

Stability of the human microbiome

Health and disease related to the time variability in the human microbiota



Daniel Martínez Martínez

Supervisor: Prof. Andrés Moya Simarro
Dr. Carlos Peña Garay

Advisor: Dra. Carolina Raussell Segarra

Doctorado en Biomedicina y Biotecnología
University of Valencia

This dissertation is submitted for the degree of
PhD of the University of Valencia

September 2018

Cuando emprendas tu viaje a Itaca
pide que el camino sea largo,
lleno de aventuras, lleno de experiencias.
No temas a los lestrigones ni a los cíclopes
ni al colérico Poseidón (...)
Aunque la halles pobre, Itaca no te ha engañado.
Así, sabio como te has vuelto, con tanta experiencia,
entenderás ya qué significan las Itacas.

C.P. Cavafis

Antología Poética.

When I heard the learn'd astronomer,
When the proofs, the figures, were ranged in columns before me,
When I was shown the charts and diagrams, to add, divide, and measure them,
When I sitting heard the astronomer where he lectured with much applause in the lecture
room,
How soon unaccountable I became tired and sick,
Till rising and gliding out I wander'd off by myself,
In the mystical moist night-air, and from time to time,
Look'd up in perfect silence at the stars.

Walt Whitman

And I asked old Jack, 'Do you remember the night
when the sky was so dark and the moon shone so bright?
When a million small children pretending to sleep
nearly didn't have Christmas at all, so to speak?
And would you, if you could, turn that mighty clock back
to that long, fateful night, now think carefully, Jack.
Would you do the whole thing all over again,
knowing what you know now, knowing what you knew then?"
And he smiled, like the old Pumpkin King that I knew,
then turned and asked softly of me, 'Wouldn't you?'

The Night Before Christmas

Dedicación

Una tesis doctoral, como tantas otras cosas en la vida, no es obra de uno mismo solamente. Como dijo Ortega y Gasset, *yo soy yo y mi circunstancia*, y esa circunstancia son todas las personas que han dejado una marca en mí a lo largo de los años. Uno se moldea en gran medida por las personas con las que interactúa, con el intercambio de ideas, opiniones, recuerdos y formas de vivir. Detrás de cada simple palabra aquí escrita hay un sinnúmero de experiencias vitales que han moldeado lo que soy hoy en día, y que sirven como base de lo que seré en el futuro. Es por ello que esta tesis no es solo mía sino de todos vosotros.

Primero, me gustaría agradecer enormemente a mis dos directores, Andrés y Carlos, sin los cuales esta tesis no tendría sentido, y de los que he aprendido tanto a lo largo de estos años. Gracias por confiar en una persona olvidadiza y caótica como yo, gran parte de mi crecimiento como científico os lo debo a vosotros.

Gracias a todos mis compañeros de la carrera y del máster, sin vosotros, todo habría sido mucho más aburrido y anodino. Entre todos ellos quiero destacar a ese grupo de personas que cambió mi vida y me hizo ver la belleza de la ciencia y la biología, y que tan buenos momentos me siguen dando hoy en día: Andrea, Dani, Marta, Jose, Silvia, Ana Rosa, Julián, Ximo, Borja, Martina, Anselm y un sin fin más, perdonad de quien me olvide. El aprecio que os tengo no se puede medir con palabras, y siempre tendréis las puertas abiertas donde yo esté.

A todos mis compañeros de laboratorio, desde los alumnos de doctorado hasta a los investigadores principales, con los que he compartido tanto tiempo y de los que he aprendido siempre mucho. Gracias por hacer del trabajo un lugar al que querer ir.

To the people at Albert Einstein's College of Medicine, specially to Aviv Bergman for his support, and Dan, for the interesting conversations. And, once again Ximo, for helping me to find myself again, and for all the support you gave me there. Also, to Andrew and Kostya, for all the funny moments. That short stay was a life-saver for me, and all of you have my eternal gratitude.

A mis amigos de la clase de aikido, por tantísimos buenos momentos, por ayudarme a evadirme de los quehaceres diarios a base de caídas y golpes, y por hacerme descubrir una afición que respeto y que espero estar practicando todo el tiempo que mi cuerpo permita.

A los amigos del foro de Hype, porque uno se da cuenta que las distancias realmente no importan cuando las personas que hay detrás de la pantalla merecen la pena.

A José Manuel, comenzaste como un compañero y has terminado siendo un gran amigo. Sé que tendrás éxito allá donde vayas, gente con el interés, el buen hacer y la dedicación que tienes es extremadamente rara de encontrar. Mucho más improbable si sumamos tus valores como persona.

A esos amigos que siempre han estado ahí. Julián, por esos años llenos de deporte, conversaciones interesantes, y chivitos matutinos en el bar de los chinos. Álex, por todos los años vividos, tus canciones del móvil a las 5 y pico de la mañana, y tu demostración de crecimiento como persona. Alberto, mi divulgador favorito, gracias por meterme ideas raras en la cabeza que después han florecido en forma de conocimiento interesante sobre esa cosa rara que haces, con números y letras, y que vosotros decís entender. Julián (Conde!), gracias por todo el apoyo, cosas frikis, y momentos de risa en medio de deliciosas masas rellenas de carne. Mariano, no conoceré en mi vida una persona más buena onda que tú, inundas de alegría cualquier rincón con solo acercarte. Fernando, conocerte ha sido como hacerse amigo de un tsunami, un tornado y un terremoto a la vez, lo que te convierte en alguien inigualable. Darkos, se me hace difícil destacar solo una cosa en la que hayas influido en mi vida, dudo que pueda encontrar amigos como tú. Carmen, tú has sido una pieza fundamental en mi vida en estos últimos tiempos, y sé que sin ti hoy sería peor persona de lo que soy. No sois solo amigos, sois parte de mi familia.

A toda mi familia, pues mi persona depende en buena medida de quién son ellos y el apoyo que me han dado.

A mis padres y mi hermano. Si he sido capaz de levantarme después de cada caída, ha sido siempre gracias a vosotros. La educación que me habéis dado, el amor y la paciencia infinita, y el saber que siempre estáis ahí os hacen las personas más valiosas de mi vida. Hoy soy quién soy gracias a vosotros, y me siento inmensamente orgulloso de ello.

A ti, Mariana, porque eres la persona que da sentido a mi vida.

Declaration

Andrés Moya Simarro, catedrático del Departamento de Genética de la Facultad de Ciencias Biológicas de la Universidad de Valencia, y Carlos Peña Garay, director del Laboratorio Subterráneo de Canfranc,

INFORMAN QUE:

La presente memoria titulada ‘Stability of the human microbiome. Health and disease related to the time variability in the human microbiota’ ha sido realizada bajo su supervisión por el Licenciado en Biología, Daniel Martínez Martínez, en el Instituto de Biología de Sistemas Integrativa, y que reúne los requisitos necesarios para ser defendida como tesis doctoral y optar al grado de Doctor por la Universidad de Valencia. Y para que así conste a los efectos oportunos, se firma el presente en Paterna el 20 de julio de 2018.

Supervisor:

Prof. Andrés Moya

Supervisor:

Dr. Carlos P. Garay

Tutora:

Dra. Carolina Raussell

Daniel Martínez Martínez

September 2018

Table of contents

List of figures	xi
List of tables	xv
1 General Introduction	1
1.1 General description of human microbiota	1
1.1.1 Origins	1
1.1.2 Limits of life, the supremacy of the tiny biosphere	6
1.1.3 Who lives where	9
1.1.4 Variation across time and space	12
1.1.5 Health and Disease related to human microbiota	17
1.2 Dynamics of human microbiota	22
1.2.1 The evolution of microbiota	22
1.2.2 Measuring time-series	24
1.2.3 Community assembly	28
1.3 Interactions	31
1.3.1 What is an interaction	31
1.3.2 Inferring interactions in microbiomes	35
1.3.3 Time stability for complex networks	37
1.3.4 Function redundancy on microbiomes, and multi-stability	42
2 Objectives	47
3 Material and Methods	49
3.1 Cohort and sample collection	49
3.1.1 Sample selection in Chapter I	49
3.1.2 Sample collection in Chapter III	50
3.1.3 Sample collection in Chapter IV	52

3.2	16S rRNA gene sequencing	52
3.2.1	Nucleic acid extraction and sequencing	52
3.2.2	Sequence analysis	53
3.2.3	OTU characterisation	55
3.3	Metagenomic sequence analysis	57
3.4	Clinical measurements	57
3.5	Diversity analysis	59
3.5.1	A primer of diversity in ecology	59
3.5.2	α -diversity indexes and calculation	59
3.5.3	β -diversity and ordination plots	61
3.6	Biomarker discovery	62
3.7	Functional analysis	62
3.8	Correlations and interactions inference	63
3.9	Temporal analysis	63
3.9.1	X-weighted power-law fit	63
3.9.2	Standardisation	64
3.9.3	Langevin Model	65
3.9.4	Rank Stability Index	66
3.9.5	Temporal analysis of V and β	67
3.10	Lotka-Volterra model with an stochastic component	67
3.11	Graph inference and error measurement	68
4	Chapter I - Temporal stability of the human microbiome	71
4.1	Abstract	71
4.2	Material and Methods	72
4.2.1	Model	72
4.2.2	Cohort	73
4.2.3	16S rRNA sequences processing	73
4.2.4	Metagenomic sequences processing	73
4.2.5	Taxa level robustness	74
4.3	Results	74
4.3.1	Gut microbiota follows the Taylor's law	74
4.3.2	The phase space of gut microbiota	78
4.3.3	Rank stability of the taxa	80
4.3.4	Time dependence of model parameters	81
4.4	Chapter Summary	84

5	Chapter II - Microbial interactions as a matter of noise	87
5.1	Summary	87
5.2	Results	88
5.2.1	Species simulations	88
5.2.2	Noise-induced and pulse perturbation experiments	89
5.2.3	Graph Inference	91
5.3	Chapter Summary	95
6	Chapter III - Dynamical analysis of children’s microbiota with acute diarrhoea	97
6.1	Abstract	97
6.2	Material and Methods	98
6.2.1	Sample selection	98
6.2.2	16S rRNA gene sequencing and read processing	98
6.2.3	Diversity and functional analyses	99
6.2.4	Temporal analysis and stability	99
6.3	Results	100
6.3.1	Bacterial composition and diversity	100
6.3.2	Taxonomic and functional enrichment	103
6.3.3	Dynamic analysis of the microbiota	106
6.4	Chapter Summary	108
7	Chapter IV: Bacterial individuality in oxidative stress profiles in the oral cavity	111
7.1	Abstract	111
7.2	Material and Methods	112
7.2.1	Saliva samples and Oxidative stress markers	112
7.2.2	16S rRNA gene sequencing	112
7.2.3	Analysis of the bacterial composition	113
7.2.4	Temporal variations of the microbiome	113
7.2.5	Correlations between salivary markers and microbiota	114
7.2.6	Longitudinal sample collection vs. single sample collection	114
7.3	Results	114
7.3.1	Bacterial composition	114
7.3.2	Temporal stability of the microbiome	115
7.3.3	Large inter- and intra-individual difference of the oxidative stress markers levels	118
7.3.4	Correlations between bacterial taxa and oxidative stress markers are individual-specific	120

7.3.5	Correlations between the bacterial OTUs	122
7.3.6	Simulation of single sample collection	123
7.4	Chapter Summary	123
8	General Discussion	127
9	General Conclusions	139
10	Resumen en castellano	143
	References	149
	Appendix A Appendix	175
A.1	Chapter I - Supplementary information	175
A.1.1	Supplementary Tables	175
A.1.2	Supplementary Figures	181
A.2	Chapter II - Supplementary information	187
A.2.1	Supplementary Figures	187
A.3	Chapter III - Supplementary information	188
A.3.1	Supplementary Tables	188
A.3.2	Supplementary Figures	189
A.4	Chapter IV - Supplementary information	192
A.4.1	Supplementary Figures	192

List of figures

1.1	Concepts about the origin of life.	2
1.2	The phylogenetic tree of life.	5
1.3	Evolution of Earth's atmospheric oxygen content through time.	7
1.4	Enterotypes in the human gut microbiome.	11
1.5	Enriched taxa at various body parts and various populations around the world.	14
1.6	Population genetic structure of three common intestinal microbes.	15
1.7	Microbiota across age.	17
1.8	An illustrative case of pathogen colonisation after antibiotic treatment.	21
1.9	Co-speciation events in <i>Hominidae</i> gut microbiota.	24
1.10	Microbiome variation in time due to change in lifestyle.	27
1.11	Microbial assembly on <i>Hydras</i>	30
1.12	Metabolic exchange between microbes.	34
1.13	Different interaction networks inferred from samples.	37
1.14	Cooperation proportion and its influence in network stability.	39
1.15	Phase diagram of ranking stability.	42
1.16	Functional redundancy in microbial systems.	44
1.17	Multi-stability in microbial systems.	45
3.1	Paired Ends	54
4.1	X-weighted power-law fits of the standard deviations versus the mean values for each bacterial genus monitored over time.	76
4.2	Taylor's law parameter space.	77
4.3	Microbiota states can be placed in the phase space $F - V$	79
4.4	Rank variation over time for the 50 most dominant elements for a special period (days 72 to 122, travelling abroad).	81
4.5	Rank variation over time for the 50 most dominant elements for an ordinary period (days 123 to 256, after the trip).	82

4.6	V as a function of time for the two subjects in Caporaso's study.	83
4.7	V as a function of time for patient $P2$ in the IBS study and patient D in the antibiotics study.	84
5.1	Interaction matrix used in our simulations.	89
5.2	Relative abundances in log scale of the most abundant species as a function of time in our simulations.	90
5.3	Interaction and correlation matrices for the normalised counts.	92
5.4	Interaction matrices inferred from the regression model for the normalised counts.	94
5.5	Error between original and inferred interaction matrix.	95
6.1	Relative abundances of gut microbiota in every individual.	102
6.2	Shannon diversity index per health status and per day.	103
6.3	Comparison of microbiotas between healthy children and children with acute diarrhoea.	104
6.4	Taxonomic biomarkers and functions enriched.	105
6.5	Taylor's Law Parameter Space.	107
6.6	Rank Stability Matrix.	108
7.1	Temporal variability of the microbiome.	116
7.2	Rank matrix for the 50 most abundant OTUs for F11 and F18.	118
7.3	Correlations between the oxidative stress markers.	119
7.4	Marker-OTU correlations on the intra-individual level.	121
7.5	Marker-OTU correlations obtained by the simulation of collection of only one sample per volunteer.	124
A.1	Overview of the comparison of different approaches based on adjacent taxonomic levels using plots in the Taylor-parameters space.	182
A.2	Detail of comparison of different approaches based on adjacent taxonomic levels using plots of X-weighted power-law fits.	183
A.3	Rank variation over time for the 50 most dominant elements (taxa) and their calculated RSI (Rank Stability Index), Rank Variability (RV) and Differences Variability (DV), for an ordinary period (days 0 to 70, before the trip) belonging to subject A in the host lifestyle study.	184

A.4	Rank variation over time for the 50 most dominant elements (taxa) and their calculated RSI (Rank Stability Index), Rank Variability (RV) and Differences Variability (DV), for an ordinary period (days 257 to 364, further after the trip) belonging to subject <i>A</i> in the host lifestyle study.	185
A.5	Taylor's law parameter space for intervals concerning gut microbiota in the host lifestyle study.	186
A.6	Error between original and inferred interaction matrix.	187
A.7	Boxplots showing the Evenness index between both health status with all times together; and separated by time of sampling.	189
A.8	Canonical correspondence analysis of all individuals and times, separated by different child.	190
A.9	Relative abundances of the KEGG pathways inferred by PICRUSt at level 2 of the hierarchy.	191
A.10	Composition of the salivary microbiome of the 26 volunteers. Boxplots show the proportion of the eight most prevalent bacterial OTUs, volunteers are ordered by the average proportion.	192
A.11	Bacterial composition test for ordination of samples.	193
A.12	Comparison of NMDS ordinations of the 573 saliva samples including all 18,146 OTUs, OTUs with average proportion >0.001 % (687 OTUs) and with average proportion >0.1 % (50 OTUs). These three ordinations are not significantly different ($p < 0.001$).	194
A.13	Boxplots of Shannon diversity index and evenness index for each of the volunteers, ordered from the lowest to the highest according to the average value.	195
A.14	Temporal variability of 50 most prevalent OTUs in 6 volunteers. The volunteers F18 and F11 (the most and the less stable) are shown in the Figure 7.5	196
A.15	Boxplots showing values of AGE, AOPP, FRAS, TAC and TBARS for each volunteer, ordered from the lowest to the highest according to the average values.	197
A.16	Correlation plots summarising Pearson's and Spearman's correlations on intra-individual level.	198
A.17	Microbial interactions between the 15 most abundant OTUs per sample. . .	200

