Development of microbial population dynamics models from omics data

David Scott

Computational Methods in Ecology and Evolution (MSc)

Imperial College London

Project Proposal

Supervisor: Dr. Alberto Pascual–Garcia

Affiliated Institute: ETH Zurich

Email: alberto.pascual@env.ethz.ch

Submitted To:
Dr. Samraat Pawar
Course Director

Submitted By:
David Scott
01602669

Keywords: Microbial Communities, Metabolics, Modelling, Genomics, Population Dynamics, ODE.

1 Introduction

The role played by microbial communities has been increasigly recognised and has attracted much attention across disciplines from soil ecology (Torsvik & Øvreås 2002) to the human microbiome (Ley et al. 2008) and biotechnological processes (Rabaey et al. 2004). These roles or ecological functions are not the product of, or characteristics of, an individual species or process, but of the community as a whole.

Microbes alter the biochemistry of their media by metabolising a resource and outputting a by product, which in many cases can too be a resource. Thus, the consequence of the co-occurance of species under constant environmental conditons is not limited to competition (two species sharing same reseource). Cooperation or metabolic complementarity (metabolites of one species consumed by another) enables communties to shape their own niche, promoting community stability (Freilich et al. 2011). Therefore the combined metabolisms of individual member species create the community metabolome which determines the ecological function ((Arrigo 2004); (Goodman & Gordon 2010); (Noecker et al. 2016)).

Hence, to understand the organisational principles that determine the behaviour of communities a mechanistic understanding of interspecies interactions is required (Henry et al. 2016). However, little is known about the processes that determine the assembly and function of these microbial communities. It is difficult to investigate functions due to the diversity of species in natural communtiies, and isolating them to an axenic culture. A solution to this problem is to identify combinations of species abundances within stable communities compatable with the observation of the function of interest (Figure 1).

The aim of this project will be to explore different strategies to develop population dynamics models of microbial communties from empirical data. Through combining 16S rRNA, metagenomics and metabolomics data, we aim to infer ecological interactions and trophic strategies, from the analysis of both the abundances of species and their genetic repertoires. There are two data sets available to the study. The first dataset consists of assemblage experiments of marine particles data on pure polymer substrates, spanning 200h (12 time points), and includes both 16S rRNA and metagenomics data (Prof. Otto Cordero lab. (Datta et al. 2016)). The second dataset contains 700 16S rRNA samples of beech tree-holes, a natural aquatic habitat, and of a second subset of 250 samples of the same communities grown in the lab for seven days (Prof. Thomas Bell lab). For this new dataset, 16S rRNA and functioning of the communities was measured.
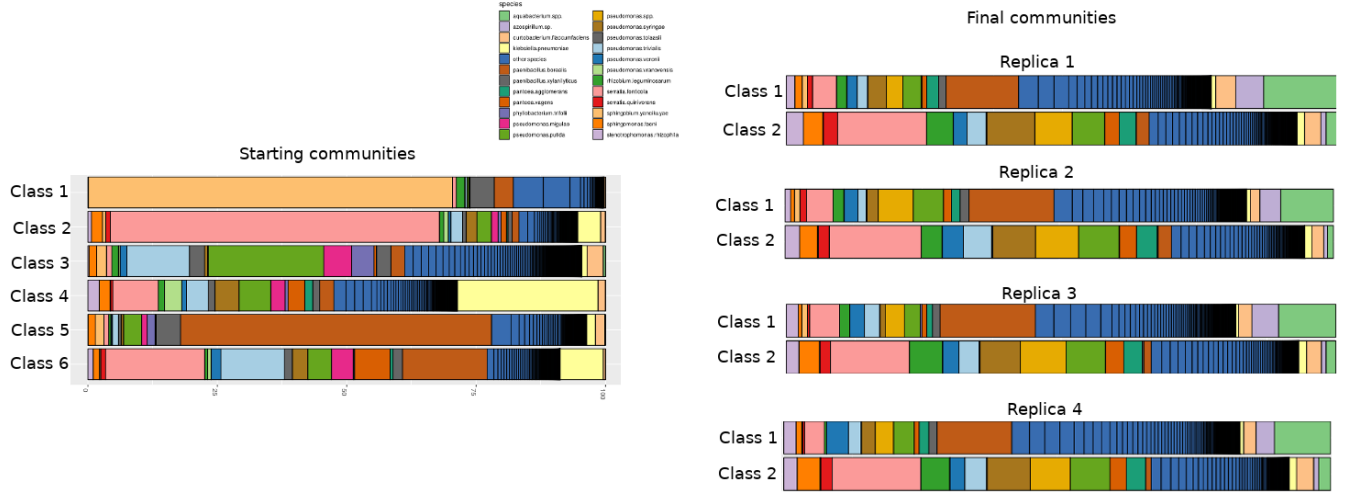
Figure 1: The 700 communties sampled from beech tree holes were classified into six classes (left panel). The size of each bar represents the relative abundance of OTU's. These are groups of species with 97% similarity. A subset of 250 microbial communities comprising of all six classes were then regrown under laboratory conditions. The final communites could be classified into two classes, with striking similarity between the four replicates (right panel).

