

PhD Project 2: Phylogeny-aware metrics for microbial community assembly driven by ecological and evolutionary principles

Supervisors:

University of Glasgow:

Dr Umer Zeeshan Ijaz

<http://userweb.eng.gla.ac.uk/umer.ijaz>

Professor Cindy Smith

<https://www.gla.ac.uk/schools/engineering/staff/cindysmith/>

Professor William Sloan

<https://www.gla.ac.uk/schools/engineering/staff/williamsloan/>

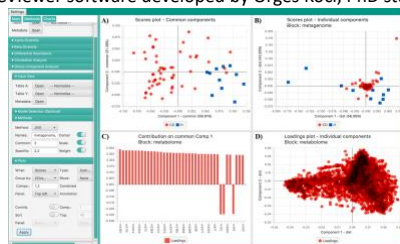
University of Sheffield:

Dr Isabel Douterelo Soler

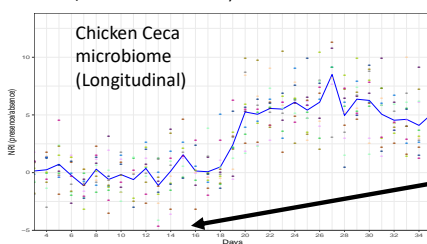
<https://www.sheffield.ac.uk/civil/staff/academic/isabel-soler>

Description: Microbial community surveys often involve alignment and generation of phylogenetic trees using Operational Taxonomic Units (OTUs) or alternatively, Single Nucleotide Variants (SNVs) as an OTU-free approach using different marker genes (16S/18S rRNA, ITS region etc.). These phylogenetic trees in conjunction with species abundances on a sample space are usually employed in distance metrics (such as Unifrac distances) to ascertain geometric sources of variation (e.g., PERMANOVA test) against extrinsic meta data. Recently, phylogenetic-aware alpha diversity measures have seen their utility in exploring stochastic and deterministic nature of microbial community assembly to delineate environmental pressures (e.g., NTI/NRI metrics). This is usually done by looking at how clustered/dispersed the phylogenetic tree is. Indeed, our recent work (Figure 1d) [1] has shown a switch from competitive to environmental drivers of microbial communities in longitudinal Chicken cecum profile creating a window of opportunity for human pathogens such as *Campylobacter* to appear.

a) CViewer software developed by Orges Koci, PhD student



c) Step change from competitive exclusion to environmental filtering (Ijaz et al. *Frontiers in Microbiology*, 2018. DOI: 10.3389/fmicb.2018.02452)



b) Phylogenetic alpha diversity measurements (NRI/NTI) implemented in CViewer (metagenomics)

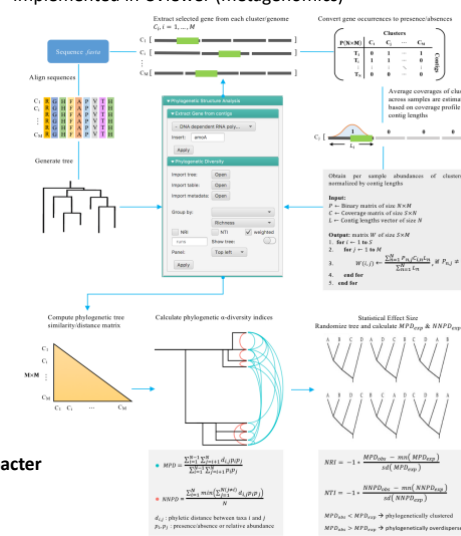


Figure 1: a) CViewer, a software for metagenomics exploration developed by Orges Koci, a PhD student that also integrated other meta*omics technologies; b) Phylogenetic assembly metrics currently available in Cviewer; and c) Recent paper by Dr Ijaz highlighting a step change from stochastic to deterministic community assemblage in chicken ceca microbiome possibly resulting in pathogen such as *Campylobacter* to appear

Other recent methodological developments include phylogenetic beta diversity variants such as β -NTI/ β -NRI [2] and various statistical moments on the phylogenetic trees [3]. In view of these recent developments, the main aims of the PhD project are:

- a) to consolidate the existing literature on phylogeny-aware metrics for microbial community analyses (borrowed from the latest in numerical ecology);
- b) to further develop information theoretic approaches looking at community assembly from a phylogenetic point of view at different granularity (from species to genera to families to taxa up the hierarchy) and by doing so assessing anomalies in the commonly used reference taxonomies;
- c) to incorporate models of molecular evolution in phylogeny aware metrics;
- d) to develop approaches for concordance of multiple phylogenetic trees all derived from different marker genes (or primers pairs), but for the same sample space;
- e) and to develop mathematical/statistical models on phylogeny that give an account of microbial community resilience to external perturbations by presence/absence of specific clades; and
- f) to apply these methods to whole-genome shotgun sequencing datasets including datasets from drinking water distribution systems.

The project team also has a vast experience in developing mathematical and statistical models to explain community assembly in microbial communities, for instance, exploration of neutral community assemblage fitting Neutral Community model for prokaryotes to the distribution of microbial taxa (Professor Sloan) [4], and recent work involving Dr Ijaz on fitting the Unified Neutral Theory of Biodiversity with Hierarchical Dirichlet Process (NMGS package [5]), as well as implementation of NRI/NTI in CViewer software (Figure 1b) for divulging stochastic nature of given gene in a metagenomics dataset. The prospective student, ideally someone with a computational background: will become part of Environmental'Omic lab within the Water & Environment group (School of Engineering); will be given access to high-performance computing facility maintained at Dr Ijaz's lab; and will be provided numerous datasets from existing and past microbial community studies to test their methods. Further, programming experience in R is required as the secondary aim of the project is to port the developed methods to microbiomeSeq package (http://userweb.eng.gla.ac.uk/umer.ijaz/projects/microbiomeSeq_Tutorial.html) that both Dr Ijaz and Professor Sloan are contributing to. Dr Douterelo (Civil Engineering, University of Sheffield) will contribute with data from real drinking water systems (physico-chemical and microbiological), analysis of data and discussion of results regarding implications for the management of drinking water systems.

References:

- [1] U. Z. Ijaz et al. Comprehensive longitudinal microbiome analysis of chicken cecum reveals a shift from competitive to environmental drivers and a window of opportunity for *Campylobacter*. *Frontiers in Microbiology*, 9:2452, 2018. DOI: 10.3389/fmicb.2018.02452

[2] J. C. Stergen et al. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME*, 6:1653-1664, 2011

[3] C. Tsirogiannis and B. Sandel. PhyloMeasures: a package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography*, 39(7):709-714, 2016.

[4] W.T. Sloan et al. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 8: 732–740, 2006.

[5] K. Harris et al. Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process. *Proceedings of the IEEE*, 105(3):516-529, 2017.