List of tables

1.1	Main interactions between organisms in an ecosystem. Extracted and modified from <i>Nature Knowledge Project</i>	32
3.1	List of the data sources of all the studies appeared in Chapter I	51
A.1	Taylor's parameters of individuals with either animal-based (A) or plant-based (P) diets.	176
A.2	Taylor's parameters for individuals taking antibiotics (Ab) in the antibiotics study, persons diagnosed with irritable bowel syndrome (IBS) in the IBS study and for special intervals concerning gut microbiota (HLS) in the host lifestyle study.	177
A.3	Taylor's parameters for the healthy subject (DH) and kwashiorkor part (DK) of the discordant twins.	178
A.4	Taylor's parameters for individuals with different degrees of overweight and obesity.	179
A.5	Rank and Rank Stability Index (RSI, as discussed in Material and Methods) over different periods for the taxa listed as <i>rank stability islands</i> regarding the gut microbiome of the subject A in the host lifestyle study.	180
A.6	Metadata with children's hospitalisation information.	188

General Introduction

1.1 General description of human microbiota

1.1.1 Origins

I believe a leaf of grass is no less than the journey-work
of the stars.
Leaves of Grass
Walt Whitman

Since its origin, around 4 billion years ago in the *Eoarchaeon* era [1], life has spread out freely on our planet, Earth. Life has grown in so many ways, shaped by the different forces of evolution, that nearly every habitat has been conquered by it. The origin of life might not have been an isolated and unlikely event as it was thought in the past. Rather, as Eric Smith says in the preface of his book *The origin and nature of life on Earth*: “A common theme was that life on Earth was not the outcome of an isolated event as suggested by the *Chance and Necessity* school but a **planetary property** that appeared early in the history of the planet and spread in a spontaneous way.” [2]. There are two principal forms to approach this problem, from biology-down, or from chemistry-up [3]. Origin of life depended on the physical-chemical conditions in the primitive Earth, so inorganic and organic chemistry, along with physics, played a key role at the beginning of it. For example, primitive chemistry was crucial in the synthesis of the primordial compounds that promoted the apparition of the complex chemistry that biological entities have in the present day. Although it is a controversial field, significant advances have been made in the last years to unravel those first steps, with hydrogen cyanide in a special position from which several compounds arise [3, 4]. Also, it seems that life might not have originated in the submarine vents but more probably in the surface due to the necessity of UV radiation for some chemical reactions [3, 5]. It is also unlikely that all the features that make something ‘alive’ were discovered

simultaneously. For this reason, an interesting idea is that life was in different degrees of ‘aliveness’ depending on the characteristics it possessed, as shown in Figure 1.1. That Figure also shows an intriguing concept, the chemical necessity versus chemical contingency. At first, it seems that life was not able to explore the vast space of chemical possibilities, and was constrained to have something similar to our system of proteins and RNA. However, at a certain degree of development, life was capable of exploring different solutions. This can be translated as that the very inner core of current biochemical reactions shares commonalities with the early chemistry (the *necessity*), but that the sequences of proteins or RNA that are responsible for these functions have different solutions in the enormous space of all the possibilities (the *contingency*). This compelling idea is not only present in the chemistry of life, but also in the capability of evolution (or evolvability) that genetic sequences show [6], something with profound implications in evolution as we will see later. Multiple steps afterwards, the protocells formed can start to operate autonomously, starting the ‘Darwinian’ machine to act upon them, and thus, giving start to the evolution of biological organisms as we know it.

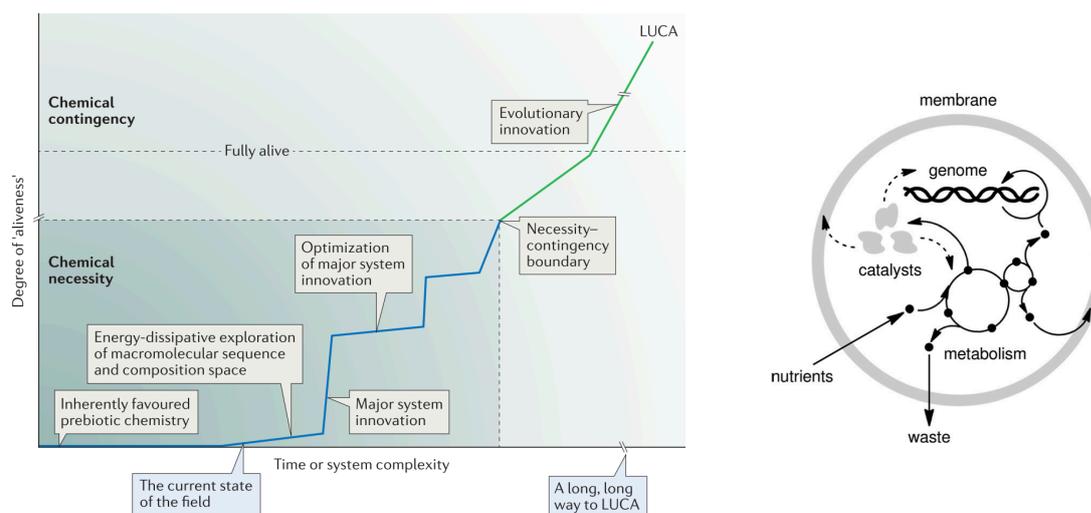


Fig. 1.1 On the left, a representation of the ‘aliveness’ versus time or complexity (from Sutherland, 2017 [3]). On the right, a schematic view of a protocell (from Sutherland, 2016 [4]).

Despite all the knowledge we have acquired about the origin of life, there are critical gaps that are hard to fill with data. The top-down approach has been extensively studied in the past decades. Several attempts have been made to see which are the minimum battery of

genes that a cell has to carry to survive, that is, the minimal genome [7–11]. This approach is related to the concept of LUCA, or the Last Universal Common Ancestor, which is the ancestor of every living being evolved afterwards. If one takes a look on the number of genes that are inferred from the studies mentioned before, it is hard to think that life had all those functions in the beginning. Thus, there is an immense gap between the protocells that we can infer now, with the model organism that arises from the minimal genome studies. What happened in-between, is hard to know. However, what seems clear at this moment is that all cellular types radiated from that common ancestor, resulting in the origin of the different **domains** of life [12].

The concept of domain had profound consequences to the evolutionary view because it changed the form in which we were classifying organisms. Biology, as a natural science field, has deep historical roots in naturalism, which was the dominant approach for the study of life for centuries. The descriptive nature that governed the examination of biological forms promoted the emergence of the first attempts of classification in Biology. The act of classification is inherent to the human nature, and, for centuries, scholars of different origin proposed their classification systems of biological entities. To be able to differentiate between some species, or between a plant and an animal is quite easy. However, to create higher orders of classification is a complicated task that requires some level of abstraction. A giant step was the creation of the notion of *genus*, made formally by the French botanist Joseph Pitton de Tournefort, around the 17th century. Later, Carl Linnaeus proposed the binomial nomenclature system that operates in the present days in Biology, made by two Latin names corresponding to the *genus* and the *species*, for example, *Canis lupus*. It is worth noting that the first time that the term “biology” appeared was in the book *Bibliotheca botanica*, by Carl Linnaeus.

Nevertheless, the dominant classification system in the past decades was the one based on the five Kingdoms proposed firstly by Herbert Copeland [13], and finally established by Robert Whittaker [14]. Briefly, the five Kingdom are: *Prokaryotae*; *Protoctista*, with algae or protozoa; *Animalia*, the animals we know like dogs or parrots; *Fungi* like mushrooms or molds; and *Plantae*, with cardiovascular plants, ferns and many other kind of organisms. A complete and beautiful book about this topic is the one written by Lynn Margulis, *Kingdom and Domains* [15]. However, this organisation is almost in complete disuse due to the concept of domain, mentioned before. This concept, based on the works of Carl Woese [12], proposed a complete revision of the five-Kingdom system in favour of a three-domain system, based on molecular evidence. He proposed the use of universal gene markers, the 16S and 18S rRNA gene, to see the evolutionary patterns of life, and thus, he proposed that there exist another domain (or super-kingdom) more than *Prokaryotae* and *Eukaryota*, being this third

group the *Archaea*. I will use Woese's system throughout this thesis because it is the most accepted system at this moment.

Independently of the system one chooses, the fact is that the prokaryotic world was here before the *eukaryotes*. The structural, molecular and genetic differences between the two worlds are evident, and the emergence of eukaryotic cells is considered one of the most important evolutionary innovations in life's history [17]. It is known that we, eukaryotes, evolved from the *evolutionary affaire* between an archaea and a bacteria that belonged to the phylum alphaproteobacterium [18]. Although recent discoveries tells that the bacteria belonged to an ancestor of the alphaproteobacteria group [19]. As can be observed in Figure 1.2, the domain of eukaryotes originated from the branch of *Archaea*, and this event might have its origin in the recently discovered *Asgard superphylum* [20]. This idea of endosymbiosis as a source of evolutionary innovations, so settled now but controversial in its conception nonetheless, came by the hand of Lynn Margulis. Her broad ecological knowledge allowed her to suggest a revolutionary idea for the origin of eukaryotic cells as a serial process of endosymbiosis, in which the bacterium was engulfed by the archaea, and became the mitochondria as we know it. It was an evolutionary innovation so important that no eukaryotic cell lack mitochondria. Without it, we never could have happened. This process was further demonstrated as the one that also originated the plastids in eukaryotes [21], and also has been reported to exist secondary and tertiary endosymbiosis in dinoflagellates [22]. In fact, secondary and tertiary processes of endosymbiosis are more common than previously thought, a fact that has been made more and more evident with the discovery of endosymbionts 'living' inside eukaryotic cells. For example, we can find the well studied case of *Buchnera aphidicola*, that inhabits several species of aphids [23, 24]; or the case of *Blattabacterium cuenoti*, that inhabits the gut of the cockroach *Blattella germanica* [25]. In the end, an enormous complexity was originated from these events, but, as we will see, bacteria are still in the top of the mountain.

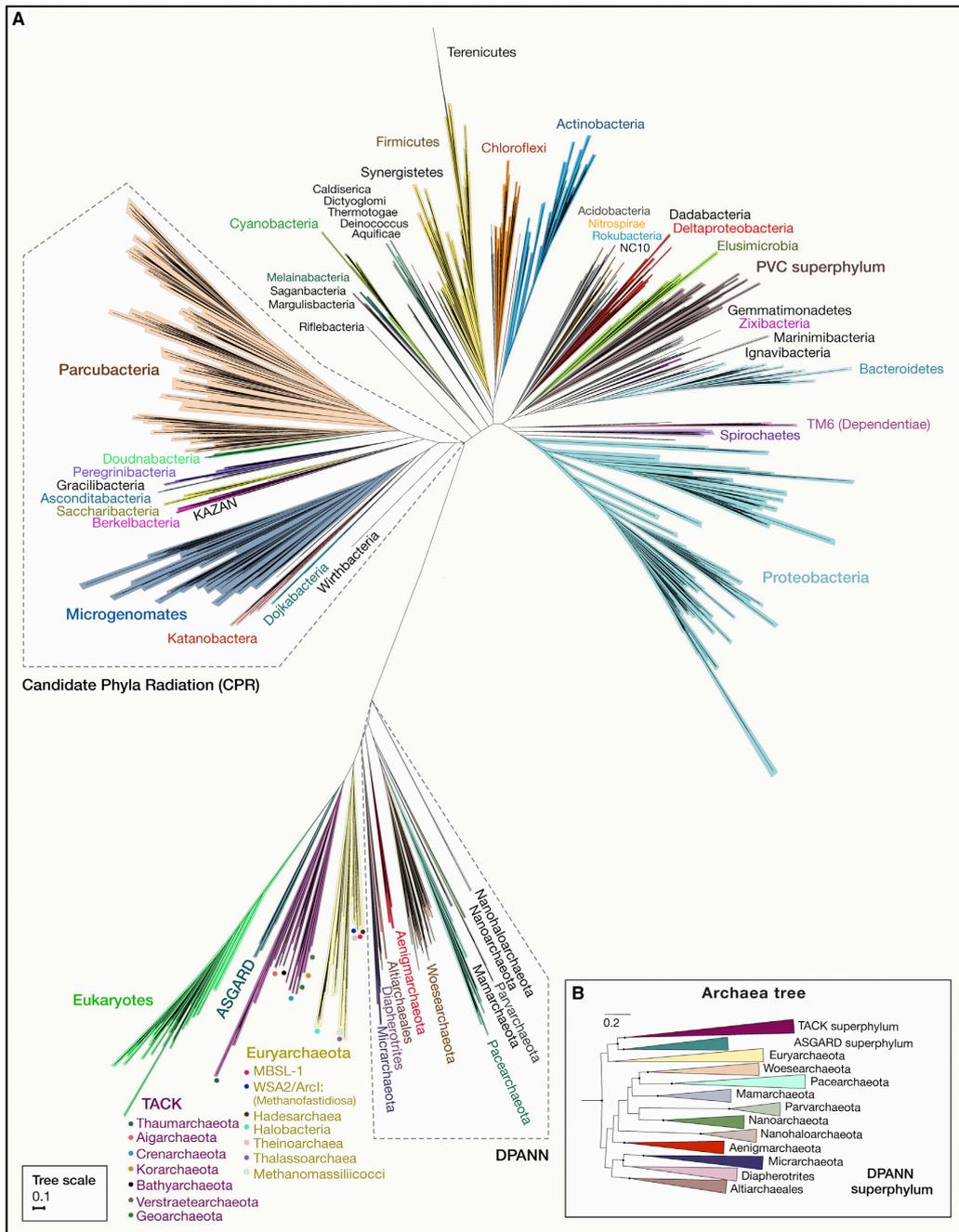


Fig. 1.2 The phylogenetic tree of life. It has a three-domain topology, with the upper part of the tree corresponding to the vast diversity of the bacteria known together with the candidate phyla radiation (CPR). In the lower part, we have the archaea, with the new groups discovered recently (Asgard and DPANN), and the eukaryotes emerging from this branch as the third domain. Part B is the phylogenetic tree of archaeal sequences. From Castle and Banfield, 2018 [16].

1.1.2 Limits of life, the supremacy of the tiny biosphere

...how many seconds of time are there in eternity, asked the King.
The shepherd boy said: 'In lower Pomerania is this diamond mountain,
every hundred years a little bird comes and sharpens its beak on it, and
when the whole mountain is worn away by this,
then the first second of eternity will be over.'

The Shepherd boy
The Grimm brothers

Bacteria and archaea have been, are, and will be, a true force of nature that has shaped this planet in a myriad of forms. The emergence of life brought the biosphere into existence as a fourth geosphere, establishing new forms in which matter and energy were organised [2]. Once they were settled, these little inhabitants have worked as unconscious tinkerers, working unceasingly and untiringly, transforming how Earth has been, is, and will be. The fact that life needs to be in disequilibrium to exist has brought up profound changes in the chemical and physical composition of this planet through history. Even though the core of metabolic functions in bacteria is between the most stable properties of matter [2], the metabolic capabilities that they have been able to search are enormous, and some of these pathways have played a vital role in the evolution of the other three geospheres [26, 27]. Yes, they are tiny, and their individual action is almost negligible, but they have been a multitude and *had all the time in the world*. The comprehension of their capabilities leads us to humbly say that we, humans, are invited organisms living in a world that belongs to the microorganisms. In the more-probable-than-you-think event in which humanity finds its end, this world will still belong to these little beings.

The most well known critical transformation that microorganisms have made to the Earth is, undoubtedly, the **rise of oxygen levels** on a planetary scale. This process, named as the Great Oxygenation Event (GOE), has been extensively studied for more than a half of a century, and albeit of the incompleteness of this puzzle, there is a solid theory about what happened. Earth's atmosphere has changed in a significant degree across time, and despite previous thoughts that it was reducing (as Miller thought in his experiments [28]), it seems that there could have been more oxidised compounds in the primary stages of this planet. Nevertheless, what seems pretty robust is that free oxygen levels, during the first half of Earth's life, were less than 0.001% of those present in the atmosphere today. However, at some point around 2.1-2.4 billion years (Gyr) ago, oxygen levels rose (see Figure 1.3). We know that thanks to the prints in the rock record, for example, with the first appearance of rusty red soils and the disappearance of easily oxidised minerals as pyrite (FeS₂) from ancient

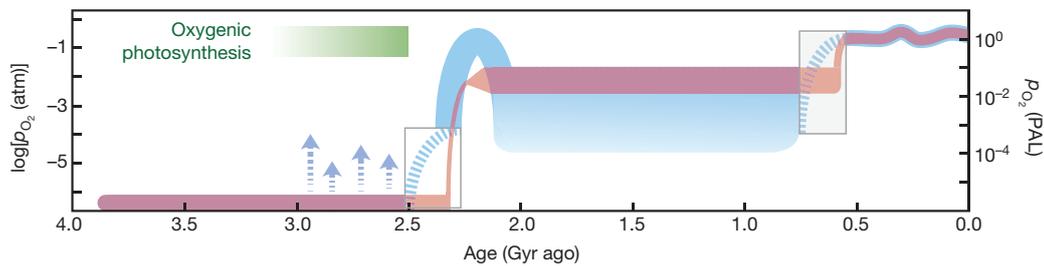


Fig. 1.3 Evolution of Earth's atmospheric oxygen content through time. In red, the classical view of two-step process; in blue, the modern view. Right axis is the relative to the present atmospheric level (PAL) of O_2 , and the left is the $\log p_{O_2}$. From Lyons *et al.*, 2014 [29]

beds [29]. Contrary to previously thought, it seems that this process of oxygenation was not a discrete step but a continuous procedure. It is thought that oxygen level rose substantially in the start of GOE event, but declined soon to reach, again, low oxygen levels until 800 million years (Myr) ago. Also, it seems that there were some spots in the Earth where oxygen was being *produced* 500 million years before the GOE start, meaning that it was present much before than previously thought. It is represented with the blue arrows that emerge upwards before the GOE in 1.3. The key word here is *produced*, because practically all the oxygen that was originated on Earth was due to the action of microorganisms with one of the greatest inventions of the evolution: the **oxygenic photosynthesis**.

