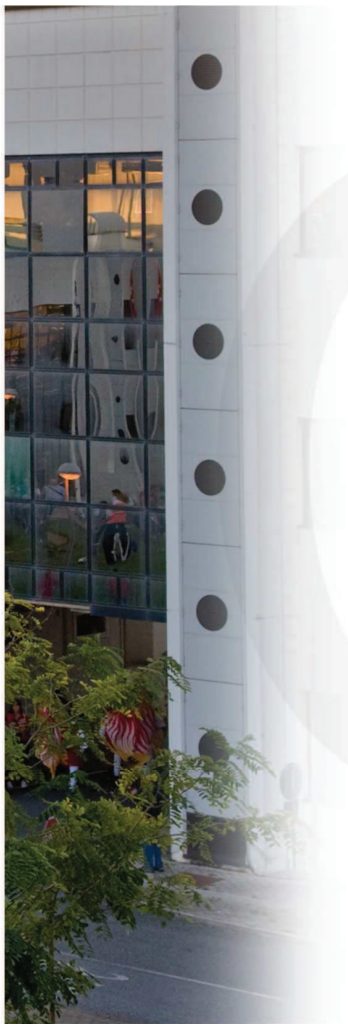


Consiglio Nazionale delle Ricerche
IFC - Istituto di Fisiologia Clinica

MAST4HEALTH
Secondment Day

Claudia Torino – Antonio Demetrio Vilasi
National Research Council (C.N.R.)
Institute of Clinical Physiology, Sede Secondaria
Reggio Calabria, Italy

My discipline area and expertise

MAST 4 HEALTH

Claudia Torino

Area: Epidemiology

Expertise:

- Clinical Epidemiology
- Statistical analysis
- Study design
- Bioinformatics

Reasons for secondment →

To deepen my knowledge in the field of bioinformatics and, as a biologist, to learn more about new frontiers in genetics.

My discipline area and expertise

MAST 4 HEALTH

Antonio Demetrio Vilasi

Area: Epidemiology

Expertise:

- Bioinformatics
- Internet security
- Disaster recovery
- Database management

Reasons for secondment →

To deepen my knowledge in the field of bioinformatics applied to genetics

Secondment venue

Valencia, Spain

Biotech Vana SL →SME in bioinformatics

Managed by researchers, covers a variety of scientific business models, including analysis of high-throughput genomic data

MAST4HEALTH activities →

- Preprocessing of raw data
- Mapping and assignation of taxonomy levels

Opportunities

Increased Knowledge in→

genetics applied to microbiota taxonomy

- Lessons on bioinformatics and sequence analysis
- Practical sessions
  - Training on software used for sequence alignment
  - Analysis of sequences obtained by Next Generation Sequencing

Activities

- Meetings on MAST4HEALTH
- Training in bioinformatics
  - Study of tools available on NCBI (National Center for Biotechnology Information) for gene analysis
  - Use of software (R and Bioconductor) for the analysis of high throughput genomic data
  - Study of DNA variants by using the Genome Analysis Toolkit (GATK) software tools
- Application of bioinformatics in Next-Generation Sequencing (NGS) for microbiota studies
  - NGS approach enables the analysis of genetic diversity

Networking

Future plans:

Maintaining contacts for future collaboration in European project proposals →

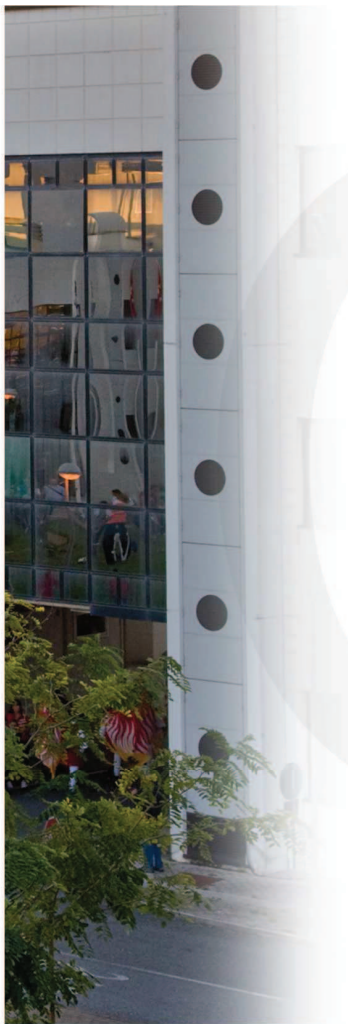
Bioinformatics applied to genetics (BIOTECHVANA)

