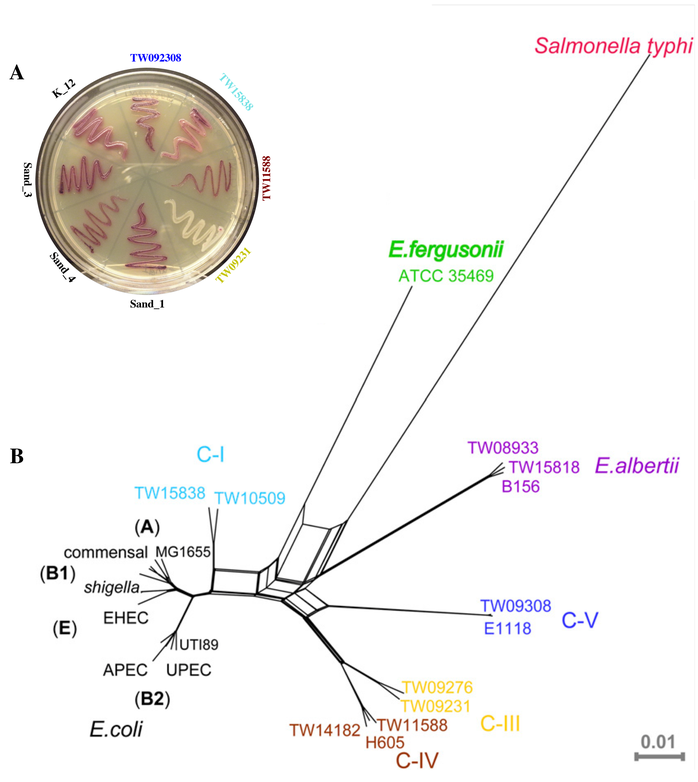
The Emergence of Naturalized *E. coli* and its Importance for Fecal Contamination Testing

Tomas Leon[1], Despina Tsementzi[1], Chengwei Luo[2], Natasha DeLeon-Rodriguez[3] and Konstantinos Konstantinidis[1,2,3]*.

[1]School of Civil and Environmental Engineering, [2]Center for Bioinformatics and Computational Biology, [3]School of Biology, Georgia Institute of Technology, Atlanta, GA.

Background

❖ The primary habitat of *Escherichia coli* is the gastrointestinal flora of humans and other warm-blooded animals (commensal organisms) and it is believed that *E. coli* populations diminish quickly outside their host.
❖ *E. coli* plays a key role in the assessment of water quality as a fecal indicator bacterium.
❖ Recent surveys have convincingly identified relatively abundant *E. coli* populations that persist over time in freshwater lakes and beaches or tropical soils (1, 2).
❖ These "atypical" environmental *E. coli* strains comprise five novel clades of the Escherichia genus, closely affiliated to typical *E. coli* (2) (Fig B).
❖ Importantly, the latter isolates were indistinguishable from typical *E. coli* based on traditional coliform test (based on the enzymatic activity of b-D glucuronidase-gene *uidA*) (Fig A).

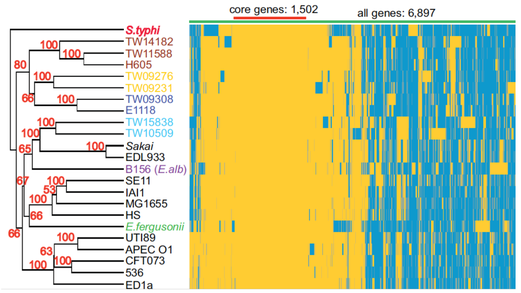


Objectives

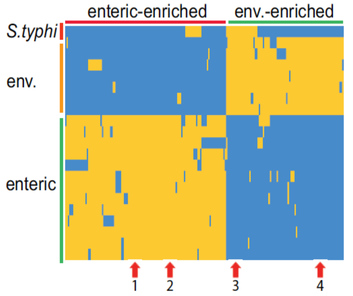
❖ Develop a new assay that targets specifically FIB (Fecal Indicator Bacteria) and is not confounded by natural environmentally adapted populations of close relatives of FIB.
❖ Test the assay in environmental samples (Lake Lanier is our case study) and compare it with the coliform test.
❖ Identify natural *E. coli* populations from environmental samples (Lake Lanier is our case study) to assess the robustness of the method in environmental samples.

Comparative genomics

❖ Gene content comparisons between 25 *E. coli* genomes with representatives from all 5 environmental clades revealed that these atypical "naturalized" *E. coli* lack several functions that are important for fitness in the human gut and are unlikely to present a health or environmental hazard (3).
❖ Out of the 5,396 assesory genes of the *E. coli* pangenome, 84 and 120 genes were found to be enriched in the environmental and the enteric groups respectively (left figure)
❖ Some of those enriched genes were reflecting the environmental adaptations and lifestyles of these organisms (right figure).
❖ These 84 enteric related genes could therefore be used to develop a more robust strategy to count the commensal *E. coli* cells in environmental samples and discriminate them from their environmental counterparts.