We have talked about the origin of photosynthesis before, as an example of evolutionary innovation through endosymbiosis process. The rise of oxygen in our planet was a direct consequence of the invention of the oxygenic photosynthesis by specific groups of *Cyanobacteria*. It seems that the abiotic origin of oxygen was insufficient to develop the aerobic respiration of late organisms [29–31]. It is worth noting that photosynthesis arose early in Earth not in the oxygenic mode but with other electron sources as hydrogen, ferrous iron or hydrogen sulfide instead of water. The use of water as electron source implies that those electrons can be used to reduce inorganic carbon (from CO_2) in order to build organic compounds, forming O_2 as a byproduct. Then, this stored redox energy can be released by oxidising organic molecules with O_2 and form water in the end [30], a process called *respiration*. The group of organisms that use photosynthesis but do not make their organic carbon are called photoheterotrophs, and is composed of purple bacteria belonging to *alpha* and *betaproteobacteria*, or green non-sulfur bacteria belonging to *Chloroflexi* phylum. In fact, only the class *Oxyphotobacteria* in *Cyanobacteria* is able to carry out the oxygenic photosynthesis, and it seems that it was a late acquisition of this particular group instead

of a general property of the whole *Cyanobacteria*. Interestingly, it has been shown that *Oxyphotobacteria*, along with the other two cyanobacterial classes *Melainobacteria* and *Sericytochromatia* developed individualised strategies for the aerobic respiration, adding support to the hypothesis that photosynthesis arose as a particular function in *Oxyphotobacteria* [31]. An important implication of the generation of oxygen is that it was toxic for most of the organisms living at that time. Life had to adapt to this new environment to survive, although a good part of microorganisms that were isolated in spaces where the oxygen did not reach remained to be anaerobic. However, despite the fact that the first rise in the level of oxygen did not yield the levels that we have today (see Figure 1.3), it was fundamental to the evolution of eukaryotes, complex forms of life that required an unusual quantity of energy. That energy was provided with the creation of the mitochondria, which further habilitated the apparition of multicellularity. Moreover, the second rise in oxygen level, the one that ended up in the levels that we have today, was performed by the growth of plants (a direct product of the first oxygen rise), that ‘coincided’ with the Cambrian radiation of animals [32]. A curious chain of events that have brought us here.

Microorganisms not only have the power to transform the world, but they also have the ability to live on it. On any place one can think. The limits of life are broader than we can imagine, and the range of temperatures, salinity, pH or radiation that can resist, or even need, is extensive. We find areas like Río Tinto (Huelva, Spain) where the conditions for life are harsh due to the acid conditions, but where we find a rich biosphere that grows there without any problem at all. A couple of examples that we can find in Río Tinto are *Acidithiobacillus ferrooxidans* or *Leptospirillum ferrooxidans*, which grow at a pH of 1-2 [33]. However, the prize to the acidophilic organism goes to the archaea *Picrophilus oshimae*, which grows in a pH of -0.2 and was found nearby to a hydrothermal spring with solfataric gases in Japan [34]. Regarding temperature, we can find exceptional examples of organisms that have their optimal growth at more than 80°C, called hyperthermophilic organisms. Several groups of archaea as *Crenarchaeota* or *Nanoarchaeota* phyla, or the order *Thermotogales* in bacteria, are part of this interesting group, and the molecular machinery they carry is thermostable and also high stable against detergents, organic solvents, and some other chemical reagents [35]. As a last example, we can find organisms with extreme resistance to radiation, as *Deinococcus radiodurans* and *Thermococcus gammatolerans*. *D. radiodurans* can tolerate radiations of 5000 Gy without any loss of viability, and *T. gammatolerans* can resist radiations up to 30000 Gy. If one thinks that radiations of 5 Gy are enough to kill a human, or with radiations of 60 Gy we can kill an entire colony of *Escherichia coli*, we can begin to understand the magnitude of this resistance. The usual mechanism that allows these microorganisms to live under high radiation levels is an exceptional ability to repair damaged DNA without loss of

viability. It is worth mentioning that life can not exist without a certain level of radiation [36].

After having seen the impressive capabilities of the tiny biosphere, it is not strange at all that a multitude of bacteria, archaea, and other kinds of microorganisms cohabit with the more complex multicellular organisms. After all, we arrived later. We are surrounded by bacteria mostly, and they have shaped our evolution in the same way that we have domesticated them. The term **microbiota** refers to the community of these little forms of life that share their destiny with the higher organisms to which they are linked. The microbiota has profound effects on the living systems, affecting nutrition, pathogen resistance, immune development and even the behaviour of the host [37–40]. Furthermore, the ways in which the microbial community is assembled in the host are far to be well known, and the dynamics of these complex communities are starting to be understood. However, before we can start making assumptions about the effects of the microbiota in the hosts, we need to know the high diversity of them.

1.1.3 Who lives where

(...) we have tried to go out too soon to discover
and colonise outer space,
when there is still much of our inner space
to discover and understand.

Solaris

Stanislaw Lem

An immense number of microbes have been unable to be cultured in the laboratory for many years, as they usually have strict growth conditions that are hard to replicate, and thus, hindering the exploration of the high diversity of microorganisms. However, the development of high throughput sequencing techniques allowed us to start exploring the unseen biosphere with higher precision and almost unlimited power. Since Sanger published his methodology in 1977 [41], these technologies have been improved in many ways to permit the massive sequencing and the discovery of a hidden diversity that was unreachable before [42]. The improvement in the sequencing produced an explosion of new statistic and bioinformatic methods to analyse the new data created. A glimpse can be seen in the methods that have been used through this thesis.

Thanks to the development of the technology, we have been able to study the microbiota in great detail. As the *naturalist* historical root is still present in our minds as biologists, the first stage of the field was mostly the description of what is there, and where are they.

This step is crucial to start knowing a system as complex as the human microbiome, which can be divided into several main sub-ecosystems depending on the physical region one is studying: *gastrointestinal*, *urogenital*, *skin*, *oral*, *nasal*, or the *respiratory tract*. The microbial composition of each of these regions differ by some degree, and we also can find subregions that are different between them even corresponding to the same main region, as can be the hands' skin and other skin parts less exposed to direct contact with the outside world [43]. Another important matter that should be taken into account is that human microbiota varies with age, geography, disease, diets, drugs, and an endless list of possible factors that can disturb this ecosystem. Even with all these sources of variation, we can extract some interesting general properties of the human microbiome composition.

We can be sure that we have a good representation of the human microbiome composition at certain higher levels as genus, but new diversity is being discovered in terms of species and strains with the analysis of whole genome shotgun data [44, 45]. The most studied system of the aforementioned is the *gut microbiome*, which has been described in a great deal for its implications in several diseases as the irritable bowel syndrome (IBS). In the gut we have three basic layers: the *lamina propria*, the mucous membrane that corresponds to the connective tissue; the *epithelial cells* that are in direct contact with the third part or *intestinal lumen*, where is located the microbiota and all the chemical components of nutrients, toxins or water. It has been observed that the gut microbiome is mainly composed by bacteria (>97%), some archaea (around 2%), viruses (about 0.2%) and eukaryotes as fungi (<0.01%) [46]. The most abundant phyla in the gut microbiome usually are *Firmicutes*, *Actinobacteria* and *Bacteroidetes*, although the proportions of these groups can vary between geographical places and other conditions [46, 47]. The three phyla dominates the majority of the gut microbiome, and we can find some interesting genus as *Bacteroides*, *Prevotella*, *Ruminococcaceae*, *Alistipes*, *Faecalibacterium*, *Roseburia* or *Blautia* to name some examples. The estimations are that we harbour more than 200 different genera of microbes in our gut, and around 600-700 species (see Figure 1.4) [46, 48], and that one strain typically dominates each species [49]. Some years ago, Arumugam *et al.*, proposed that there existed three major groups of human the gut microbiomes, called enterotypes, which were centred around a different genus: *Bacteroides*, *Blautia* or *Prevotella* [50]. Although an attractive idea, enterotypes are not a clear separation between different microbiome modes, and it seems that there are many intermediate steps between the three main groups [48]. However, as we will see in further parts of this thesis, the concept of enterotype entrails interesting ideas regarding the multiple states in which the gut microbiota can be present. Nevertheless, one essential issue (that we will also develop after) is that, even if the taxonomical composition of the

gut microbiome changes between persons or places, the functions encoded in the genomes remain highly stable [43].

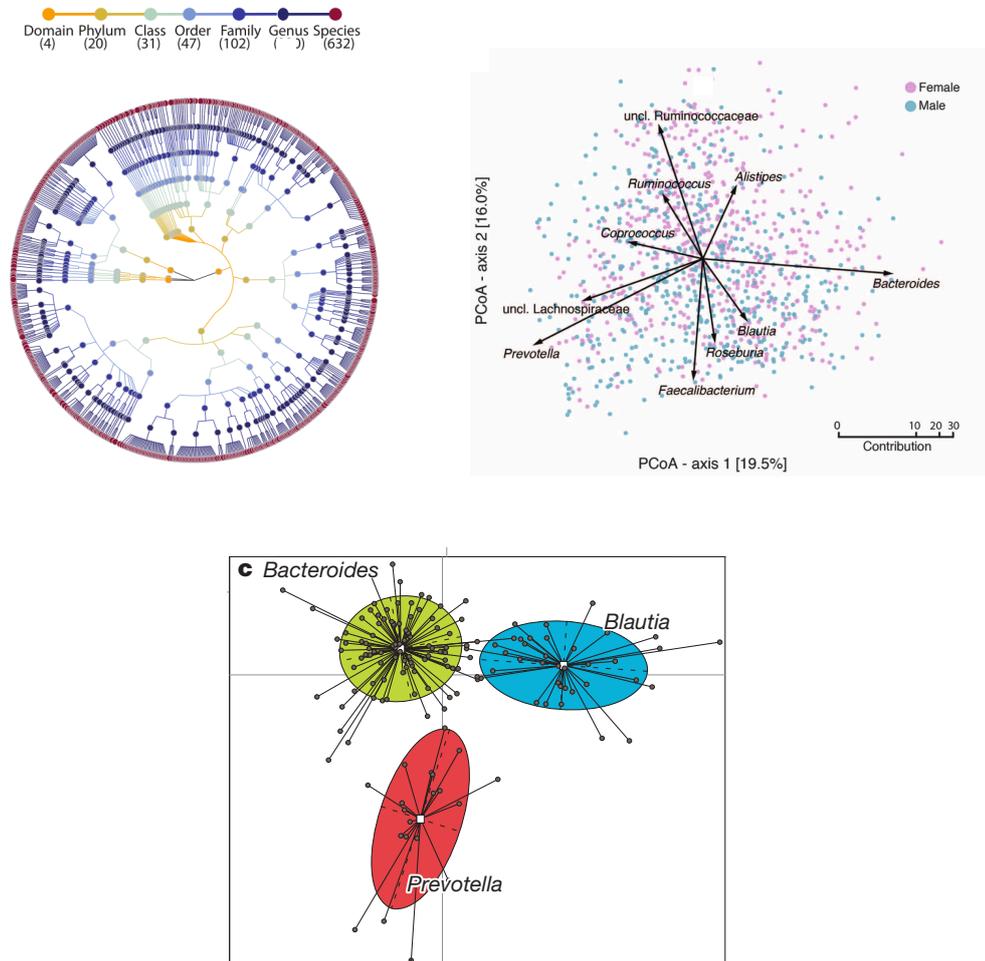


Fig. 1.4 On the upper left, the estimated number of different taxonomical levels measured in a cohort of 1135 humans. From Zhernakova *et al.* [46]. On the upper right, the top 10 contributors to community variation in a Principal Components Analysis (PCoA) with Bray-Curtis dissimilarities, from Falony *et al.* [48]. On the below line, the three different enterotypes defined by Arumugam *et al.*, in a PCoA with distances based on the Kullback-Leibler discrimination information [50].

Microbes also live in more harsh environments as is our *skin*, which is considered the larger organ of the human body. As large as it is, we find numerous different habitats where the microbes can grow that can be categorised by their physiological characteristics: sebaceous, moist or dry [51]. The skin is composed of two layers, the epidermis and the

dermis. The skin also has components as sebaceous glands, hair follicles and sweat glands that can alter the local conditions, and also harbour a great bunch of microbes. All these conditions shape the microbiota associated with the different habitats. The three most abundant and important groups of bacteria, the dominant domain in the skin microbiome, depend on the type of habitat: in sebaceous sites, *Propionibacterium* spp. dominates thanks to their ability to live in anaerobic and acidic conditions, and the humid places are dominated by *Staphylococcus* spp. and *Corynebacterium* spp. members. Among the specialisations that these microbial groups have developed, we find the ability to use the sebaceous contents of the skin or strategies for surviving at high salt concentrations [52, 53].

We harbour a great diversity of microbes in the mouth, mainly because it is a warm and moist environment where the host also offers some nutrients, such as saliva proteins, glycoproteins, and gingival crevicular fluid [54]. Although the mouth is smaller than the gut or the skin, we find several different ecosystems as teeth, gingival sulcus, tongue, cheek, lip and the palate to name some. The mouth works as a connection between our inner world and the outside, and contrary to one could think, the oral microbiota remains highly stable in time if no other perturbations are present [55, 56]. In fact, contrary to what was thought, the oral microbiome is a highly diverse and rich environment populated by hundreds of different taxa such as *Rothia*, *Aggregatibacter*, *Streptococcus*, *Haemophilus* or *Actinomyces* [43, 57].

As the reader should have noticed in this brief look of the human microbiota composition, we are populated with a great diversity of microbes, without taking into account the *fourth domain* composed of viruses, which is highly diverse and complex. However, how is this diversity translated into a raw number of things that we have in our body? The usual claim that has been done in research literature was that we had 10 microbes per eukaryotic cell in our body, it is, an order of magnitude higher. About two or three kg in an ‘average person’ weighting 70 kg. Well, it seems that the number was rather inflated, and recent estimations put the number of microbes in a ratio between 2:1 and 1:1 with respect to eukaryotic cells [58]. We are talking about $3.0 \cdot 10^{13}$ bacteria. More extraordinary is the number of different functions that they perform, which seems far from saturating in the number of gene families discovered.

1.1.4 Variation across time and space

The previous section was intended to be a general view of the highly diverse human microbiota, an observation from above of a very complex system. However, in order to disentangle its subtleties, we should study the principal sources of variation, and what are the components that shape the community. For this purpose, we are going to see the variations of the microbiota in the four dimensions: the spatial and the temporal component, for they are

critical elements that entrails a good part of the shift concerning population. Time and space are general assumptions of a significant number of minor components that are associated with them. For example, we know that the microbial composition changes with geography, but that is a consequence of the change in diets, health systems, and surrounding animals. Also, the temporal component has many minor factors as the way in which the child is born, the immune system maturation, or the change in habits. In the end, both main components are intimately linked, as your lifestyle can change when you move to another place [59]. In the next lines, we will try to make a brief explanation of the significant effects that these components have on the human microbiota.

Focusing on the *geographical component*, we have seen that there are apparent differences in the microbial composition of different populations around the World. As can be seen in Figure 1.5, we find differently enriched taxa in the world, at different body parts [60]. In an article from Yatsunenکو *et al.*, they sequenced the V4 region of the 16S rRNA gene and, using a PCoA analysis with UniFrac distances, they showed that there were a clear separation between US, Malawian and Amerindian gut communities [61]. In a more in-depth study, Truong *et al.* analysed the microbial communities of more than 1500 gut metagenomes at strain level, and found interesting properties about their geographical distribution [49]. Analysing the strain genetic structure from the most abundant species, they observed that some of them have a continuous distribution, as *Faecalibacterium prausnitzii*, while other strains seem to have a discrete distribution, as *Eubacterium rectale* (see Figure 1.6 for more details). These types of distribution may account for the way in which some bacteria are spread in human populations, meaning that some species might have more ability to spread both vertically and horizontally, and others could only be transmitted vertically. Interestingly, there seems to be a highly personalised oral microbial composition associated with ethnicity, a property that allows identifying people's ethnicity with machine-learning classifiers [62]. It could happen that, with the required degree of resolution, we will also be recognised by our microbiota with the same accuracy that our genomic markers could identify us.