Genetic epidemiology (CNR)

Suggestions

A dedicated platform / organization for a networking of all secondees to:

- Maintain relations
- Share their experiences



Consiglio Nazionale delle Ricerche
IFC - Istituto di Fisiologia Clinica

MAST4HEALTH
Secondment Day

Marco Laurino
National Research Council (C.N.R.)
Institute of Clinical Physiology, Pisa, Italy

My discipline area and expertise

MAST 4 HEALTH

Area: Bioengineering

Expertise:

- Biomedical signal processing
- Biomedical systems modelling
- Control system
- Statistics
- Neuroscience

Biotechvana

MAST4HEALTH

SME in bioinformatics created in 2006 as spin-off of the Instituto Cavanilles at the University of Valencia

Managed by researchers for researchers it covers a variety of scientific business models →

offering know-how

IT infrastructures

Software and R&D products

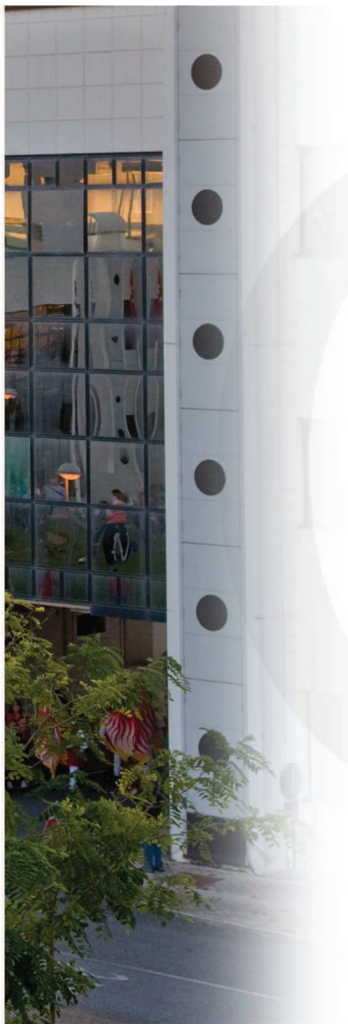
Personal evaluation & Opportunities

- Positive experience
- Dealing with new topics, forwards my research line
- Acquiring a quick, good training by experts
- Working within a company
- New scientific collaboratations

Activities

Training in genomic & proteomic data analysis obtained by "High-throughput sequencing" technologies & "RNA sequencing" (RNAseq).

- Quality analysis and filtering of genomic and proteomic data
  - DNA/RNA Read mapping (TopHat/Bowtie2 algorithms)
  - Analysis of the differential expression (Coset/Cufflinks and Deseq/EdgeR/CuffDiff algorithms)



Consiglio Nazionale delle Ricerche
IFC - Istituto di Fisiologia Clinica

MAST4HEALTH
Secondment Day

Francesco Sansone
National Research Council (C.N.R.)
Institute of Clinical Physiology, Pisa, Italy

My discipline area and expertise

MAST 4 HEALTH

Area:

ICT – Biotechnology

Expertise:

Computer Scientist

Software Developer

Secondment venue

MAST4HEALTH

BIOTECHVANA

Spanish SME in Valencia, created in 2006 as a spin-off of the Insituto Cavanilles at the University of Valencia