Heatmap of gene presence (yellow) and absence (blue) in 20 selected genomes, using all non-redundant genes that were found in at least two of the genomes as reference.



1, acetylglucosamine transporter;
2, fructose transporter;
3, diol utilization operon;
4, lysozyme production.

Distinguishing environmental from enteric

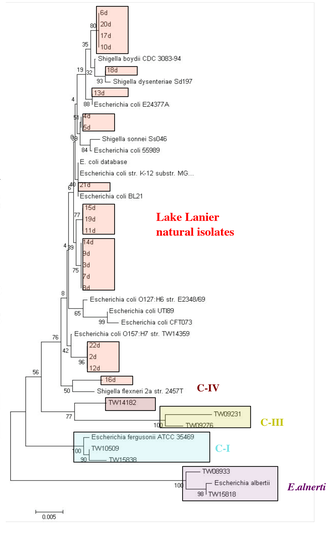
❖ Enteric and environmental specific (present in all strains) genes were found only when we were grouping the C-I clade with the enteric ones.
❖ 46 environmental and 41 commensal specific genes were identified.
❖ Those 41 genes were the ones chosen PCR based identification of the commensal strains.
❖ In combination with some C-I specific genes, we could accurately distinguish all environmental versus enteric strains.

Isolation of natural *E. coli* strains

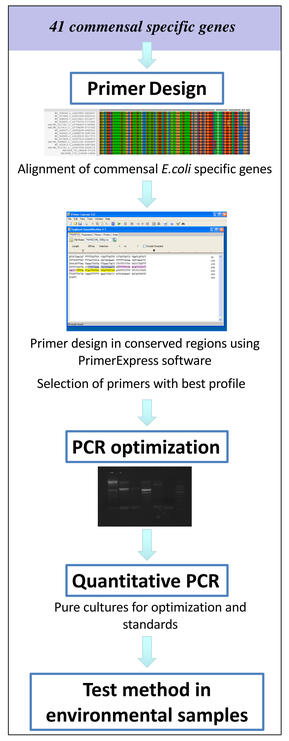
E. coli isolates were obtained using the coliform test and their species assignment was confirmed by 16S rRNA gene sequencing.

Multilocus sequence typing (MLST) was used to more precisely assign isolates to the known *E. coli* clades (environmental vs. commensal).

Most of the isolates were grouping with the commensal strains, and were therefore be used as positive control natural strains for our PCR based identification assay.

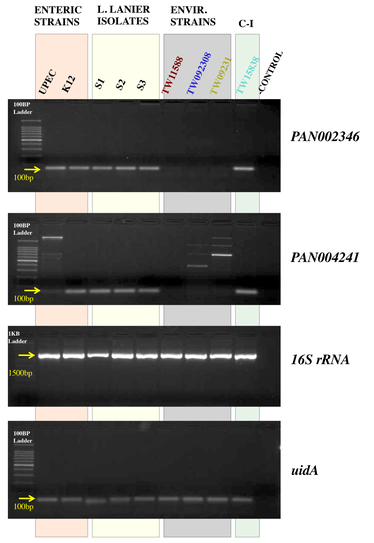


Developing the molecular assay



All 41 genes were analyzed for conserved regions and primer design, and two of them were tested in PCR assays (PAN002346 annotated as predicted permease and PAN004241 which is predicted to be a transcriptional regulator)

Our primers were more specific for PAN002346, and so they will be used in the quantitative PCR assay, which will be the method of choice for the quantification of the enteric *E. coli* cells



Conclusions

❖ Comparative genomic analysis of 25 *E. coli* genomes revealed that the environmental "atypical" *E. coli* lack several important gene that characterize the enteric strains, thus are unlikely to present health or environmental hazard.

❖ 41 genes can distinguish the enteric from the environmental strains and can be used for the development of more robust quantitative assays for the identification of coliforms.

❖ *PAN002346* is a predicted permease found only in the enteric strains (and also in the C-I clade strains) which makes it a good target for PCR characterization of the enteric strains.

❖ Collectively, our results advance current understanding of the genetic and ecologic diversity of *E. coli* and provide means toward the development a new, improved molecular assay for coliform.

References

1. S. Ishii, W. B. Ksoll, R. E. Hicks, M. J. Sadowsky, Appl Environ Microbio 72, 612 (Jan, 2006).
2. S. T. Walk et al., Appl Environ Microbiol 75, 6534 (Oct, 2009).
3. C. Luo et al., PNAS 108, 16 (April 2011).