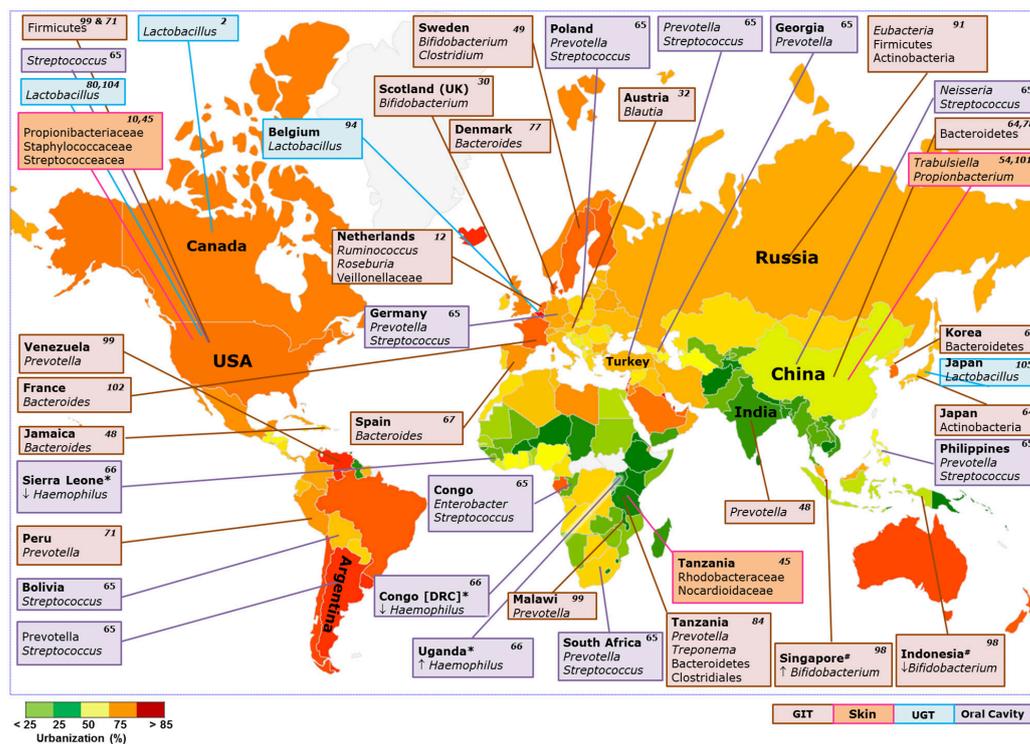


Fig. 1.5 Enriched taxa at various body parts and various populations around the world. The square colour represents the body part, and the map colour represents the level of urbanisation according to Unicef (visited in 2016). From Gupta *et al.* [60].

Geography is also interesting because it sheds light on human evolution because there are differently developed populations across the world such as a) hunter-gatherers population as the Hazda from Tanzania, or the Pygmies from Central Africa; b) traditional farming as the Bantus of Africa; and c) high developed countries from the western world in Europe or North America. It is interesting to observe that microbial diversity *in gut* decreases from the hunter-gatherers to the developed countries [63, 64], an effect that lacks a robust explanation but can be interpreted as the loss of metabolic plasticity and defence capacities for living in a more controlled world [65]. One fascinating thing is that it seems that there is a *Prevotella*-*Bacteroides* trade-off between some populations that could explain, in some degree, the existence of the three enterotypes mentioned above [61]. Which is more interesting is the fact that this trade-off is one of the most remarkable differences between hunter-gatherers populations and western countries, usually more developed. Western countries have a higher abundance of *Bacteroides* while the hunter-gatherers have a higher degree of *Prevotella*, a difference that could be due to the diet specialisation in the more advanced communities [60, 63]. Also, the role of human genetics in shaping the microbiome is poorly understood,

and it seems that the main effect on gut microbial composition is related to external agents as the lifestyle. An example of this is that the Bantu of Africa can form two types of communities, based on fishing or based on farming, and even though they share the same genetic ancestry and ethnicity, their gut microbiotas are quite different [65].

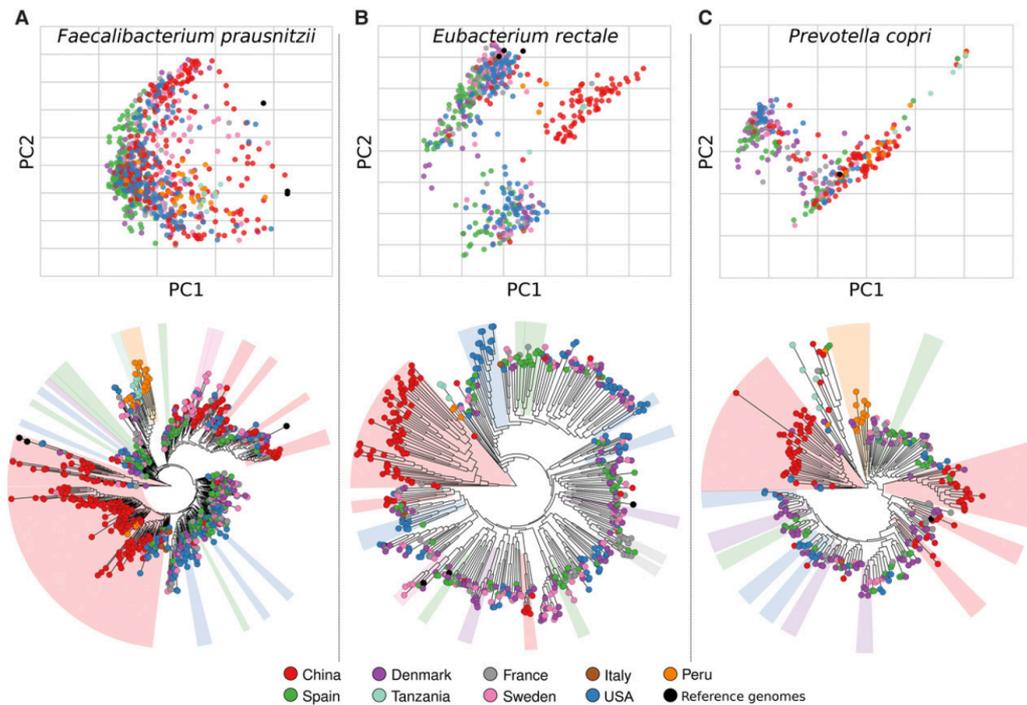


Fig. 1.6 Population genetic structure of three common intestinal microbes: *Faecalibacterium prausnitzii*, *Eubacterium rectale* and *Prevotella copri*. On the upper row, it is shown the Principal Component Analysis of the strain markers of those species studied in relation to the geographical source to see the possible clusters. On the lower row, it is represented the phylogenetic trees of the marker genes. From Truong *et al.* [49]

The other essential direction of variation in microbiota is *time*, it is, the age of subjects. In fact, and as we will see in later parts of this introduction, the time dimension opens up the ability to study microbial dynamics, which is one of the main points of this thesis work. Nevertheless, we will see now a different time scale, the one that affects human populations through their entire lives. The usual dynamics from this point of view are that children microbiota is unstable, and it evolves fast toward a more stable state in the first three to five years. Then, and until the elderly, it seems that microbiome is highly stable through time, becoming more variable in the last part of human life [61, 66, 67]. Those changes do affect not only microbial diversity but also the microbial community (see Figure 1.7, upper row).

How the microbiota is assembled into newborns is a subject of high interest and importance, and although it seems that the form of delivery (cesarean or vaginal delivery) matters [68], it is still in discussion, as it is the place of birth (home or hospital) [48].

To know how this assembly occurs is essential for the understanding of other properties of the system. The microbial transitions in the first years of life are well studied, and we have enough data to form a general view of the major changes [67]. After birth, gut microbiome of a newborn is transiently enriched with *Enterobacteriaceae* and *Staphylococcus*, a community that lasts a brief period until the establishment of a *Bifidobacterium* enriched population, due to the type of food ingestion [69]. Once the child starts eating solid food, the microbiota continuously evolves toward an adult-like conformation with a dominance of *Bacteroidetes* and *Firmicutes*. Then, at the age of three to five years old, the child reaches the typical adult stability [61], with a dominance of *Prevotella* or *Bacteroides* depending on the type of foods, as we have seen in this section [67]. The period of the first years of life is critical in the correct development in children. It was observed that intestinal microbiota helps to the proliferation and differentiation of some T and B cells in the immune system [70]. In the opposite side of the spectrum, it seems that the microbiota in older adults changes from the adult-like composition into an elderly-type [66]. In Figure 1.7, on the lower row of images, we can see that people of the same age tend to cluster together, adding evidence for the change in microbial composition in older ages.

The main point of this section is clear: human microbiota is a highly dynamical system that never ceases to change. It can be affected by a great deal from several types of elements, from diets, animals that live nearby, antibiotics, age, immune system maturation and an endless list which is growing every day. Moreover, as it happened to this planet, where microbes shaped its geochemical cycles, they also affect the human body in a great deal. In the next section, we will see that some of these aforementioned changes have profound impacts on the human health.

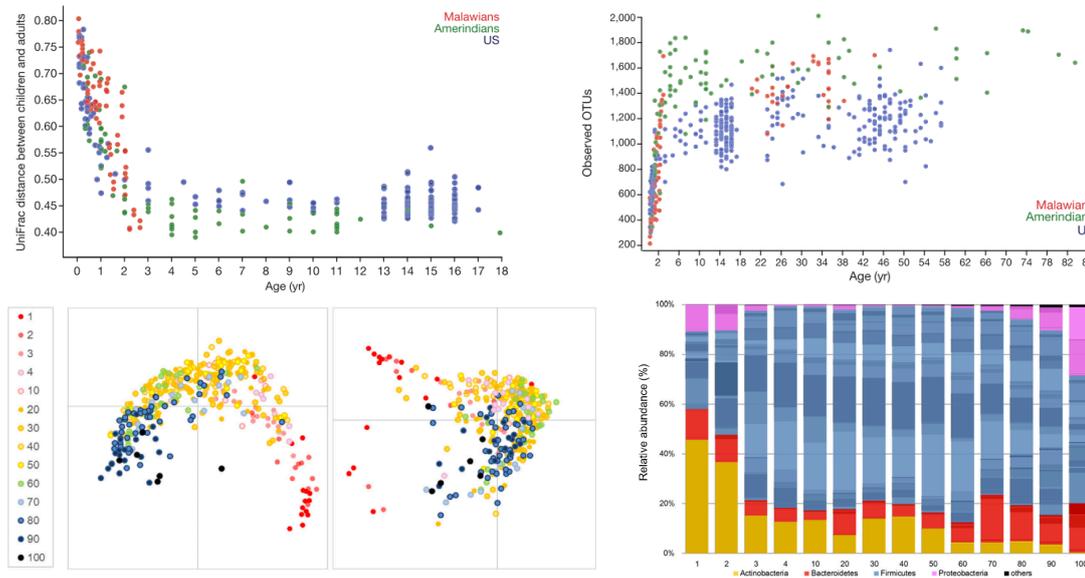


Fig. 1.7 Microbiota through age. On the upper left, it is represented the UniFrac distances between children and adults in three different populations through time. On the upper right, it is represented the observed number of OTUs in the same populations through time. From Yatsunenکو *et al.* [61]. Left below it is represented a PCoA of UniFrac distances (left, unweighted; right, weighted) between groups of people clustered by age. Right below, the average microbial composition through age in the same groups. From Odamaki *et al.* [66].

1.1.5 Health and Disease related to human microbiota

Better the huntsman
 Should perish in the swamps,
 For in the Dead Mountains
 There is a deep, deep chasm,
 Where dwelleth Ygramul the Many,
 The horror of horrors.
The Neverending Story
 Michael Ende

One of the leading forces in the study of the human-associated microbiota is the search of the agents that cause diseases on us. In 1890, Robert Koch published his postulates changing the paradigm in medicine. From that moment, it was accepted that microorganisms caused some diseases, and he provided a general scheme to prove whether the candidate bug was

the responsible for the ailment or not. However, with the apparition of the studies in human microbiota, the paradigm shifted toward a more complex point of view of the aetiology agents. The Koch's postulates are still valid for some limited cases as cholera, malaria or tuberculosis, but the steadily growing knowledge about the complex ecosystem of microorganisms we carry has changed our point of view. Think about it, we live with approximately 10^{13} microorganisms in our body, they are living and doing their functions, and we deal with them. It seems rather clear that, in the case that certain conditions are matched, we are going to suffer the consequences. In other words, if the microbial community changes, we might suffer particular diseases as a cause of this disequilibrium, also named as *dysbiosis*. Also, if there is a disequilibrium, that means that we have a 'healthy composition', a concept that has been broadly studied and currently discussed. What is the healthy composition of microbiota? What it had be described in the previous sections could be considered as a healthy microbiota, but as it has been noted, this composition changes in a great deal with forces as geography or age. Moreover, as some authors agree, it might be that there is not only one health condition because different microbial communities could be possible, an idea that has deep roots on functional redundancy [71] and multi-stability in ecology [72].

However, in a case of dysbiosis, it is often hard to assess whether the imbalance has provoked the disease or is the abnormal condition that induces the rupture of the healthy microbial community. Besides, if it happens to be the second possibility, the dysbiosis might help to worsen the symptoms in the subjects, and creates a positive feedback loop. It has been observed that gut microbiota is related to a significant number of well-defined diseases such as irritable bowel syndrome [73], Crohn's disease [74], obesity [37, 75], colorectal cancer [76], or HIV [77, 78] to cite only a very few number of examples. In other cases, it will happen that a change in the community will occur as a consequence of other factors as antibiotics, diets or change in lifestyle that will, in the end, make the subject more prone to infections as it is the broadly studied case of *Clostridium difficile* [79], a pathogen that needs a compromised microbiota in order to succeed. Some of these cases are going to be described with some detail in this section, to prove the tightness between our microbial world and us.

The irritable bowel syndrome (IBS) is the most common functional gastrointestinal disorder in Western Countries, and it is a condition for which researchers have been looking for its microbial origin with more or less success. It is a disease with a complicated symptoms onset, heterogeneous both in clinical presentation and in aetiology and pathogenesis [73]. IBS can appear in many grades, and based on many causes, so a gut-microbiota-centred hypothesis is unlikely to apply to all patients to IBS [80]. Nevertheless, a growing number of studies state that IBS is related to with an altered diversity, stability and metabolic functions compared to healthy people. The main differences are the increased abundance of *Firmicutes*,

especially of *Ruminococcaceae* spp. and *Clostridium* cluster XIVa, with the reduction of *Bacteroides* [81]. The low-grade mucosal inflammation in IBS might be due to the imbalance of bacteria as *Clostridium* cluster XIVa or members from *Lachnospiraceae* family, that express flagellin proteins that can help to cause it [82]. Moreover, it seems that altered levels of the archaea *Methanobrevibacter smithii* might be important in intestinal transit [83]. **Crohn's disease** (CD) is just one of the two main forms that can take the irritable bowel diseases (IBD). It has been observed that *Faecalibacterium prausnitzii* and *Escherichia coli*, in particular, were decreased in patients suffering from CD [84]. The other major subtype of IBD is the **ulcerative colitis** (UC), and to differentiate between them is a non-trivial issue. It seems that dysbiosis is greater in CD than in UC, and that some microbial groups as *Faecalibacterium*, *Anaerostipes* or *Collinsela* could help to distinguish between CD and non-CD patients [84].

Regarding **obesity**, it is a public health problem that affects around 2 billion adults, in which a 39% of them are in the range of overweight. We can define obesity as the excess of body fat, and it is related to the energy harbour mechanisms that are encoded in the microbiota along to several other factors such as host genetics and environment [85]. However, even with the complexity of this affection with many such factors operating, it seems that there are general patterns of microbial differences between lean and obese people, with some essential functions or metabolic pathways implicated. At phylum level, there is a difference in abundance between *Bacteroides* and *Firmicutes* abundances, being the later increased both in obesity and type 2 diabetes. This imbalance has been related to the production of short chain fatty acids (SCFAs), in which butyrate is critical due to its central functionality as gut hormone [86]. Butyrate has protective effects against diet-induced obesity and insulin resistance. Other differences were in the expression levels of adenosine monophosphate-activated protein kinase (AMPK), which acts to regulate energy via fatty acids oxidation, ketogenesis, glucose uptake and insulin secretion. It has been observed that microbiotas in obese people could deplete the activation of AMPK, and also the NAD⁺ levels are also lower in these microbiotas, which also upregulates AMPK [87].