Opportunities

MAST4HEALTH

Very interesting experience

Great opportunity of seeing the world with my eyes

Know something new about the work of some new colleagues

Activities

MAST 4 HEALTH

Target: Manage a Bayesian Network pipeline like a Service

Result: A new (web) component

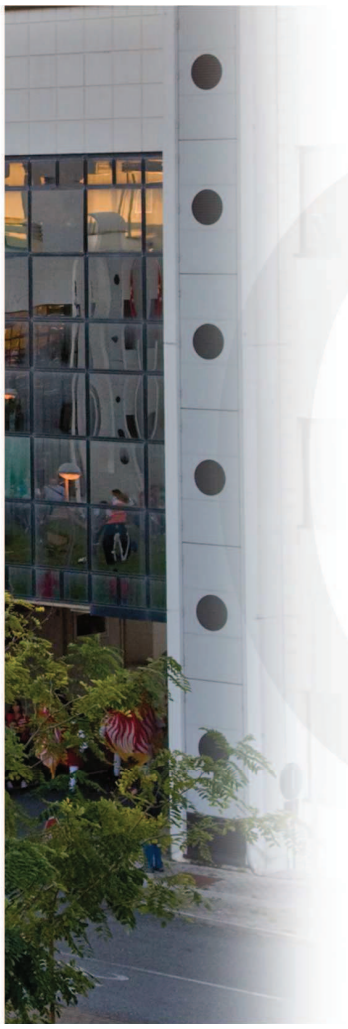
Suggestions/Changes

Suggestion

Preplanning the activities before the departure

Future exchange programs?

I would participate for sure!



Consiglio Nazionale delle Ricerche
IFC - Istituto di Fisiologia Clinica

MAST4HEALTH
Secondment Day

Marco Scalese
National Research Council (C.N.R.)
Institute of Clinical Physiology, Pisa, Italy

My discipline area and expertise

MAST4HEALTH

Area: Statistician

Expertise:

- Mathematical techniques to analyze and interpret data and draw conclusions
- Social epidemiology
- Population survey
- Health care (prevention and diagnosis of disease)

BIOTECHVANA

MAST4HEALTH

Biotech Vana SL (Biotechvana) is a SME in bioinformatics created in 2006 as a spin-off of the Instituto Cavanilles at the University of Valencia. The Enterprise is managed by researchers for researchers and covers a variety of scientific business models offering know-how, IT infrastructures, software and R&D products to out customers.

Opportunities

Positive experience in this new scientific collaborations, the topics covered could be useful in the future even if these ones were different from the work I usually do at CNR

Activities

MAST4HEALTH

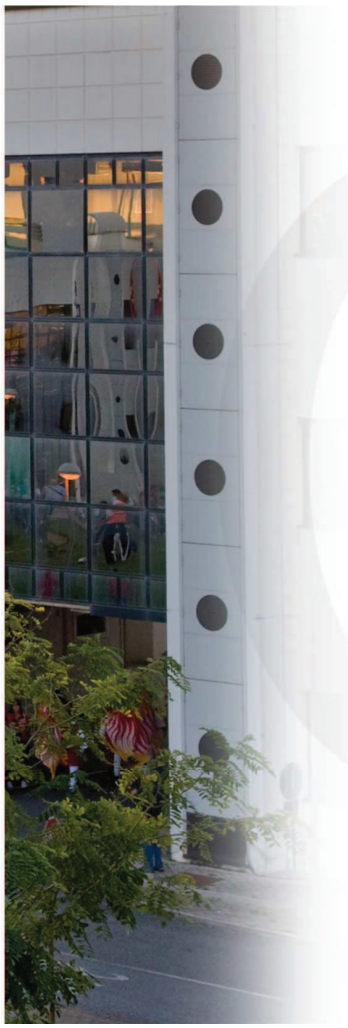
applies a suite of genomic technologies and bioinformatics tools to study the composition and diversity of microbial populations present in human body (microbiota) through their genetic content. Tries to understand how environmental factors can unbalance the homeostasis of this system and how this unbalance is related with the development of certain pathologies

Suggestions

Future exchange programs

mitoFOIE GRAS: Non-invasive Profiling of Mitochondrial Function in Non-Alcoholic Fatty Liver Disease (NAFLD).

Together with several PhD students, the team will share expertises and work synergistically along the value creation chain to address the unmet medical need of more informative NAFLD assessment.



Consiglio Nazionale delle Ricerche
IFC - Istituto di Fisiologia Clinica

MAST4HEALTH
Secondment Day

Letizia Guiducci, PhD
National Research Council (C.N.R.)
Institute of Clinical Physiology, Pisa, Italy

Secondment in July 2019

My discipline area and expertise

MAST 4 HEALTH

- Area:

Microbiota and pathophysiology of fetal development and developmental age. Impact on Autism Spectre Disorders

The role of Microbiota in Obesity, diabetes and metabolic diseases and related cardiovascular.

- Expertise:

Clinical research projects:

Gut to brain interaction in autism. Role of probiotics on clinical, biochemical and neurophysiologic parameters.

