The specificity - abundance relationship in microbial communities

Mahendra Mariadassou, Dieter Ebert, Samuel Pichon

Proposed Minisymposium Title:

Modeling Microbial ecosystem using Meta-Omic data

Organisers:

Dr. Beatrice Laroche
National Institute for Agricultural Research
UR341 Applied Mathematics and Informatics
Domaine de Vilvert
FR-78352 Jouy-en-Josas
France
Ph: +33 (0)1-34-65-2843
Em: beatrice.laroche@jouy.inra.fr

Dr. Sophie Schbath
National Institute for Agricultural Research
UR1077 Mathematics and Genome
Domaine de Vilvert
FR-78352 Jouy-en-Josas
France
Ph: +33 (0)1-34-65-2890
Em: sophie.schbath@jouy.inra.fr

Main contributors:

Dr. Julie Aubert
National Institute for Agricultural Research
UMR518 INRA AgroParisTech
Mathématiques et Informatique Appliquées
16 rue Claude Bernard,
F - 75 231 Paris Cedex 05
France
Ph: +33 (0)1 44 08 1673
Em: julie.aubert@agroparistech.fr

Pr. Thomas Curtis
University of Newcastle
School of Civil Engineering and Geosciences
Cassie Building
Newcastle upon Tyne, NE1 7RU
United Kingdom
Ph: +44 (0) 191 208 6690
Em: tom.curtis@ncl.ac.uk

Dr. Mahendra Mariadassou
National Institute for Agricultural Research
UR1077 Applied Mathematics and Informatics
Domaine de Vilvert
FR-78352 Jouy-en-Josas
France
Ph: +33 (0)1-34-65-2900
Em: mahendra.mariadassou@jouy.inra.fr

Pr. William Sloan
University of Glasgow
School of Engineering
Environmental genomics and bioinformatics
Rankine Building
Glasgow G12 8LT
Ph: +44 (0) 1413304076
Em: william.sloan@glasgow.ac.uk

Session Description:

Technologies make it now possible to obtain a large number of data concerning microbial ecosystems, such as a set of genes (genomics), RNA (transcriptomics), proteins (proteomics), metabolites (metabolomics), etc. “Meta”-omic approaches are used to understand microbial communities as a whole. These communities can be composed of a small number of microorganisms or a huge crowd, depending on the type of ecosystem considered, but in all cases the complexity of these data is a challenge to achieving their integration, analysis and modelling.
This session therefore focuses on modelling and model-based applications in microbial ecology using various modelling techniques ranging from statistical to mechanistic models, with meta-omic data to support the modelling and model analysis research presented.
A secondary goal of the session is to provide a broad cross section of the area, rather than focusing on specific application areas, leading to fruitful comparison and discussions.

The proposed presentations illustrate the use of meta-genomic and meta-transcriptomic data for the modelling of

- Environmental Microbial Ecosystem and Microbial Bioprocesses
- Animal or human microbiome
- Food Microbial Ecosystem.

Schedule: Monday, June 19th, 16h-19h

List of presentations (speaker):
16h00-16h30: The specificity-abundance relationship in microbial communities. (Mahendra Mariadassou)
16h30-17h00: Predicting microbial community assembly. (William Sloan)
17h00-17h30: Energy and Rates of Evolution in Bacteria. (Thomas Curtis)
17h30-18h00: Metagenomics data analysis using a latent block model: application to plant-microbial communities interactions in the rhizosphere. (Julie Aubert)
18h00-18h30: Metagenomic data analysis and integration in a functional population model of fiber degradation by the human intestinal microbiota. (Sebastien Raguideau)
18h30-19h00: Study of the stability and functional redundancy of a food microbial ecosystem using physico-chemical and transcriptomics data. (Frederic Fer)

Mahendra Mariadassou
INRA, Jouy-en-Josas
e-mail: mahendra.mariadassou@jouy.inra.fr

Samuel Pichon
University of Basel
e-mail: samuel.pichon@unibas.ch

Dieter Ebert
University of Basel
e-mail: dieter.ebert@unibas.ch

The specificity-abundance relationship in microbial communities.