One of the cases in which it is not already known whether the microbiota affects the host, or is the host disease that alters the microbiota, is related to **acute diarrhoea** in kids, one of the leading causes of morbidity, hospitalisation, and mortality worldwide [88]. The rotavirus genus is the most common cause of gastroenteritis in children, and recent studies have shown that infections can cause changes in the composition of intestinal microbiota [89]. The rotavirus can infect intestinal epithelium villi cells and lead to watery diarrhoea, resulting in intestinal dysbiosis that destroys the microbial barrier and, in the end, worsens diarrhoea and leads to fatal consequences in some cases [90]. This case is interesting because

it has been seen that the use of probiotics could improve the recovery of the children [91]. *Lactobacillus rhamnosus* GG, *Saccharomyces boulardii* and *Lactobacillus reuteri* DSM 17938 are the most common probiotics studied and used in children with acute gastroenteritis (AGE) [91–93]. *S. boulardii* can reduce the duration of diarrhoea by approximately one day, shorten the initial phase of watery stools, and decrease the length of hospital stay [92]. Numerous mechanisms on the action of *S. boulardii* in acute infectious diarrhoea have been proposed, where *S. boulardii* seems to play a role in the regulation of intestinal microbial homeostasis and the stabilisation of the barrier function [94]. This precise example of disease and probiotic improvement will be extended and studied in Chapter III of this thesis, and more details will be provided then.

Another example that has a great interest in the field is the effect of antibiotics in gut microbiota. It is already known that the use of antibiotics functions as a critical perturbation to the microbial community, generating profound changes that can produce undesirable effects as the uprising of hidden colonisers, i.e. the case of *Clostridium difficile*, leading to colitis in some cases [95, 38]. *C. difficile* is a spore-forming, gram-positive and anaerobic bacterium that produces two toxins, TcdA and TcdB, that cause colitis through colonocyte death, loss of intestinal barrier function, and neutrophilic colitis [38]. The colonisation of *C. difficile* is possible only when the microbiota is compromised, and it has a reduced diversity (see Figure 1.8). Besides, *Salmonella typhimurium* often take part in the infection of *C. difficile*. Both bacteria catabolise sialic acid, a product excreted by other bacteria as *Bacteroides thetaiotaomicron*, a process that is enhanced by the application of antibiotics [96]. It is important to take notice of a critical issue; it is very well known the fact that antibiotics (bad usage of them mostly), speeds up the apparition of antibiotic-resistant strains that become hard to manipulate and are the cause of death for many people. In 2015, the number of people dead by antibiotic-resistant was around 50,000 in Europe and USA. The estimations are that around 10 million people will face death worldwide by 2050 [97]. The discovery of new forms of dealing with these pathogens will become one of the hardest challenges that science will need to overcome in the next years.

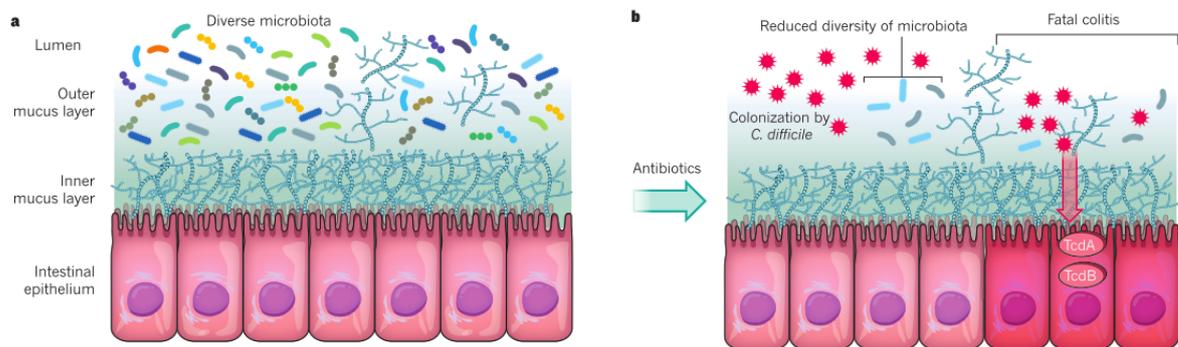


Fig. 1.8 An illustrative case of pathogen colonisation after antibiotic treatment. In the left part, it is represented a healthy microbiota located in the intestinal lumen. In the right part, it is represented a colonisation by *C. difficile* due to the decrease of diversity in microbiota. From Baumer and Sperandio [95].

Lastly, and changing of habitat, it would be interesting to talk a little bit about caries, as it is the central issue explored in Chapter IV. Recent metagenomic studies showed that the oral cavities affected by periodontitis, gingivitis, halitosis or dental caries are colonised by a variety of microbial species, including those found in healthy oral microbiomes [57, 98]. It suggests that the oral diseases are not caused by an overgrowth of a single pathogen as previously thought, such as *Streptococcus mutans* in dental caries [99], instead, they are caused by a dysbiosis of the oral microbiome [98]. However, despite rigorous metagenomic sequencing efforts, the specific microbes in a dysbiotic community that cause a particular oral disease have not been identified yet. Nearly 100% of the population have experienced or will experience dental caries or other oral diseases during their lifetime. It means that the pathological conditions in the oral cavity are regularly occurring. When the human leukocytes combat pathogens, reactive oxygen species (ROS) are formed [100]. Overproduction of ROS results in oxidative stress, which triggers structural and functional changes of proteins, lipids and nucleic acids [101]. Several studies have reported that the saliva from patients with periodontitis and/or dental caries exhibited elevated levels of oxidative stress markers and protein inflammatory markers [55, 102, 103]. It has been demonstrated that the levels of the oxidative stress markers rise temporarily after consumption of meals containing ROS stimulating bacterial strains [104]. The composition of the oral microbiome also shows intra-individual temporal variations [55]. For all these reasons, oral health is a complex issue in which several factors are involved, and where it is happening a change in the paradigm of one-bacteria, one-disease, mostly related to caries.

1.2 Dynamics of human microbiota

The previous section was intended to describe some snapshots of the human microbiota. However, microbial communities are dynamical, and they do not exist in only one stable (and steady) form. They evolve in different time-scales with different implications to the stability and shape of their ecosystem. In the current section, it will be exposed some general evolutionary concepts, and most important, the utility of knowing the short-scale dynamics of microbiotas with follow-up studies. This latter part has been critical for the major part of the results presented in this thesis, and it will be defended that measuring microbial systems in time opens the way to powerful analyses for understanding essential properties of these ecosystems.

1.2.1 The evolution of microbiota

There is grandeur in this view of life, with its several powers,
having been originally breathed into a few forms or into one;
and that, whilst this planet has gone cycling on according to the fixed law of gravity,
from so simple a beginning endless forms most beautiful
and most wonderful have been, and are being, evolved.

On the Origin of Species

Charles Darwin

Evolution is the backbone of Biology. It is the theory that unites all biological knowledge and makes it understandable in the light of history. Dobzhansky said that *nothing in biology makes sense except in the light of evolution* [105]. Since the publication of Darwin's book, *On the Origin of Species*, in 1859 [106], the field of Biology has grown in many ways and shapes, allowing us to interpret biological entities according to their ancestry forms. Moreover, evolutionary theory has a central role in understanding microbiome systems along with ecological theory. In this section, we will only get a glimpse of the some evolutionary forces acting upon the human microbiome.

As it has been exposed in the previous section, human microbiota is primarily composed of bacteria with a minor fraction of archaea and viruses. Microorganisms have the ability to change their genomes with great freedom, allowing to create new traits without a general loss of fitness. It is said that microorganisms have a high 'evolvability', it is, a significant ability to evolve. The mechanisms that permit these innovations go further than punctual mutations to more sophisticated modifications as duplication and drift of existing genes, changes that can be dramatic and well tolerated by bacteria [107]. However, one of the best and quickest

ways that prokaryotes have to achieve new functions is through horizontal gene transfer (HGT), an ability that eukaryotes seem to have lost [108]. In an exciting work by Smillie *et al.*, they proved that one of the principal forces shaping the human gut microbiota was HGT [109]. In their work, the authors observed that bacterial genomes are shaped mainly by the ecology rather than geography or taxonomy relatedness. This statement means that even if millions of years of evolution separate the bacteria that live together, the genetic traffic through HGT will be a great force shaping their genomes if they share the same ecosystem. The HGT allows, then, rapid transit of functions between microbiome components as can be antibiotic resistance, an issue that can completely modify the behaviour of such organisms [110]. Moreover, the fact that new functions arise quickly, and are easily shared among the community, has deep impacts in what we consider as an *organism*, and whether studying organisms alone is an accurate approximation for the analysis of microbiomes [111].

Looking at the evolutionary scale involving both hosts and their related microbiomes is a difficult task. We have seen that microbiotas are highly mobile and dynamic, and can change significantly with several factors. Thus, is it possible to see evolutionary patterns in the microbiomes that can be related to the evolution of the hosts? In a work of Groussin *et al.*, they studied how the gut microbiomes of mammals have changed over evolutionary time and came to the conclusion that dietary shifts have driven horizontal acquisition of ancestral clades, and host phylogeny only can predict the presence of recent bacterial lineages [112]. Here, we have an essential concept that we should take in mind, which is whether the components of human microbiome are transmitted horizontally or vertically and as often happens in biology, the answer is a mixture of both ways. The acquisition of specific clades due to dietary shifts are based upon the horizontal transfer, as they have to be included from the outside to be able to colonise the gut. However, the microbial members that have shown co-speciation in the gut are thought to be transmitted vertically, which is the reason for which they have been able to conserve an evolutionary structure that can be distinguishable from other related hosts. In this matter, it has been observed that co-speciation have occurred to some members of the gut microbiota from the *Hominidae* family, the same that include us with chimpanzees, bonobos and gorillas [113]. Measuring the divergence of specific strains from *Bacteroidaceae* and *Bifidobacteriaceae* with molecular clocks, it seems that their divergence time converges with the evolutionary times of host phylogeny (see Figure 1.9). Nevertheless, the existence of co-speciation does not involves necessarily co-evolution, the testing of which should be assessed in a more mechanistic way [112].

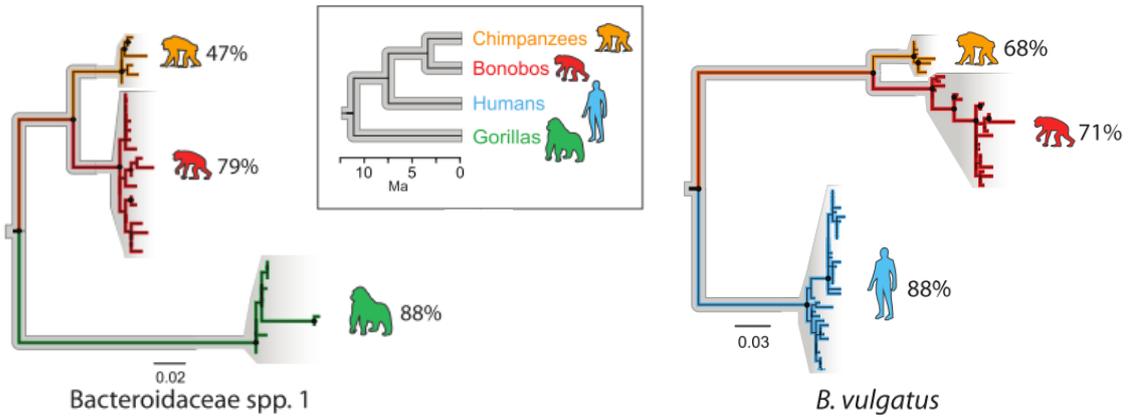


Fig. 1.9 Co-speciation events in *Hominidae* gut microbiota. On the left part, maximum-likelihood phylogeny of *Bacteroidaceae* lineages that codiversified with the African apes but went lost in humans. On the right part, maximum-likelihood phylogeny of a *Bacteroidaceae* clade that co-speciated with humans, chimpanzees, and bonobos. Modified from Moeller *et al.* [113].

1.2.2 Measuring time-series

(...) profesaba que la historia es un círculo
y que nada es que no haya sido y que no será.

Los teólogos

Jorge Luis Borges

If in the last section we have seen the importance that evolutionary theory has in defining the mammals' microbiota, we need to go a step further and jump into another scale of the spectrum: the short-term dynamics governed by temporal series of data. As we will see through this thesis work, we need time-series to properly study a *highly mobile* system as it is the *human microbiota*. By the observation of the ecosystem in time, we can study fundamental properties that can enable better and more in-depth knowledge about how microbiota behaves in stable conditions, how the community changes when a perturbation is present, and how rapid (and complete) can be the restoration of the original state, to mention some examples. In this section, we will review some works from the literature when the temporal dimension is available, and what is the potential for such analyses.

One of the leading and most interesting works that is published was the study of a male and a female, in four body parts (gut, right and left hands, and the tongue), by Caporaso *et al.* in 2011 [47]. A critical part of this work is that they measured more than 300 time-points in the case of the male, and approximately 150 time-points in the case of the female, giving us a tremendous quantity of samples to study. In the original article, Caporaso *et al.* did very basic analyses regarding the depth of the sequencing, and studying the core OTUs present in the different body parts and how they changed in time. However, many other works have been based on this one, making use of this very useful dataset to do other kinds of analyses as inferring microbial interactions through Lotka-Volterra system of equations [114], or studying the universality of the dynamics in human microbiota [115]. We will study in depth these two examples mentioned above in further parts of this introduction. Besides, this work was also used in the Chapter I of this thesis. What is interesting here is that, first, temporal datasets from other studies, that are under public accession, can be used to apply other dynamical tests; and second, the temporal dimension opens new kind of information about the system as can be the interactions between components.

Nevertheless, the main limitation of studies as the aforementioned, is the limited number of subjects under analysis. As Faust *et al.* pointed out in their review in 2015, the works that investigate temporal data are rare, and are either with few subjects and considerable amount of time points, or with many subjects with few time points [116]. In the same review, they have a large table with many time series analysis techniques used in microbiome profiling studies as the Local Similarity Analysis, used in one of the main chapters of the present thesis. Many of these techniques rely on the knowledge about dynamical systems studied with linear and non-linear differential equations, such as Lyapunov exponents to measure whether the community behaves with chaos or not [117]; the bistability analysis, that detects the existence of multiple stable states in the population [118]; or the predictability analysis, that tell us whether the community follow linear or non-linear dynamics [117]. The number of methods available to analyse temporal data is quite large, mostly due to the inheritance of methodology that was already developed in other fields as Physics. Still, these approaches still have to be fine-tuned and integrated within an ecological body of knowledge in order to be of biological interest. For example, it was reported that a lake under a massive perturbation recovered its initial state [119], a concept known as *resilience*, and such recovery was not entirely present in other kinds of perturbations like antibiotic treatment [120] or a pathogen invasion [59], both happening in the gut microbiota. Analysing these results with the techniques mentioned before tells us that there exist different microbial communities assemblages that can be valid, there is multi-stability, but it does not say anything about what mechanisms are allowing this, or what ecological consequences might have.

Another study, carried out by David *et al*, in which they measured the gut and salivary microbiotas from two subjects (A and B) for an extended period, along with daily life measurements [59]. What makes this work attractive is the fact that both subjects suffered a perturbation that altered their microbiotas in different ways: subject A travelled abroad from a developed country to a developing world, changing him/her lifestyle and diet entirely; while subject B suffered a *Salmonella* infection that caused disease. In Figure 1.10, in the upper row, we can easily observe the changes in the microbial community produced by both perturbations. The differences between subjects were higher than within subjects, except in the case of the *Salmonella* infection and the travel abroad, that changed scene entirely, at least for some time. In fact, the most exciting result is highlighted in the lower row, in Figure 1.10, where we can observe how both perturbations altered the original microbiota. In the case of subject A, the diet changed the community structure completely increasing the *Bacteroidetes* to *Firmicutes* ratio during the travel abroad, but returning to the original composition after the subject went back to the original country. On the other hand, the subject B that suffered a *Salmonella* infection changed its microbial community structure and did not return to the original composition. The authors theorise that both cases could represent different scenarios in microbiome variation, one in which the microbiota passes through a temporary state that is reversible, and the other in which the microbiota change its composition into another stable conformation. Again, a case of multi-stability that has been able to be studied only with temporal series.