(Project Code: GR-2011-02348280, Project Type: Young Researcher, under 40 years)

MEDUSA Study: The perception of emotion in women: the menopause as a model of the relationship between microbiota and sexual steroids, stress and emotions.

Secondment venue

MAST4HEALTH

VENUE: BIOTECHVANA (supervisor, Carlos Llorens)
Lab. Parc Cientific Universitat de *Valencia*

CONTACT: FISABIO-Public Health develops multidisciplinary research aimed at increasing the quality of Public Health

Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana

The Center has close ties with policy makers, hospitals, primary health centers, universities and biomedical research institutions, as well as collaborative partnerships with pharmaceutical and biotech companies

Research focuses on several areas:

- Environmental Health
- Cancer and Public Health
- Vaccines
  Genomics and Health
- Health Inequalities
- Rare Diseases
- Food safety
- Health Services
- Global Health

HOST: M. Pilar Francino
Head, Department of Genomics and Health FISABIO-Public Health
Valencian Government, Health Department

Opportunities

The secondment has largely benefited my skills, knowledge and experience both professional than personal

The initial objectives have been achieved
1. to obtain some basic theoretical knowledges on microbial genomics and metagenomics
2. to learn the research lab protocols in microbioma analysis from human feces samples

Activities

MAST4HEALTH

Theoretical lessons on microbial genomics and metagenomics

- Introductory lectures on microbial genomics and metagenomics
- Brief introduction to the Unix system
- Sequence processing and QC
- 16S rRNA analysis
- Read mapping and variant calling
- Sequence assembly and annotation
- Integrated Microbial Genomes (IMG) Data Management and Analysis System

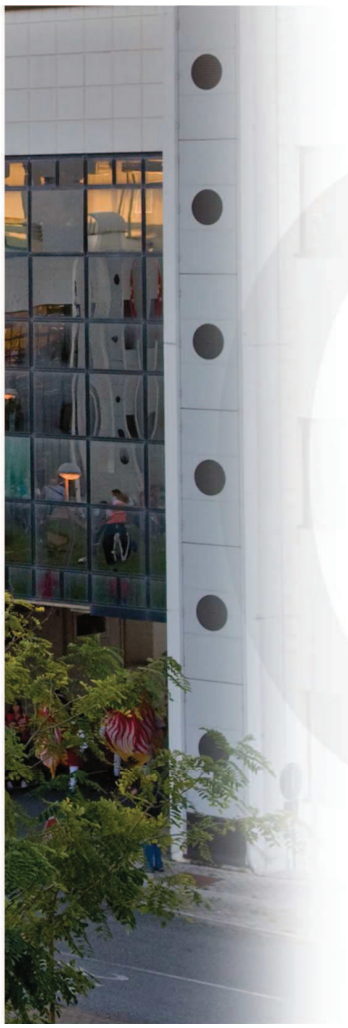
Microbioma analysis from human feces samples

- Preparation samples for DNA extraction,
- DNA Extraction in Magna Pure LC (Roche)
- Qualitative and Quantitative analysis of DNA extracted by electrophoresis and QuBit Fluoromiter.
- Amplicon PCR: to amplify template out of a DNA sample using region of interest-specific primers with overhang adapters attached
- PCR Clean-Up: this step uses AMPure XP beads to purify the 16S V3 and V4 amplicon away from free primers and primer dimer species.
- Index PCR: this step attaches dual indices and Illumina sequencing adapters using the Nextera XT Index Kit.
- PCR Clean-Up 2: this step uses AMPure XP beads to clean up the final library before quantification.
- Library Quantification, Normalization, and Pooling

Networking/Suggestions

Maintain contacts

Collaborate on projects of common interest: involvement of the microbiota in the health of the child (obesity and autistic spectrum disorders) and women's health



Consiglio Nazionale delle Ricerche
IFC - Istituto di Fisiologia Clinica

MAST4HEALTH
Secondment Day

Egeria Scoditti
National Research Council (C.N.R.)
Institute of Clinical Physiology, Lecce, Italy

My discipline area and expertise

MAST4HEALTH

Area:

Nutrigenomics, Vascular Biology.