Microbial communities are found in wide array of diverse environments, ranging from the human skin to built freshwater lakes. The species-abundance curve of these communities is usually highly skewed and typical of communities where the diversity mostly comes from a lot of rare species but a few abundant ones contribute the bulk of the individuals. Several community assembly processes have been suggested: they generally assume that local communities are assembled by sampling in some way from a meta pool of species. Neutral processes constitute a useful null model to test observed community structures against expected ones. They assume that both species and samples from different environments are exchangeable.
Here we use an approach based on the local specificity of a species to its environment to explore this assembly process within a set of communities coming from closely related environments. Our analyses reveal a non-random assembly process, with locally abundant species much more specific to their environment than expected under a neutral model, either using or not overall abundances. A more sophisticated model that accounts for environmental filtering and environment-specific carrying capacities to bias the sampling probabilities achieves a better fit to the data. This underlines the need to incorporate ecological processes when studying assembly processes. At the same time, it suggests that abundant species are the cornerstone of differences between conditions and that shallow sampling efforts are sufficient to identify differences between communities, in accordance with numerous empirical findings.

William Sloan
Glasgow University,UK
email: William.Sloan@glasgow.ac.uk

Predicting microbial community assembly.

The functioning of complex and diverse communities of microorganisms is critical in agriculture, engineering and medicine. Yet the coming together, or assembly, of these communities is poorly understood. When the microbial communities fail to function as we hoped they might it is often attributed to the 'wrong' community microbes having assembled. But we currently have no means of ensuring that the 'right' community assembles. Thus many economically important bio technologies are engineered by a process of trial and error. The ability to predict and hence manage the process of community assembly would thus have wide set applications. We use high resolution time series pyrosequenced amplicons from multiple bioreactors to show that a simple birth-death-immigration process can capture much of the variance in community assembly and that by controlling immigration the probability that key species form part of the community can potentially be controlled.

T.P. Curtis
Newcastle University, UK
email: tom.curtis@newcastle.ac.uk

L. Øvreås
Bergen University, Norway

W.T. Sloan
Glasgow University, UK.

Energetics and evolution.

It is now widely accepted that the diversity of the microbial world is very large indeed. However, the rate of evolution appears to remarkably slow and possibly variable. The putative diversity of differing functional groups implies differing rates of evolution, for the age of the functional group does not match the diversity of the group. What then controls the rate of evolution? There appears to be a relationship between the putative diversity of differing functional groups and the free energy available to those organisms. Interestingly, there are four established patterns in genetic variations: larger genomes have fewer mutations, larger genomes have fewer deleterious mutations, deletions are favoured over mutations and AT mutations are favoured over GC mutations. Each of these established methods can be related to the energy available to the cell. We hypothesise that there is an energetic burden to the carriage of a mutation. And therefore organism with more energy are better able to support that mutation. We further hypothesise that organisms that carry more mutations can evolve more quickly. To prove this hypothesis we must establish the levels of mutations in differing functional groups. This can be most easily achieved using metagenomic data, ideally reconstructed into genomes. However, it is far from clear that we have the mathematical tools to do this.

Julie Aubert
INRA, Paris, France
AgroParisTech, Paris, France
email : julie.aubert@agroparistech.fr

Metagenomics data analysis using a latent block model: application to plant-microbial communities interactions in the rhizosphere.

Binary or count matrices are widely used in numerous fields and namely in ecology. Metagenomics which studies microbial communities directly from environmental samples, provides abundance matrices where rows correspond to bacteria and columns to biological samples. One major goal is to find associations between bacteria communities and biological samples. We will propose a model and tassociated inference procedures for a simultaneous clustering: one on the populations of bacteria constituting the metagenome, and the other one on the samples. We will use the latent block model framework introduced by Govaert and Nadif (2010). As metagenomics data are abundance matrices, it seems natural to use a Poisson distribution. Nevertheless recent studies show that metagenomics data are sparse and overdispersed. We will thus propose a Zero-inflated Poisson in order to take this excess of zero counts into account. Since latent variables (unknown labels of groups respectively in rows and columns) are not independent conditionally on observed variables, the classical maximum likelihood inference is impossible. We will present an inference algorithm based on a variational approach (Wainwright and Jordan (2008)). We will apply this model to metagenomics data in order to study the plant-microbial communities interactions in the rhizosphere, the region of soil directly influenced by root secretions and associated soil microorganisms.

Keywords : mixture model, clustering, count data, metagenomics

References :

Govaert, G. et Nadif, M. (2010), *Latent Block Model for contingency table*, Communications in Statistics – Theory and Methods, 39, 3, 416-425.
Wainwright, M.-J. And Jordan, M. I. (2008), *Graphical models, exponentials families, and variational inference*. Foundations and Trends in Machine Learning, Vol. 1, Numbers 1-2, pp. 1 :305.