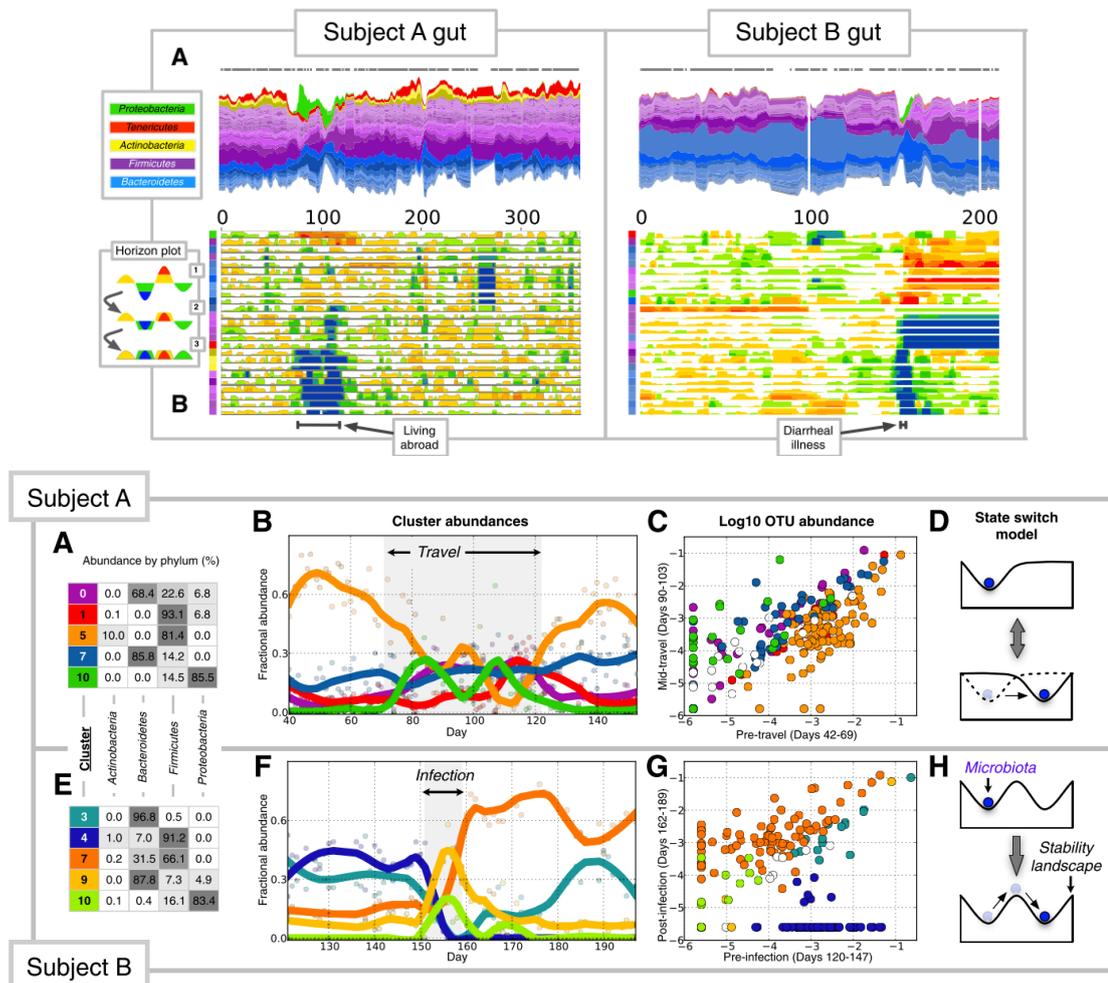


Fig. 1.10 Microbiome variation in time due to change in lifestyle. On the upper row, it can be observed stream plots showing OTU fractional abundances over time (A), and horizon graphs of most common OTUs' abundances over time. On the lower row, it is represented the cluster abundances over time during both perturbations (B,F), median log₁₀(abundance) of OTUs in each cluster before and after perturbation (C,G), and theoretical models proposed to explain the change in the microbial communities. From David *et al.* [59].

Another kind of analyses and tools that makes use of temporal series is the network inference of the community, an issue that will be greatly reviewed in the third section of this introduction. Summarising, with the availability of temporal series, we can use models based on simple correlations as it is the case of Pearson's correlation, or more sophisticated mathematical models based on systems of differential equations, as it is the generalised Lotka-Volterra system, used in an elevated number of publications [79, 114, 121–123]. The frequency of sampling is a critical factor that will determine the characteristics under study.

While large frequencies will enable seasonal or annual properties of the system [124], smaller frequencies will allow the researcher to study chaotic behaviours of the microbiotas or critical shifts in detail as we have seen in the previous example [59]. However, the sampling with human subjects have some problems regarding both the regularity and the availability. The first problem is a general one; it is essential for some studies to have evenly-spaced time points to apply some analyses techniques as classic Fourier transformations. The second one is related to the randomness of life itself, as it is complicated to be regular in large times in human studies [116]. What is important from this point is that time series help us to make more sophisticated analyses, uncovering hidden essential features, and letting us to avoid false positives and false negatives. As we are about to see in depth in the following sections, dynamics are essential to understanding the fundamental nature of these ecosystems.

1.2.3 Community assembly

Dijo Tennyson que si pudiéramos comprender una sola flor
sabríamos quiénes somos y qué es el mundo.

El Zahir

Jorge Luis Borges

The first steps in the establishment of an ecosystem is known as **community assembly**, and its understanding is essential because it has a great impact on microbial communities. After many years and experiments, we still don't know whether the order in species inclusion is important or not for the final result. In other words, first stages dynamics in ecosystems might be different depending on the way, and order, we put the individuals into the community. Another question is the incidence of the host in the process of community assembly.

From the theoretical point of view, we have numerous works that explored the dynamics of community assembly processes. The classical work from Drake explored the idea of assembly and succession by means of the generalised Lotka-Volterra model and its relationship with invasion resistance [125]. With the use of an algorithm to measure the resistance to invasions, he observed that ecosystem formation was different in each iteration and that each configuration was differently prone to invasions. This work has been the basis of many other studies after it, in which the idea of ecological assembly was treated from the theoretical point of view. That is the case of the work from Capitán *et al.*, that explored the dynamics of assembly using tools from statistical mechanics [126]. They used a variation of the generalised Lotka-Volterra model focused on trophic levels, and the main conclusion was that, under certain circumstances, the process only had a single stable solution. Therefore, there is only one stable solution, and it is independent of the initial conditions one may have, a

conclusion that seems controversial with real data. Lastly, another interesting work is the one of Dickens *et al.*, that studied the ecological dynamics of an ecosystem with many species with analytical solutions [122]. In their work, they specify that in community assembly many forces are acting as immigration, speciation, selection and drift. They develop a model based on the presence and absence of species that they call the Presence-Absence (PA) model. In their model, they take into account both the carrying capacities of the different members of the ecosystem and the rates of immigration and extinction. Among the many conclusions that they can infer with this simple model, it seems that community assembly can be related with the three distinct phases in which an ecosystem can be (total coexistence, a few species dominating, or all species extinct for high noise), with the Hubbell point playing a pivotal role in the dynamics.

Despite the usefulness of the theoretical works commented above, it is clear that they have limitations. The main limitation is the discontinuity between the aforementioned models and real data, or experiments. For example, it is hard to believe that independently of the initial conditions, you will end up with the same configuration, as the work from Capitán *et al.* states [126]. As it would be hard to believe that there is total and perfect control by the host/environment to select the members of your community. There are other kind of works that explore at what extent the host is selecting the microbiomes, with animal models (in the end, it is difficult to experiment with humans). The work from Napflin *et al.* is a good example [127]. They used a host that has a constrained population of microorganisms, the bumblebee *Bombus terrestris*. These insects are known to gain their microbiomes mediated by the environment in a vertical transmission mediated by the other insects in the colony. This behaviour is very common in nature and also happens with cockroaches and termites, for example [128]. They exposed germ-free insects to a mixture of all the microorganisms present in a large selection of colonies to see whether the host filtered them or not. The results, even being not conclusive, hint toward a certain selection pressure dependent of the host, which was different in some of the replicas, meaning that host genetic background could also be playing an important role.

Another quite interesting work is the one from Franzenburg *et al.*, in which they explored the community assembly of *Hydra* hatchlings [129]. *Hydras* are quite simple animals that have been broadly used to study microbiomes. Most importantly, their microbiomes remain stable in the adult stage and is easy to follow the steps of the community assembly, what makes them good candidates to study this process. The striking results reflect a robust process of assembly and succession that was repeated with high precision in the four replicas of the experiment. As it is observed in Figure 1.11 on the left part, principal component analysis of the bacterial composition place the samples in highly differentiated regimes in

the space, meaning that all replicas were comparable between each other at all times, but different between the times measured (with the remarkable exception of the second week and the adult stage). Additionally, they proposed a toy model based on the replicator-mutator equation, represented on the right part in Figure 1.11, to uncover principle rules controlling the microbial assembly process in the *Hydra*. By changing the fitness and colonisation parameters, they observed that under specific values, the output reassembled to the real data of microbial assembly.

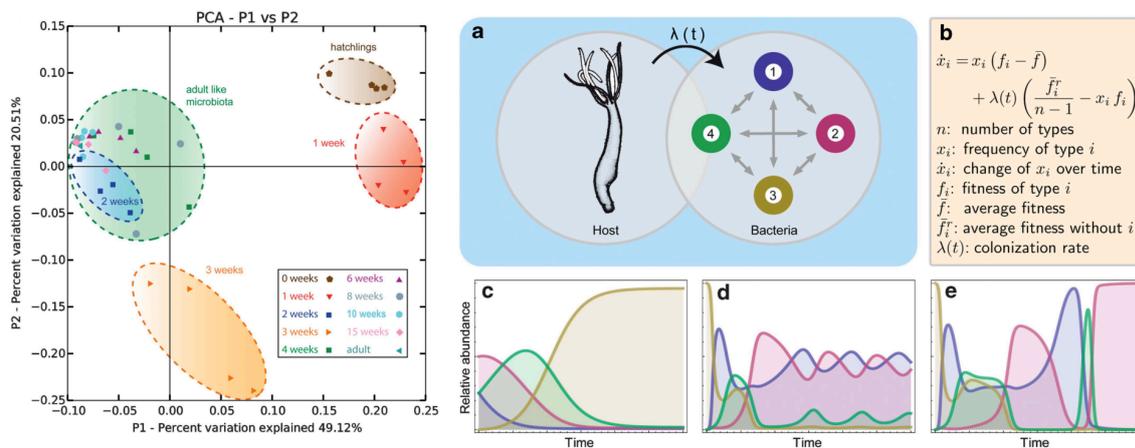


Fig. 1.11 Microbial assembly on *Hydras*. On the left part it is represented a PCoA of Pearson's distribution between *Hydra* microbiota at different time points, in all replicas. On the right part, it is represented the mathematical model based on the replicator-mutator equation (b), with three different scenarios: constant interactions and colonisation parameters (c); bacterial fitness modelled as abundance-dependent values (d); and colonisation rate assumed to be time-dependent, decaying over time (e). From Franzenburg *et al.* [129].

All the studies commented in this section are only a glimpse of the highly complicated panorama which is present in more complex microbiome ecosystems. We have seen in this introduction that the first steps of bacterial colonisation in humans are more or less well understood, with specific periods marked by the presence of distinct microbial groups. It might be that host genetics plays a role in the selection process, but it seems that its effect would be minor nevertheless [130]. Microbial assembly in humans will be defined not only by host genetics but many other forces as selection, extinctions, functional redundancy, interactions and stochasticity, among many other unknown forces. Moreover, the final solution will not be the only possible one. In the final section of this introduction, we will explore one of the most important issues in microbial ecological theory, the interactions, that will play a major role in shaping the microbiomes.

1.3 Interactions

Up to this point, we have studied what is a microbiome, how it is in different scenarios, that these populations are highly mobile, and that temporal series help us to discover the dynamics governing this complex system. The final section of the introduction will focus on the concept of *interaction* between the individuals in microbial systems. The individuals present different responses to other members of the community, maybe helping their relatives or fighting against possible invaders, but also developing more complex and sophisticated actions of treachery, changing sides or suicide, as would happen in a classic Greek tragedy. We will see how this battery of interactions can be modelled and also inferred, and how they help to uncover essential properties of population stability.

1.3.1 What is an interaction

Principles

You can't say A is made of B
or vice versa.

All mass in interaction.

Richard Feynman

Defining what is an interaction is not a trivial question. To interact is to establish a connection between two elements, by means of transmission of matter, energy, or both, that will end in three different classes of effects to each *interactor*: positive, negative or neutral. In our case, the elements that interact are the biological entities that are present in an ecosystem, behaving in different manners depending on the other organism. This concept is among the most essential subjects treated in the field of Ecology, and we can classify the types of interaction depending on the gain/loss output. We know many examples taught in schools for years, like the classic *trophic chains* in which we have producers and predators of a different hierarchy, like the herbivores that eat plants, or hunters that eat herbivores (or other hunters). However, defining interactions in terms of trophic chains has limited value if we try to extend this idea to other types of organisms. Instead of this way, we usually describe interactions concerning gain or loss in fitness after the *interaction event*. For example, the predatory behaviour of a lion and a gazelle will likely end up with the herbivore dead, and the hunter with food to eat. Thus, we can say that a predatory behaviour has a positive effect on one individual, and a negative effect for the other (+/-). With this generalisation, we can specify a limited range of possible interactions that can happen in an ecosystem, that are summarised in Table 1.1. Although helpful, we should not forget that a partition like this does not capture

Interaction	Species X	Species Y
Mutualism, co-operation	+	+
Commensalism	+	0
Competition	–	–
Exploitation, predation	+	–
Amensalism	–	0
Neutralism	0	0

Table 1.1 Main interactions between organisms in an ecosystem. Extracted and modified from *Nature Knowledge Project*.

the real complexity of the interactions in biology, and, sometimes, organisms can change *sides* and shift from one type of interaction to another with ease.

At a higher scale, and focusing on the subject of this thesis, we find three different types of effect between microbiomes and hosts: microbe to host, host to microbe, and between microbes [131]. Even though we will focus on the third type of effect (microbe to microbe), it will be worth knowing what we currently know about the interaction between the host and the microbiome, which has different properties depending on the interaction direction. In fact, we studied the effects of the microbiome to the host in the first section of this work, when we saw how dysbiosis could cause diseases like IBD [73] or Crohn’s disease [74]. Nevertheless, there are also positive effects from the microbiome to the host that arise both acting directly or indirectly. For example, it is well known the fact that microbial species confer *colonisation resistance* to the host by niche competition [132]. Furthermore, human microbiome has an important role in host metabolism, providing short-chain fatty acids like butyrate, which acts as signalling molecule on the host that promotes beneficial properties as seen in this introduction concerning insulin resistance [87, 133]. The immune system maturation is also affected by the microbiome [134]. There are more examples of this kind of interaction, from the microbiome to the host, like the ability to increase the energy extraction from food, increase nutrient harvest, and alter appetite signalling [135]. With respect to the other way of interaction, from host to the microbiome, we have studied how the host could be selecting the microbes to live with it in the process of microbial assembly [131]. Furthermore, there are other types of effects from the host to the microbiome, such as compartmentalisation of microbes, letting some parts without microbes (i.e., the brain) [132], feeding certain types of microorganisms by epithelial cells [136], or monitoring which actions they do [131].

Either way, the interesting part for us in this section is the kind of interaction that happens between microbes. These interactions can take multiple forms, like simple predatory

behaviour of *Bdellovibrio bacteriovorus*, a highly mobile delta-proteobacteria that invade the periplasm of other larger gram-negative bacteria, killing and digesting them [137]. Another typical and simple interaction is the competition for nutrients, which also confers the resistance to colonisation that we described before. However, we also find more sophisticated ways to interact as are cross-feeding metabolic exchanges between microbial species. In other words, some metabolites that are produced by specific microbes will be the metabolic substrate for others, becoming a co-operative type of interaction between microbiome components, which also has significant impacts in the host. In Figure 1.12 we can see that metabolic exchange can lead to the formation of short-chain fatty acids as acetate, butyrate or propionate, essential elements for us [138]. Microbes can also change their behaviours depending on the conditions. On an exciting work by King *et al.*, they observed that bacteria belonging to *Enterococcus faecalis*, that are mild pathogens to the worm *Caenorhabditis elegans*, rapidly evolved after few generations in the presence of a more pathogenic bacteria, *Staphylococcus aureus*, to confer protection for the host [139]. The main changes in their genomes were related to the production of antimicrobial molecules against *S. aureus*. This change of interaction type, from parasitism to mutualism through evolutionary time, highlights the complexity that is present when microbes interact. These interactions can be highly precise, bacteria acting as microbial surgeons, affecting only selected objectives. That is the case of the example studied in the work of Wexler *et al.*, where they described how the strain *Bacteroides fragilis* NCTC 9343 interact directly with other bacteria using the Type-6 secretion system [140]. This strain establishes multiple connections with the surrounding bacteria, injecting antimicrobial components that will be lethal for other *B. fragilis* that do not have the immunity genes encoded in that specific strain. In other words, *B. fragilis* NCTC 9343 is making a strain-specific interaction, killing other members from its same species. And, this is only one example in one specific strain; the possibilities in a whole microbiome are quite large if we think about the thousands of strains that inhabits our body, with many strategies to deal with each others.