Expertise:

- Cell and tissue culture
- Gene/miRNA expression analysis
- Signaling pathway analysis

Secondment venue

FISABIO

Valencia - Spain

Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana

Unit of Public Health

Research areas

- ✓ Environmental Health
- ✓ Cancer
- ✓ Vaccine
- ✓ **Genomics and Health**
- ✓ Inequalities in Health
- ✓ Rare Diseases
- ✓ Food safety
- ✓ Health Services
- ✓ Global Health

Lines of Research

- ✓ Genomics of microorganisms
- ✓ Metagenomics of microbial communities
- ✓ Molecular epidemiology and evolution of microorganisms
- ✓ Bioinformatics and Biostatistics

Head

M. Pilar Francino, PhD.

Opportunities

MAST4HEALTH

It was a great experience both from professional and personal perspective.

I was followed in each step by expert researchers and technicians in a open, friendly and cooperative environment.

Sharing of knowledge, research efforts and problems, skills and experiences was easy and productive.

It gave me the opportunity to increase my scientific knowledge, competence and skills, research objectives and network.

Activities

My secondment was centered on human gut microbiome studies, from samples processing, genomic analysis to bioinformatics data analysis with state-of-the-art tools and methods.

I participated in theoretical training in the field and actively took part in laboratory procedures as well as data analysis.

Networking

Potential partnership for future joint research activities and projects.

Participation in workshops/seminars organized by the host institution.

Suggestions

Secondment within the project H2020-MSCA-RISE mitoFOIE GRAS

Measuring the complexity of microbial communities

GIULIANO RESCE

Secondment Venue Biotechvana, Contact Fisabio

Area of working

- Giuliano has a Ph.D in Economics, and Master's degree in Political Science and Decision-Making Processes
- His research focuses on Public Economics, Inequality, Health, Education, and Environmental Sustainability
- He has a consolidated expertise in the use of Big Data, Text Mining, Machine Learning, Applied Optimizations, Multi Criteria Decision Analysis, and Impact Evaluation tools

Background

- This research proposes the Method of Reflection to evaluate diversity, ubiquity, and complexity of microbial communities. Two novel indices are proposed: the **Community Complexity Index (CCI)** and **Bacterium Complexity Index (BCI)**, which methodologically reflect respectively two influential measures in the recent economic literature: the Economic Complexity Index and Product Complexity Index developed by Hidalgo and Hausmann (2008, 2009).
- As first application this study uses abundance contingency data at genus level for 13 pairs of mother-infant in 7 time points (1 week, 1 month, 3 months, 7 months, 12 months, mother before delivery, mother 12 months after delivery).

Application in Economics

- According to Hausmann et al. (2014), the complexity of an economy is related to the multiplicity of useful knowledge embedded in it. Economic complexity is there expressed in the composition of a country's productive output.
- Some goods, like medical imaging devices or jet engines, embed large amounts of knowledge and are the results of very large networks of people and organizations. By contrast, wood logs or coffee, embed much less knowledge, and the networks required to support these operations do not need to be as large.

How do Hidalgo and Hausmann (2008, 2009) go from what a country makes to what a country knows?

- First, countries whose residents and organizations possess more knowledge have what it takes to produce a **more diverse set of products**. In other words, the amount of embedded knowledge that a country has is expressed in its productive diversity, or the number of distinct products that it makes.
- Second, products that demand large volumes of knowledge are feasible only in the few places.
- To the aim of measuring this, Hidalgo and Hausmann (2008, 2009) define ubiquity as the number of countries that make a product.
- Using this terminology, they observe that **complex products are less ubiquitous**. The ubiquity of a product, therefore, can reveal information about the volume of knowledge that is required for its production.
- Hence, **the amount of knowledge that a country has is expressed in the diversity and ubiquity of the products that it makes**.

Diversity and ubiquity are, respectively, crude approximations of the variety of capabilities available in a country or required by a product

- It is worth noting that ubiquity of a product can be due to both scarcity (like rare natural resources, such as uranium or diamonds) or complexity of the economy.
- One way to see whether low ubiquity originates in scarcity or complexity is by looking at the diversity of the countries making those products. If these countries have low diversity, then it is likely that rarity explains the low ubiquity. However, if the countries that can make these rare products are, in general, more diverse, then it is likely that the low ubiquity of the product reflects complexity.
- Diversity can therefore be used to correct the information carried by ubiquity, and ubiquity can be used to correct the information carried by diversity. This process can be taken a step further by correcting diversity using a measure of ubiquity that has already been corrected by diversity and vice versa. This can be done an infinite number of times using mathematics. This process converges after a few iterations and represents the Hidalgo and Hausmann (2008, 2009) quantitative measures of complexity.
- For countries, they refer to this as the Economic Complexity Index (ECI). The corresponding measure for products gives the Product Complexity Index (PCI).