Béatrice Laroche
INRA, Jouy-en-Josas, France
email: Beatrice.Laroche@jouy.inra.fr

Marion Leclerc
INRA, Jouy-en-Josas, France
email: Marion.Leclerc@jouy.inra.fr

Sébastien Raguideau
INRA, Jouy-en-Josas, France
email: Sebastien.Raguideau@jouy.inra.fr

Metagenomic data analysis and integration in a functional population model of fiber degradation by the human intestinal microbiota.

The human intestinal microbiota is a complex microbial ecosystem that plays a crucial role in several aspects of human health. It is particularly involved in the metabolism of residual fibers, through anaerobic digestion, thus providing significant energy (Short Chain Fatty Acids, simple sugars) and vitamins to the host. Whole Genome Sequencing (WGS) data from metagenomic analyses give an insight of the content in terms of genes of an entire microbial community, even if the organisms that compose it cannot be cultivated. In addition to the potential of conventional molecular inventory techniques (such as targeting ADNr16s), which allows an analysis of diversity, WGS approaches provide an access to the functions.

Our main goal is to integrate this functional information in a previously developed higher level mechanistic model of the microbiota carbohydrate trophic chain[1]. Thus, we investigate how the concept of functional population is relevant in order to synthesize WGS based information.
We formulate a problem similar to a regularized Blind Source Separation (BSS) approach which allows us to create functional profiles. As we want theses profiles to be biologically meaningful, the regularization term is critical since it let us incorporate prior information during the profiles creation process. Therefore we explore a variety of classic penalization (sparsity, lasso and elastic net lasso) while developing our original criteria.

[1] R. Muñoz-Tamayo, B. Laroche, Éric Walter, J. Doré, and M. Leclerc. Mathematical modelling of carbohydrate degradation by human colonic microbiota. J. Theor. Biol., 266(1): 189-201, 2010.

Frédéric Fer
INRA, AgroParisTech, UMR 1319 MICALIS, France
INRA, AgroParisTech, UMR 518 MIA, France
frederic.fer @agroparistech.fr

Jean-Marie Beckerich,
INRA, AgroParisTech, UMR 1319 MICALIS, France

Béatrice Laroche,
INRA, UR 341 MIA, France

Sébastien Gaucel
INRA, Agroparistech, UMR 728 GMPA, France

Julie Aubert
INRA, AgroParisTech, UMR 518 MIA, France

Combining kinetic modeling and transcriptomic analysis to study the resilience on a model cheese reduced ecosystem

Keywords: cheese ecosystem, mathematical modeling, differential analysis, RNA-seq, resilience, deterministic model.

The emergence of "omics" methods is an opportunity to understanding cheese microbial ecosystems functions, and their resilience, which represents a major scientific and economic challenge. To study its resilience, capacity for an ecosystem to maintain its function following a perturbation, a reduced and controlled ecosystem of a washed rind cheese [1] was challenged by modifying either the salinity of the curd or by omitting one of the two major yeast species. Metatranscriptomics and physiochemical data was acquired during each ripening. We propose a two-step approach combining these different kind of data to study the resilience and apply it on casein proteolysis.

In order to identify the role of each species in the proteolysis resilience, we have developed a kinetic model integrating the biochemical data of proteolysis and the population dynamics. Model parameters were estimated by Metropolis-Hastings algorithm and MCMC chains using the MCMC Matlab Toolbox [2]. With this approach, we have highlighted the major role of the yeasts *Debaryomyces hansenii* and *Geotrichum candidum* in this phenomenon.

Then we have focused transcriptomic analysis on these major species. We selected genes declared differentially expressed using the R package DESeq2 [3] between a disturbed condition and the normal one at 5% after correcting for multiple testing. As the number of selected genes was exceeding the number of observations, we have used a lasso penalized regression model to identify genes potentially predictors for the proteolysis.

Our approach has identified a set of implicated genes. This work is an example of an approach which integrates diverse data (microbial growth, biochemical, physiochemical and genomic data) in order to understand the microbial ecosystem

resilience. Our approach is a general one for studying microbial ecosystems with high-throughput methods and has shown the feasibility of a functional microbial ecology.

[1] Mounier, C. Monnet, T. Vallaeys, R. Arditi, A.S. Sarthou, A. H elias, and F. Irlinger. *Microbial interactions within*
a cheese microbial community. Applied and environmental microbiology, 74(1) :172–181, 2008
[2] H. Haario, M. Laine, A. Mira and E. Saksman, 2006. *DRAM: Efficient adaptive MCMC*, Statistics and Computing **16**, pp.
339-354.
[3] Simon Anders, Wolfgang Huber: *Differential expression analysis for sequence count data* Genome Biology **11** (2010) R106, J.