UNIVERSIDAD ADOLFO IBÁÑEZ

DOCTORAL THESIS

Contributions to Metagenomics: Classification, Dynamics and Applications

---

Author:
Dante Travisany

Committee:
Eric GOLES
Alejandro MAASS
Gonzalo RUZ
Pedro MONTEALEGRE
Pablo MARQUET

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

April 15, 2019

Abstract

A microbiome is a community of microorganisms that inhabit a particular environment like soils, oceans or the human body. Those communities are made of a trillion of microorganisms divided into a dozen or even hundreds of different species which interacts among them. The community behaves as a complex adaptive system and fluctuates in response to changes in environmental factors such as acidity, pressure, temperature or, when are host related it is sensible to perturbations like antibiotics, changes in the diet and lifestyle factors. The recovery of the genetic material of the microbial communities directly from their environment, the classification and study of these samples is called Metagenomics. The digitalization of the genetic material is done by high throughput sequencing technologies, such as Illumina which shreds the DNA into fragments called reads.

In the first part of this work we propose a alignment-free method capable of assign taxa to each read in a metagenomic sample by analyzing the statistical properties of the reads. Given an environment, we collect genomes from public available databases and generate synthetic genomic fragments libraries. Then, statistics of k-mer frequencies are computed and stored in an environment-associated dataset used to build a robust machine learning procedure based on multiple CART trees.

In the second part we study the dynamics of a gut microbiome under antibiotic perturbation and *Clostridium difficile* infection (CDI). Here, interactions play a key role in the development of the disease. Using a previously Boolean network model for CDI we demonstrate that this model is in fact a threshold Boolean network (TBN). Once the TBN model is set, we further explore the space of possible interactions generating an evolutionary algorithm to identify alternative TBNs. Allowing the construction of a neutral space conformed by a set of models that differ in their interactions, but share the final community states of the gut microbiome under antibiotic perturbation and CDI. We organize the resulting TBNs into clusters that share similar dynamic behaviors and the most relevant interactions are identified. Finally, we discuss how these interactions can either affect or prevent CDI.

In the third part we examined the microbial community across a transect of few kilometers long, where there is a remarkable abiotic variations like pH, temperature and humidity gradient, with acidic soils, to name a few. We studied the compositional structure of the community and constructed two co-occurrence networks representing two sections that divided the transect. Using network analysis, we examined changes in putative ecological interactions among microbial Operational Taxonomic Units (OTUs), as well as their associations to physicochemical and nutritional variables. Network comparisons allowed us to examine the nature of the ecological rearrangements that take place in the microbial community when facing contrasting environments. L-GRAAL, the graph alignment method we used provides a comprehensive way to understand topological shifts among members from two networks. We show here that this method provides a glimpse into the nature of the changes in microbial communities that can foster resistance and resilience to contrasting environmental conditions.

Acknowledgements

I would like to express my sincere gratitude to Eric for his continuous support and constant bullying. Also for his patience, motivation and charisma. His guidance construct outstanding researchers, but more important is the confident environment generated by his presence.

I am profoundly grateful to Alejandro for his example of rightness, methodology and professionalism. He is an exemplary leader.

I would like to thank Pedro and Patricia for their warm welcome in Orleans and their selfless aid and help during my internship there.

I greatly appreciate to Dr. Gonzalo Ruz and Dr. Pablo Marquet for being interested in my work and being part of the committee.

I gratefully acknowledge the funding received towards the Centro de Modelamiento Matemático CONICY PIA 170001 and BASAL PBF003, to the Centro para la Regulación del Genoma FONDAP 15090007, the Adolfo Ibañez excellence scholarship and CONICYT PFCHA/Beca Doctorado Nacional 2015/FOLIO 21150895.

Many thanks to my lab mates, María Paz, Vicente, Ricardo, Natalia for their time, good vibes and recomendations. Many thanks also to the former lab members Alex, Marko and Karina.

Thanks to my family, Carolina for her unconditional support; to Sofía, Florencia and Dante for their love and joy.

Thanks to my parents for being there always.

Contents