Application to microbial communities

- As the economic systems, microbial communities, i.e. metagenome, can be represented on the base of diversity and ubiquity of the bacteria within the feces sample obtained by subjects with different characteristics. Diversity is given by the number of bacteria present in each subject sample. Ubiquity is given by the spread of the bacterial strain among the group of subjects. Complexity can be captured by the interaction between ubiquity and diversity.
- Adopting the model of Hidalgo and Hausmann (2008, 2009) to microbial communities we can explain intra- and inter-subject bacteria distribution and variability by two indices of complexity respectively for bacterial communities (intra-subject), **Community Complexity Index (CCI)** and for bacterial strain (inter-subjects), **Bacterium Complexity Index (BCI).**

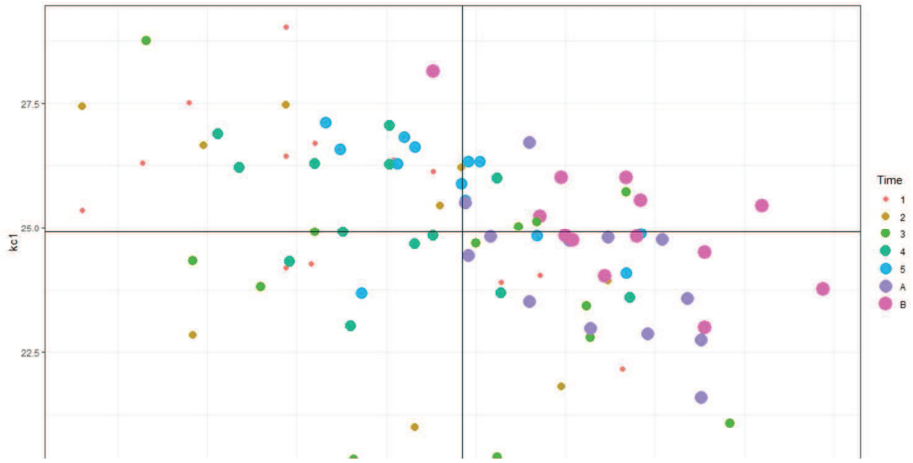
Case study

Data for this analysis are contingency tables for 13 series of samples corresponding to 13 pairs of mother-infant (MIP). Estimates are done on taxa (genus level). Each series has 7 time points:

- I1: infant 1 week;
- I2: infant 1 month;
- I3: infant 3 months;
- I4: infant 7 months;
- I5: infant 12 months;
- MA: mother before delivery;
- MB: mother 12 months after delivery.

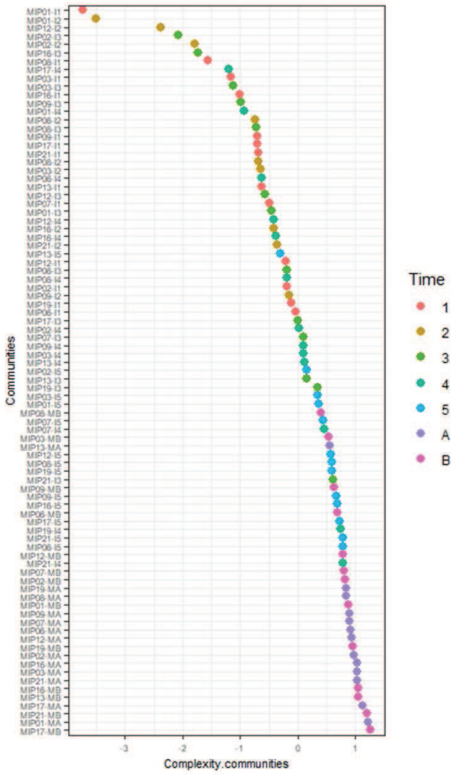
$\vec{K}_{i,0}$ versus $\vec{K}_{i,1}$ by communities

There is a negative correlation between $\vec{K}_{i,0}$ and $\vec{K}_{i,1}$, the most populated quadrants are the second and the forth, meaning that diversified communities tend to have less ubiquitous bacteria in average. In the upper left quadrant (non-diversified communities having ubiquitous bacteria) there are many babes at time 1 and time 2 (infant 1 week and infant 1 month). On the contrary, the lower right quadrant (diversified communities with exclusive bacteria) is mainly populated by babies at time >3 (infant 3 months, infant 7 months, infant 12 months) and mothers both before delivery and 12 months after delivery (A and B).