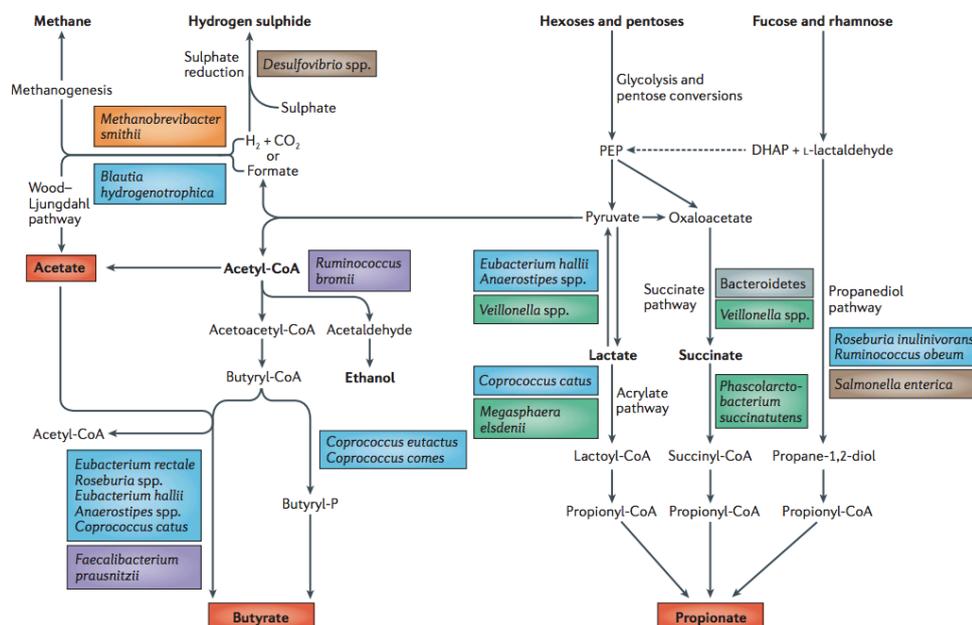


Fig. 1.12 Metabolic exchange between microbes. In this figure are represented the pathways that are responsible for the biosynthesis of the major microbial metabolites that result from carbohydrate fermentation and bacterial cross-feeding. From Louis *et al.* [138].

These examples of interaction complexity are only the tip of the iceberg, and how life forms can interact are practically endless. The ecosystems become war zones where microbes battle each other with the production of antibiotics, that also leads to the creation of counter-measures by protecting themselves from these molecules [141]; altering the production of quorum sensing molecules as autoinducer-2 promotes a shift in the Firmicutes/Bacteroidetes ratio [142]; or that physical properties like adhesion in biofilms can work as a weapon against other members that are less efficient to surface attachment [143]. Phages could also promote diversity because they spread through host bacteria that are abundant, and can give rare bacterial species an advantage [131], or they also help the horizontal gene transfer, a process that fastens the transmission of helpful and advantageous functions across microbiomes [109]. Up to this point, we should ask ourselves whether such complexity can be adequately studied and integrated into models, or not. Microbial behaviour can shift from one type of interaction to another, as seen in this point, and the same two bacteria can behave in different ways between them by changing the environmental conditions. However, simplifying all this mess and specifying the effects between microbes with only three different effects (+, -, 0) is handy when we want to model it using mathematics. So, even using this simplification, we

are able to extract central properties of microbial ecosystems as we will see in the following sections.

1.3.2 Inferring interactions in microbiomes

What I cannot create, I do not understand.

Richard Feynman

After the picture we draw about the complexity of interactions, just imagine how hard it is to infer what is happening without seeing it. In the macroscopic world, we can observe how the lion hunts the gazelle, or ants taking care of aphids to eat their ‘poop’, a sugar-rich substance. However, it is a lot more complicated to observe a bacteria injecting antimicrobial molecules into another. If we want to address this problem in an rich ecosystem like the human microbiota, we should use methodologies that use the power of time series that can robustly infer microbial interactions.

One of the most basic approaches is to use correlations to extract interactions. The idea behind is quite easy to understand: if two species grow in time, they will have a positive correlation; if two species have opposite behaviours (one grow, the other reduce its abundance), the correlation will be negative. The usual form is by the use of Pearson’s correlation, that takes the covariance of the two variables and divides it by the product of their standard deviations. Other forms of correlation exist, like Spearman’s rank correlation coefficient, which is a form of Pearson’s correlation between the rank values of the two variables. The main problem of using correlation is due to the way in which we have microbiome data, that is not by absolute counts but rather by relative abundances. This problem leads to compositional effects in microbial abundances, where the increase or decrease of determined species can mask accurate positive or negative correlations. There are some software that tries to amend the compositional problem, like SparCC [144]. Also, correlation coefficients are not optimised to use temporal data, an obstacle that other techniques like LSA (from Local Similarity Analysis) try to overcome using a sophisticated form of correlation [145]. Besides, another problem that arises from correlation analyses is the classical motto: *correlation does not imply causation*, that in this case implies that a correlation between two species could be an artefact instead of a real interaction, something that has been extensively proved [114]. However, not everything are bad news, correlation analyses are still used because they are a quick analysis that can shed light in the first steps of a pipeline, and correctly used, they can even be a powerful tool to extract real interactions as we will see in Chapter II.

Although the community had used correlation coefficients for a good time, there has been a shift toward the use of more complicated models to make the inferences. The most used

model is, without any doubt, the classical Lotka-Volterra model. This model was proposed first by Alfred Lotka in 1910 for chemical systems [146], and further developed by him in 1920 to address biological problems [147]. Vito Volterra used the same system of equations in 1926, and later on, the system acquired the name of both scientists. The classic form of this model measured how a population formed of preys and predators behaved in time, and then it was further extended to other types of biological problems, giving rise to the *generalised* Lotka-Volterra model. The equation in 1.1 is an example of a generalised Lotka-Volterra model with parameters for individual growth (μ_i), and an interaction network (M_{ij}) that specifies how the members of the ecosystem interact with each other. This model has the form of a system of differential equation and has been massively studied since its conception, one of the main reasons for why it is among the most used models in ecology nowadays. In fact, this model has appeared many times through the introduction. However, being this a model to see what happens in an ecosystem given an interaction matrix, how can we use it to infer precisely the interactions of a given ecosystem? It is a case of reverse engineering, and some works have explored this way of interaction inference.

$$\frac{d}{dt}x_i(t) = \mu_i x_i(t) + x_i(t) \sum_{j=1}^L M_{ij} x_j(t) \quad (1.1)$$

For example, Stein *et al.* proved that they could robustly infer interactions by linearising the Lotka-Volterra system, and estimating the model parameters by the Tikhonov regularisation [79]. A refinement of this methodology came out with the work of Bucci *et al.*, where they improved the model by adding various parameter estimators based both on maximum-likelihood or Bayesian algorithms [121]. Furthermore, Fisher and Mehta also used the Lotka-Volterra system of equations together with an algorithm they created to infer interactions in microbiome samples [114]. Some examples from these works are represented in Figure 1.13. As seen in all examples, the number of species from which interactions are inferred does not exceed the dozen or so. This aspect is one of the main limitations of the aforementioned methods; they need a significant number of samples to make the interaction inference. Another limitation concerning the Lotka-Volterra system of equations is its use as a rough approximation to measure the inextricable nature of microbial interactions. The use of model parameters helps to make it more realistic, but on the other hand, it also makes it harder to analyse due to overfitting problems, to name one limitation. Either way, assuming a more realistic model like the Lotka-Volterra, instead of simpler ones like Pearson's correlation, is a significant advantage which is helping to elucidate interesting biological interpretations from metagenomic data.

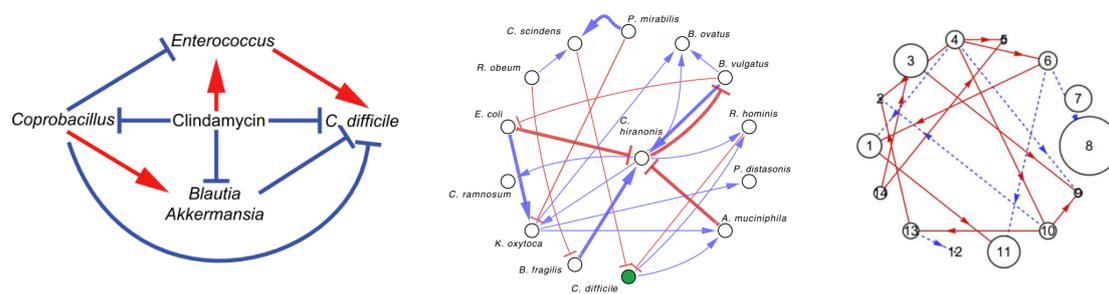


Fig. 1.13 Different interaction networks inferred from samples. On the left, it is represented an example of the work by Stein *et al.* [79]; on the center, it is represented an example from Bucci *et al.* [121]; on the right, it is represented an example of the work by Fisher and Mehta [114].

In any case, researchers are exploring complementary ways like an equation-free method that does not rely on any equation like the generalised Lotka-Volterra model [148]. This method relies on a modification of the algorithm from Fisher and Mehta [114], generating a sparse interaction network without assuming any particular equation. This methodology also allows for non-linearities, a considerable limitation in other types of interaction inferences. Furthermore, there is a good body of literature about Bayesian estimation of microbial interactions, used in several works [78, 149, 150]. Nevertheless, Bayesian estimation will not be treated in this introduction because we have used the Lotka-Volterra system of equations due to its simplicity and because it is one of the most studied systems in ecology.

1.3.3 Time stability for complex networks

The problem was how to keep the wheels of industry turning without increasing the real wealth of the world. Goods must be produced, but they need not be distributed.

And in practice the only way of achieving this was by *continuous warfare*.

1984

George Orwell

From the dynamical analysis of microbial populations we can derive essential properties as the *ecosystem stability* over time, and robustness to different classes of perturbations. The study of stability in dynamical systems is a well described body of knowledge, extensively

treated in mathematics and physics, from where ecology utilises some of their tools [151]. Besides, microbiome stability seems to be related to the health status of the host, as we describe in Chapter I. To know how the microbial species interact with each other enables interesting approaches to study this key property, an issue that has been explored since decades ago.

One of the most known facts about ecosystem stability was presented in a seminal work by Robert May, where he described that ecosystems with a large number of species tended to be unstable [152]. This fact resulted striking to the ecologists back in the 70s because it was thought that a high diversity helped to stabilise an ecosystem, and May's work came up as a paradox for this observation. It is a paradox because we know now that systems as our gut microbiome can be stable even with hundreds of species cohabitating the same space. How can this be possible? Today, we know that May's assumptions were failing in one key point: he was studying community networks with randomly distributed types of interactions, meaning that cooperation (+, +) and competition (-, -) were occurring with half the probability of exploitation (+, -). Then, the question is whether the types of interactions are important or not to promote stability in ecological networks, an issue treated in a great work from Coyte *et al.* [153]. In their work, they explored the ecological stability by different analyses, varying the proportion of cooperative interactions as represented in Figure 1.14. First of all, measuring the network's eigenvalues, a classic stability measure, they observed that increasing the number of species created instabilities in the network. Besides, networks with only cooperative interactions were unstable meanwhile increasing the competition proportion helped to stabilise the network. They proved the same fact with two new types of analyses: a *permanence analysis* that measured how communities behaved away from equilibrium; and an *individual-based*, less general but more explicit that allow them to follow each species per separate.

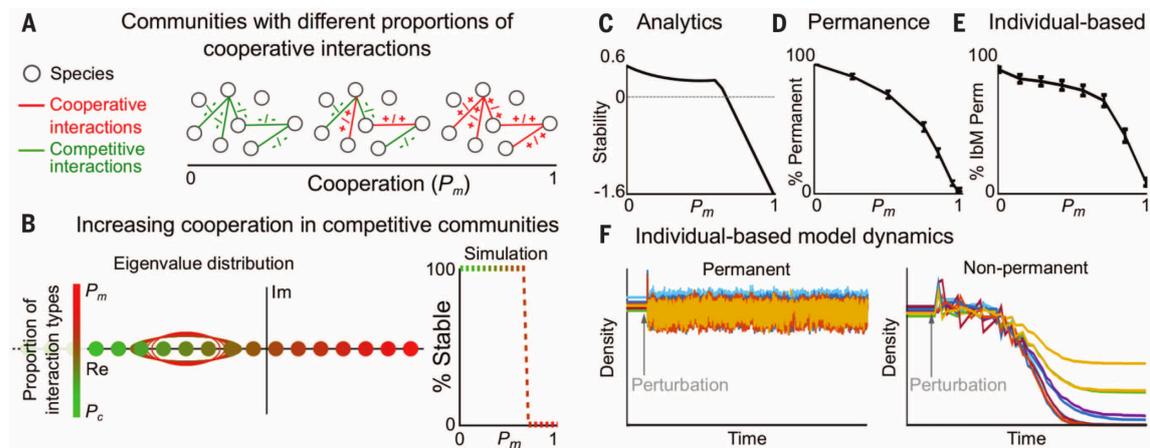


Fig. 1.14 Cooperation proportion and its influence in network stability. In **A**, how the proportion of cooperation (P_m) is changed in the analyses. In **B**, numerical analysis of Linear Stability Analysis. From **C** to **E**, analyses of networks with different P_m and their stability by the three different stability measures. In **F**, two different results in the *individual-based* analysis after a perturbation in a permanent or a non-permanent population. After the perturbation, the species remain stable only in the permanent population. Figure from Coyte *et al.* [153].

The fact that competition could, in principle, be a network stabiliser, seems counterintuitive. We might think that cooperation between the parts leads to an improvement of the system, but it instead seems that positive feedback loops can destabilise the network via coupling between species. This fact means that if a species decreases for some reason, it will pull down the other species and thus, destabilise the ecosystem [153]. However, it is also true that microbial cooperation promotes metabolic efficiency in microbiomes [154, 155], so, in the end, we will have to reach a compromise between metabolic efficiency (and other positive effects arising from cooperation) and their destabilising action. Then, a rich ecosystem like the human microbiome should have a proper proportion of competitive interactions among other factors that help to increase network stability. Coyte *et al.* [153] explore three different factors that increases this stability: the action of the immune system can reestablish the normal microbial abundances by an abundance-dependent action, something that helps to stabilise the network from other positive loops [156]; the strength of the interactions is another essential factor, and weakening them (with compartmentalisation, for example) leads to stability [157]; and, lastly, it seems that host epithelial feeding can also lead to stability by allowing to grow specific microbes, and by weakening their interactions. If you have a carbon source in abundance, you do not have to compete with others for food [131].

Up to this point, we have seen that cooperation is a double-edged sword, which can help the microbiome in terms of metabolic efficiency but also can decrease the stability at high species abundance. However, both cooperation and competition are needed to maximise the ‘fitness’ of the entire ecosystem. Indeed, in a complex population like a microbiome, there will be members that will benefit other microbes to gain advantages [158], but competition will always exist because microbes will always fight fiercely for the resources [159]. However, there is an essential component in ecosystems that we have not explored in detail, the self-regulation. Barabás *et al.* explore the importance of self-regulation in large ecological networks, finding that most of the species should have strong self-regulation in order to confer stability to the network. This concept of self-regulation can be understood concerning the carrying capacity of an ecosystem, meaning that the ecosystem can allow a limited number of biological beings to live in there. But the concept is more complicated than that, because, sometimes, bacteria can present complex (and somewhat counterintuitive) behaviours as ecological suicide [160]. Contrary to what we could think, it seems that this specific behaviour can rescue the microbial population to get extinct in specific conditions, a practice which is very common in Nature. Finally, as Valverde *et al.* points out in their wonderful work, the architecture of mutualistic network could have been evolved as an evolutionary spandrel [161]. Given some set of rules to build ecological networks, it seems that mutualism is a by-product of the main structures in the network. Among the interesting questions that arise from this point, we should ask what properties of cooperation and competition are due to the way in which Nature builds complex networks in microbial ecosystems.

Another way in which dynamic system analysis helped to understand the behaviour of complex systems is by defining, theoretically, that there could be different regimes in which the ecosystem can be. This idea is strongly related to the methodology and results from Chapter I, and it has been explored in different ways. The main idea is based on the assumption that every system in which there is a noisy component, can be in different phases depending on the noise magnitude along with the robustness of the system [122, 123, 162]. We have just seen that in terms of cooperative interaction proportion, or with the self-regulation strength. Fisher and Mehta explored the different regimes in which an ecosystem could be by the study of Lotka-Volterra system of equation, and a model of presence-absence of species further developed and analytically solved in other work [122, 123]. The main conclusions were that there are two different regimes, a niche and a neutral regime, depending on whether the stochastic component was important or not. We use the same idea in the Chapter I by using a model based on the Langevin equation [162]. The Langevin equation, represented in equation 1.2, has been used to model the Brownian motion and has two

main components: on the one hand it has a deterministic part that measures the ‘fitness’ of the system; and on the other hand, a stochastic component that measures the amplitude of fluctuations. In the original work, they study the ranking dynamics, and the equation 1.2 measures the ranking of a given object over time (\dot{x}_i). The solution of the equation gives three different regimes: a *rank-stable* regime where neither the ranking nor the score of each item varies over time; a *score-stable* where the score is stable, but the ranking can change; and the *unstable* region, where the system is unstable under any perturbation. In Figure 1.15 it is represented in this phase diagram, where the authors explored some systems as Wikipedia entries, Twitter citations and English use of words among others.

$$\dot{x}_i = B_i \cdot x_i^\alpha + A \cdot x_i^\beta \xi_i(t) - \phi(t) \cdot x_i, \quad (1.2)$$

Both works, the one by Fisher and Mehta, and the other by Blumm *et al.*, converge in saying that there will be a *stable* regime dominated by deterministic forces, and an *unstable* one dominated by stochasticity [122, 162]. In physical terms, it is more or less clear, but from a biological point of view is more tricky. In principle, the stable regime will be dominated by forces like the species abundance and interactions, as seen through this section; and the unstable regime will be dominated by stochasticity, in which neutrality can be an important factor. Neutrality in ecology shares the same principles as in evolutionary biology: a neutral change in an ecosystem means that you can interchange specific species for others without changing the global properties of that ecosystem. A neutral change that will root mainly in the topic of the next point. In fact, one of the main results from the Dickens work is that neutrality can be acting as a special point where the regimes converge, where all species are identical, and their dynamics are uncorrelated. Nevertheless, it seems that ecosystems transitions from one regime to the other depending on the conditions, an aspect that will be essential in the first chapter.

