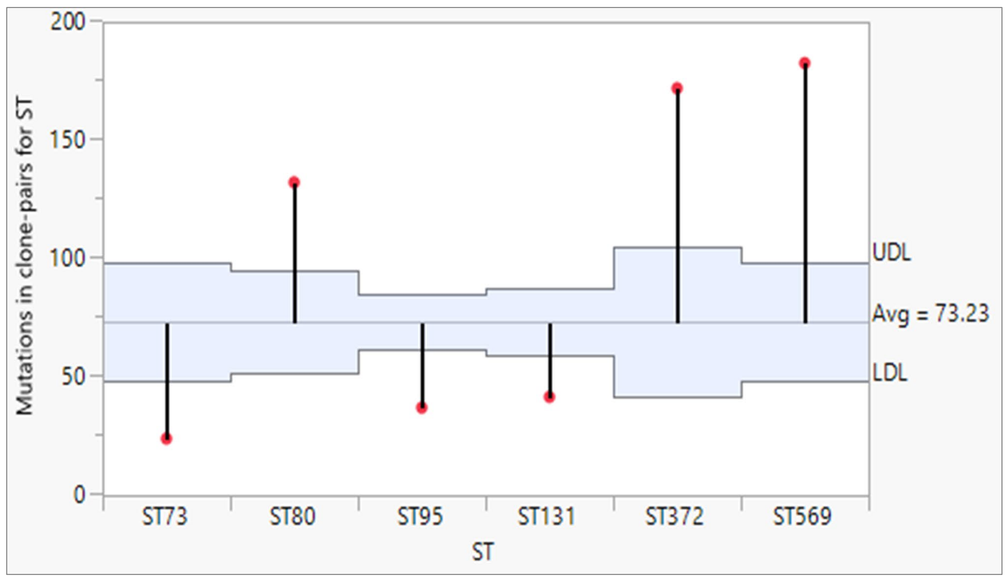
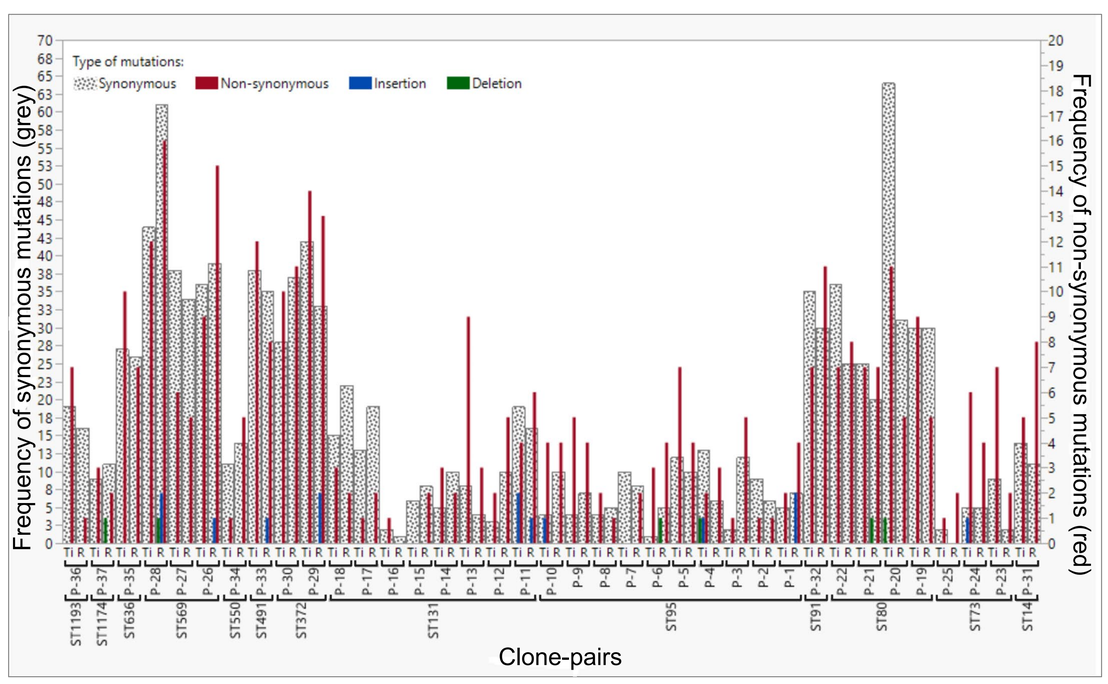


Figure- 5: Frequency of mutational types in clone-pairs.



d) Approximately 76% of the mutations detected between clone-pairs were synonymous, 23% were non-synonymous, and the remaining were indels (insertions and deletions) [Figure- 5]. However, neither Ti or R strains were more likely to harbor more synonymous or non-synonymous mutations than the other ($p>0.05$), and no non-synonymous mutations were common to all Ti or R strains.

Mutational differences may be an outcome of within-host genomic evolution of *E. coli* to facilitate adaptation to a specific niche, as physiological variations exist between the Ti and R. However, given the lack of any gene signature in either Ti or R strains, most of the non-synonymous mutations were likely accumulated neutrally.

Conclusion and significance

Clone-pairs of *E. coli* isolated from different gut regions showed genomic variation that may account for site-specificity in the gut. Mutational differences in clone-pairs are not as common in human, as opposed to non-human-associated STs, suggesting that strains belonging to human-associated STs are better able to survive in multiple environments.

References

DENAMUR, E., CLERMONT, O., BONACORSI, S. & GORDON, D. 2021. The population genetics of pathogenic Escherichia coli. *Nature Reviews Microbiology*.

GORDON, D.M., O'BRIEN, C.L. & PAVLI, P. 2015. Escherichia coli diversity in the lower intestinal tract of humans. *Environmental Microbiology Reports*.

Acknowledgement:
Participating patients; Staff of Gastroenterology and Hepatology Unit, Canberra Hospital; ANU College of Health and Medicine

Contact:
rasel.barua@anu.edu.au