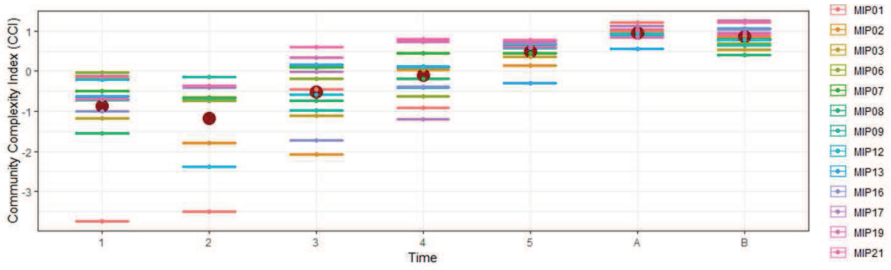
Community Complexity Index

It is interesting to note that the bottom of the raking is mainly populated by babies during the early days of life (time 1 and time 2), when their bacterial flora is still developing. On the other hand, mothers, both before and after delivery (A and B), have the highest CCI (figure 2). MIP01 at time 1 and time 2 shows the lower complexity with a relevant distance from the others.



Community Complexity Index (CCI)

Complexity in babies is increasing with time, but mothers are in average (black dots) more complex than babies, with the average complexity slightly higher in the group A (before delivery) than in the group B (12 months after delivery) (CCI group A = 0.948, CCI group B = 0.855). While babies at time 5 are in average more complex than babies at time 1 (CCI = 0.495 at time 5 v.s. CCI = -0.862 at time 1), the increase in average complexity does not follow monotonically the increase in the age of babies (CCI at time 1 = -0.862, CCI at time 2 = -1.185, CCI at time 3 = -0.512, CCI at time 4 = -0.106, CCI at time 5 = 495). The spread of CCI decreases monotonically with the age (Figure 3), and in mothers the CCI spread is lower in the group A (before delivery) compared with the group B (12 months after delivery).



Bacterium Complexity Index (1/2)

The top 10 complex bacteria			
	$\vec{K}_{j,0}$	BCI	Present in
Aeribacillus	1	2.502321	MIP17-MB
Alcaligenes	1	2.502321	MIP17-MB
Ornithobacterium	1	2.414702	MIP01-MA
Malonomonas	4	2.228189	MIP01-MA, MIP17-MA, MIP17-MB, MIP19-MB
Nonomuraea	1	2.194176	MIP17-MA
Saccharothrix	1	2.194176	MIP17-MA
Taylorella	3	2.037035	MIP01-MA, MIP12-MA, MIP16-MA
Blastochloris	1	2.005941	MIP16-MB
Sedimentibacter	4	1.913263	MIP06-MA, MIP17-MA, MIP17-MB, MIP19-I4
Tetragenococcus	2	1.889289	MIP01-MA, MIP12-MB

Bacterium Complexity Index (2/2)