2   Proposed Methods

Despite of the complexity of the dataset, the reproducibility of the experiments is quite remarkable (see Figure 1 & 2). Given this reproducibility, I aim to parameterize population dynamics models following different strategies. The main rational behind the strategy is that it may be possible to infer interactions from correlations between either abundances, genes or both. Therefore, the first part of the project will attempt to infer networks of putative interactions, reducing its complexity with network analysis and studying if it is possible to interpret network motifs ecologically. This would allow us to build models of Lotka-Volterra type ((Allesina & Tang 2012); (Taillefumier et al. 2017); (Marsland III et al. 2019)) where the information about the interplay between species and resources is kept implicit. For this task, the beech-tree dataset will be used.

A second approximation consists of predicting the underlying resources underlying the population dynamics. Starting from the metagenomics information, we aim to infer community-level metabolic networks as outlined by (Henry et al. 2016). In addition, if metabolomics data is available, metagenomics composition can be linked with the underlying resources (Noecker et al. 2016). We will make this prediction for each time point of the marine particles data, for which metabolomics information is available. Then, we will analyse the dynamics of resources looking for correlations, as it was done in the previous task. Combining the network of resources and the one of species abundances, we will attempt to parameterize population dynamics models with both resources and consumers explicit ((Marsland III et al. 2019); (Taillefumier et al. 2017) (Posfai et al. 2017); (Goyal & Maslov 2018); (Tikhonov & Monasson 2017)).
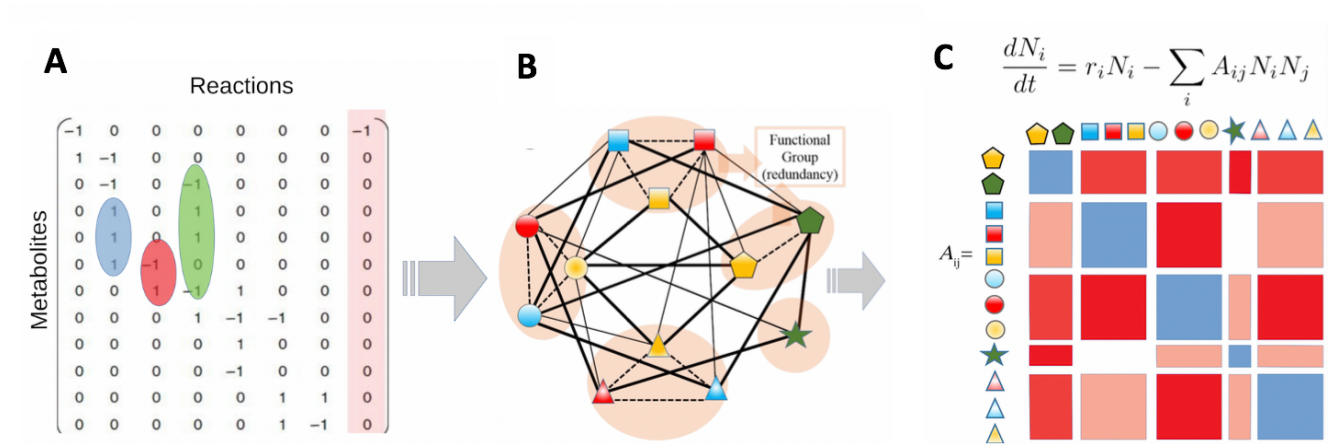
Figure 2: A) Community level metabolic matrix of the contribution of each species. B) Shape's represent unique species contributions. These are used in large communities, identify those with the same contribution to determine functional groups. C) An interaction matrix (A_{ij}) represents this analysis with competition within functional groups in blue and cooperation between them in red.

3   Outcomes

This project will culminate in a report produced in the style of a scientific paper, in partial fulfillment of an MSc in Computational Methods in Ecology and Evolution at the Imperial College London. The findings of this project will be presented and defended in a viva during the month of September 2019.

4   Project Timeline

The project will run over a six month period beginning the 1st of April 2019 and ending the 28th of September 2019. I will be based in ETH Zurich for the months of May, June and July where I will work directly with my supervisor Dr. Alberto Pascual-Garcia.

Gantt chart	April				May				June				July				August				September			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
The Project																								
Project Proposal	■																							
Review Methods	■	■	■																					
Model Metabs				■	■	■	■																	
Model Pop Dyns								■	■	■	■													
Analyse Results													■	■	■	■								
Literature	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■				
The Report																								
Introduction	■	■	■	■	■	■	■	■	■	■	■						■	■	■	■				
Methods					■	■	■	■	■	■	■													
Results													■	■	■	■								
Discussion															■	■	■	■	■	■				
The Assesments																								
Submission																			■					
Presentation																					■	■	■	■
viva																					■	■	■	■

5   Budget

Costs calculated in pound sterling. Accomodation outlines the cost for three months accomodation in Zurich for the months of May, June and July 2019. Travel covers the cost a return flight. Mobility fund is 291.36 pounds contributed by the Swiss government each month I am there. This is part of the Swiss-European Mobility Fund. In total, a deficit of £1031 will be incurred.

Table 1: All expected expenses and income associated with the project. Currency is pounds Sterling.

Details	Expense	Income
Accomodation	1705	
Travel	200	
Mobility Fund		874
Course Budget		??
Total	1905	874