The bottom 10 complex bacteria			
		BCI	Present in
Kluyvera	1	-6.316	MIP12-I2
Serratia	15	-3.619	MIP01-I1, MIP01-I2, MIP02-I3, MIP06-I2, MIP08-I1, MIP09-I3, MIP12-I1, MIP12-I2, MIP12-I3, MIP16-I4, MIP17-I1, MIP17-I3, MIP17-I4, MIP21-I1, MIP21-I2
Brachymonas	1	-3.418	MIP17-I4
Erwinia	25	-3.285	MIP01-I1, MIP01-I2, MIP01-I3, MIP01-I4, MIP02-I2, MIP02-I3, MIP03-I3, MIP06-I2, MIP08-I1, MIP08-I2, MIP09-I3, MIP12-I1, MIP12-I2, MIP12-I3, MIP12-I4, MIP13-I1, MIP16-I1, MIP16-I2, MIP16-I3, MIP16-I4, MIP17-I1, MIP17-I4, MIP19-I1, MIP21-I1, MIP21-I2
Pantoea	19	-3.243	MIP01-I1, MIP01-I2, MIP02-I2, MIP02-I4, MIP02-I5, MIP03-I1, MIP06-I2, MIP08-I1, MIP09-I3, MIP09-I4, MIP12-I1, MIP12-I2, MIP13-I1, MIP13-I5, MIP16-I1, MIP16-I3, MIP17-I1, MIP21-I1, MIP21-I2
Haemophilus	12	-3.158	MIP01-I1, MIP01-I2, MIP01-I4, MIP06-I4, MIP07-I1, MIP08-MB, MIP08-I1, MIP08-I3, MIP09-I1, MIP12-I3, MIP16-I5, MIP17-I4
Salmonella	30	-3.099	MIP01-I1, MIP01-I2, MIP01-I3, MIP01-I4, MIP02-I1, MIP02-I2, MIP02-I3, MIP03-I1, MIP03-I2, MIP03-I3, MIP06-I2, MIP06-I4, MIP08-I1, MIP08-I2, MIP08-I3, MIP09-I3, MIP12-I1, MIP12-I2, MIP12-I3, MIP12-I4, MIP13-I1, MIP16-I1, MIP16-I2, MIP16-I3, MIP16-I4, MIP17-I1, MIP17-I4, MIP19-I1, MIP21-I1, MIP21-I2,
Klebsiella	10	-3.081	MIP08-I1, MIP08-I2, MIP09-I3, MIP12-I2, MIP16-I1, MIP16-I2, MIP16-I3, MIP16-I4, MIP17-I1, MIP21-I1
Raoultella	4	-3.054	MIP09-MB, MIP12-I2, MIP16-I3, MIP17-I1
Cronobacter	29	-3.024	MIP01-I1, MIP01-I2, MIP01-I3, MIP01-I4, MIP02-I1, MIP02-I2, MIP02-I3, MIP03-I1, MIP03-I2, MIP03-I3, MIP08-I1, MIP08-I2, MIP08-I3, MIP09-I3, MIP12-I1, MIP12-I2, MIP12-I3, MIP12-I4, MIP13-I1, MIP16-I1, MIP16-I2, MIP16-I3, MIP16-I4, MIP17-I1, MIP17-I4, MIP19-I1, MIP19-I5, MIP21-I1, MIP21-I2

Conclusions

- This is a pilot study to demonstrate that the Hidalgo-Haussmann's algorithm previously employed in economics, can also be applied to analyse the complexity of metagenome distribution in human samples, thus allowing to derive new information that can be used to predict health status.
- It is well known that bacterial diversity is associated with a better health status. However, the complexity of bacterial strain and their interaction make difficult to identify which bacterial strains are harmful without taking into account at the same time inter- and intra-subject diversity, ubiquity and complexity. The Hidalgo-Haussmann's algorithm applied to metagenome of human feces samples was able to identify differences in samples obtained at different times of life (new-borns vs before and after delivery of their mothers), bacteria diversity, ubiquity and complexity, in a fully data-driven context.
- The algorithm confirmed that subject complexity, measured by the CCI, increases with age and its standard deviation monotonically decreases (figure 3). At the same time bacterial complexity, given by BCI, increases with age and mother shows the highest BCI as expected (Table 1). The less complex bacteria are present only in babies during the first days of their life, i.e. first month (Table 2).

Next steps

- From pilot to scale the idea, the method of reflection will be applied to Mast4health data

- After the validation process developed on the training database in Valencia, both the **Community Complexity Index (CCI)** and **Bacterium Complexity Index (BCI)** will be the appropriate tools to evaluate effects of Mastiha on the randomized double-blind placebo controlled (parallel arm) clinical trial

References

- Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570-10575.
- Antonelli, C., Crespi, F., Mongeau Ospina, C. A., & Scellato, G. (2017). Knowledge composition, Jacobs externalities and innovation performance in European regions. *Regional Studies*, 51(11), 1708-1720.
- Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. Mit Press.
- Hidalgo, C. A., & Hausmann, R. (2008). A network view of economic development. *Developing alternatives*, 12(1), 5-10.