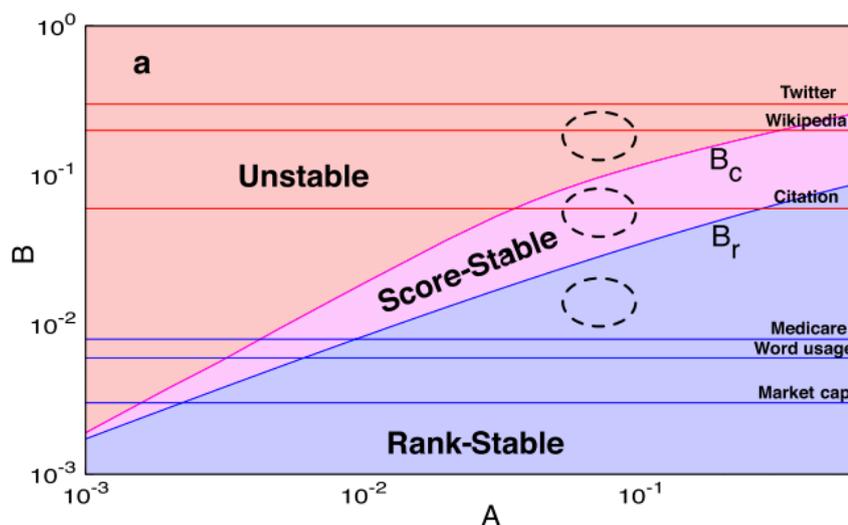


Fig. 1.15 Phase diagram of ranking stability. Three different phases were found, and some systems were explored and placed in the Figure according to their stability. Figure from Blumm *et al.* [162].

1.3.4 Function redundancy on microbiomes, and multi-stability

‘What a stupid exercise,’ said her dæmon after five minutes.
‘Words belong in contexts, not pegged out like biological specimens.’

La belle sauvage

Philip Pullman

We have talked about the different interactions that can be present in an ecosystem, and how they shape its dynamics. These interactions are performed by individuals, and we can explore this problem from another perspective instead the one based on the individual; it can be explored by the functions coded in the genomes of the individuals. A crucial issue to have in mind when we talk about microbiomes is that it is not only important *who* is there, but also *what* are they doing. This approach based on the functional capability of a microbiome has deep biological implications. Microbiome studies commonly refer to species or strains counts as a proxy for what they are doing, but it is important to notice that even two strains from the same species can perform quite differently in similar conditions [140, 163]. This fact is mostly due for horizontal gene transfer events, a powerful microbial tool to transfer genes rapidly among the population [109], and the spread of antibiotic resistance is one of its most

known effects [164]. However, there is another essential property of microbial systems, their *functional redundancy*, or how microbes can perform similar actions in different manners. Functional redundancy is the product of two main forces mainly: the HGT which allows functions to flow in the community; and the existence of non-unique solution for a problem given that the same function can be carried out by different enzymes, for example. The second force is also known in evolution as an *analogy*, when different organisms develop a similar feature by different solutions, mostly due to distinct evolutionary trajectories. For example, we find that many microorganisms have developed different mechanisms for the same functions as energy-yielding, oxygen respiration or carbon fixation [111].

The important point of the functional redundancy is given by the fact that, even though a microbial community can be different in abundance and presence of taxa, it could be similar regarding functions carried by this community. Within the microbiome realm, this fact means that even changing the microbiome composition through time, we could still have the same functions as before. We can be unstable regarding taxa, but stable concerning functions. This issue was recently reviewed by Moya and Ferrer [71] and has been observed in several systems as children developing type 1 diabetes [165], across the Human Microbiome Project (HMP) data [166], across body parts [43], or even the functions in all oceans on Earth [167]. In Figure 1.16 we can see that KEGG gene categories or the metabolic gene groups are pretty stable given the instability of their taxonomic assignments [111]. Functional redundancy has, thus, profound implications in ecosystem dynamics. As Moya and Ferrer state in their review, the notion of stability in microbiotas can behave in different ways depending on the output after a perturbation. For instance, a microbiota that does not change its composition is said to exhibit *resistance*, a microbiota that changes its composition but can restore it is said to exhibit *resilience*, and a microbiota that changes its composition but not in their functions is said to exhibit *functional redundancy* [71]. Functional redundancy also reveals a truly outstanding fact about microbial systems: you can go from a taxonomically described community to the realm of functions, but not from the function description to the taxa [111].

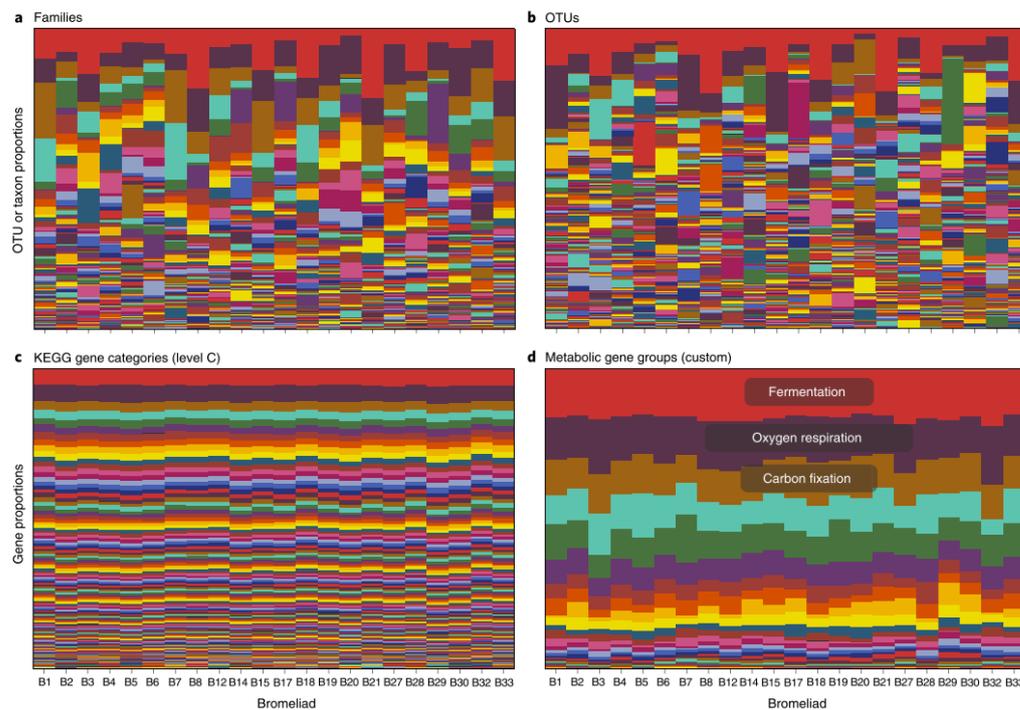


Fig. 1.16 Functional redundancy in microbial systems. In **a** are represented the families and in **b** are represented the OTUs at 99% of gene similarity. In **c** it is represented the KEGG gene categories, and in **d** custom metabolic groups from the taxa from the above row. From Louca *et al.* [111].

The implications of functional redundancy have, also, a strong connection with another key concept, *multi-stability*. The disentanglement between functions and taxa allows having several community configurations with the same functionality. There is not only one possible solution for the stability problem in microbial systems. In other words, there can exist different microbial communities equally stable in the same conditions. This is the idea behind concepts as the enterotypes [50], where different stable and healthy microbe configurations can exist at the same time. The work of Gonze *et al.* explored this idea with a toy model of bacterial interaction between three species [72]. They used a model based on Hill equations, represented in Figure 1.17, where they observed that depending on the initial conditions, or the change in the equation parameters (as bacterial growth), the system could have different stable states. They extended the idea to a system with 15 species, connected in three different modules, and concluded that any of the three clusters could be the most abundant and stable depending on the conditions..

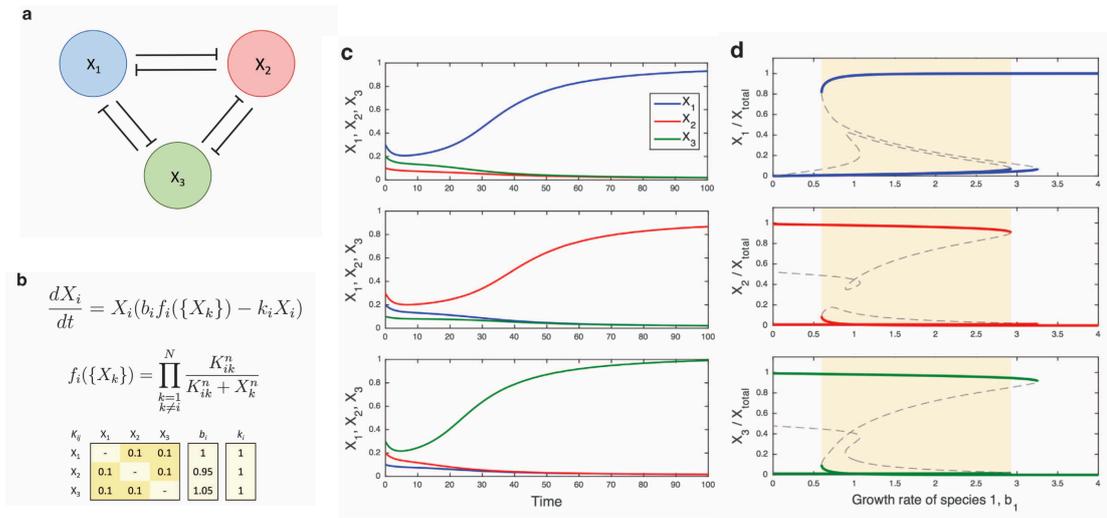


Fig. 1.17 Multi-stability in microbial systems. In **a**, the three species of the toy model. In **b**, the Hill equations that govern the system, with parameters b_i for bacterial growth, k_i for death rate, and inhibition term f_i with inhibition coefficients K_i . Here, $n = 2$. In **c**, numeric simulations that shows that the abundant species depend on the initial conditions. In **d**, three bifurcation diagrams show for each species the range of the three stable states. From Gonze *et al.* [72].

Together, the ideas of functional redundancy, neutrality, and multi-stability are a powerful tool to explore biological implications in microbiomes. Taking the extension of Waddington Landscape proposed by Moya and Ferrer [71], the events that affect microbiotas shape them, but also they have the possibility of ‘jumping’ between states in the future thanks to the properties we have reviewed through this introduction. Nevertheless, we should be aware of an important point in all this mess, measuring neutrality is not trivial. We can easily rely on arguments saying that the species interchange conserving the functionality of a microbiome are neutral changes, but the system complexity is far higher than we can observe at this moment. Even though two different microbiotas can develop similar functions, we do not know much information about the reaction kinetics of the enzymes, the ability to get nutrients from the media, the genomic plasticity that some species can have, and a myriad more details we cannot know, and, in the end, shape how are the interactions between the system components [111]. Because of this, we should never lose focus in the species interactions, for they are one of the most powerful factors that shape the behaviour of a microbiome.

Objectives

Microbial systems are highly dynamical, always adapting to the ever-changing conditions that shape them. This variability is also present in the human microbiome, a diverse microbial system which is influenced by several external forces. The interplay between host and microbiota derives in changes to host's homeostasis that could lead to diseases. Although it seems that a stable microbiota is connected, somehow, to a healthy host, it is still unclear what means to have 'a stable microbiota'. Thus, uncovering the relationship between microbiome variability and host health status would be an improvement in the dynamical analysis of human microbiota.

The present thesis work aims to elucidate the conditions for stability in the human microbiome, propose a quantitative definition for this concept, and connect it to the health status of the host. Once established the definition of stability, we aim to use it on our experimental data of follow-up studies, and try to prove its validity on other kinds of diseases. Finally, we also aim to use, and control, the variability of an *in silico* system emulating a human microbiome to uncover the possible interactions that exist in the community.

The present thesis contemplates the following four objectives, that are defined as follows:

Objective 1: Establish a definition of stability in human microbiome, and relate it with the health status of the host.

The lack of a formal definition of what constitutes a stable microbiota, along with the variability present in different scenarios that influence the human microbiome, are of critical importance to establish a connection between host's health status and stability. The availability of metagenomic data from follow-up studies and the existence of mathematical tools based on systemic variability analysis makes possible the theoretical exploration of this issue and its validation. The *first objective* aims to establish a formal definition of microbial stability with a model based on the Langevin differential equation, used to describe the

motion of elements given deterministic and stochastic components, and to apply this model to real microbiome data where the subjects were subjected to different perturbations or diseases.

Objective 2: Uncovering microbial interactions by noise induced experiments.

Microbial interactions is one of the main forces that shapes the communities. The type of interaction pattern determines the stability and robustness of the entire network, and thus, it is essential to be able to study them. However, the inference of bacterial interactions using temporal data is not trivial, with methods that rely on extensive data series, and which are limited by the models used. The *second objective* aims to uncover microbial interactions with noise-based experiments. As system perturbations help these models to have a better inference of the interactions, we propose a methodology by which one can alter the behaviour of a single specific bacterial species to study the interactions in the community.

Objective 3: Case study 1: time series analysis of the microbiota of children suffering from acute infectious diarrhoea and their recovery after treatment.

Acute infectious diarrhoea is one of the leading causes of mortality and morbidity in children worldwide, and its main cause is an infection with rotavirus. The gut microbiota seems to be implicated in the course of disease. Thus, understanding the dynamics of children recovering from this disease is of clinical interest. The *third objective* aims to study a cohort of 10 children suffering from acute infectious diarrhoea, and 6 healthy children, across a month of time with 5 time samplings. We aim to use the tool from the first objective to assess whether the children suffering the rotavirus infection were less stable than the healthy ones, and study this case from other points of view as the functional part of the human microbiome.

Objective 4: Case study 2: dynamics of oral microbiota and its relationship with oxidative stress.

The oral microbiome is in constant contact with the external world, and, it is still a highly stable system in time. Previously, it was thought that cavities and other oral diseases were caused by the presence or absence of specific bacteria such as *Streptococcus mutans*. Nevertheless, recent data seems to point toward a polymicrobial theory of disease in the mouth, defining the cause of oral disease as bacterial dysbiosis. The *fourth objective* aims to explore the dynamics of the oral microbiome and its correlation with oxidative stress markers that reflect the inflammatory processes in the oral cavity. For this purpose, we study 26 volunteers oral microbiomes during one month on a daily basis.

Material and Methods

3.1 Cohort and sample collection

In this section it will be explained the sample selection and subject enrolment that has been carried out in chapters I, III and IV. Briefly, the Taylor's law study in Chapter I was made of data publicly available from other research articles, with a total of 99 subjects with variation in their time-sampling; in the study of the gut microbiota of children with acute infectious diarrhoea, in Chapter III, we had a total of 16 children with 5 time points per child; finally, in the study of the salivary microbiota, in Chapter IV, we had a total of 26 volunteers with an average of 21 time points per voluntary. Chapter II did not have any sample selection procedure due to its *in silico* nature.

3.1.1 Sample selection in Chapter I

This study was based on data which was publicly available, except for the work regarding the Irritable Bowel Disease carried out by Durban *et al.* [73], which was directly provided by the authors. Although the material used can be seen in each reference, a table is provided where each data source can easily be found (Table 3.1). The separation between healthy and non-healthy periods was important, and the reasons are explained in the next lines.

We chose studies about relevant pathologies, containing metagenomic sequencing time data series of bacterial populations from humans in various healthy and non-healthy states. Only those subjects who had three or more time points of data available in databases were selected. The study by Caporaso *et al.* [47] was selected as it featured two healthy subjects measured over a very long time-span, with almost daily sampling. The study of Faith *et al.* [168] was selected given the BMI differences between subjects. Moreover, some of them followed diets which could be treated as system perturbations. Only those subjects whose BMI (body mass index) was normal or overweight were considered healthy. The study by Smith *et al.* [169] was selected for both the age of the patients and the rare disease.

Regarding kwashiorkor, we considered only the discordant twins and deemed subjects unaffected by kwashiorkor as being healthy in each pair of patients. The study by David *et al.* [170] was selected for its differential diets. The healthy period was considered to be the initial samples of each subject before starting the diet, while the remaining time points were considered as perturbations. Dethlefsen and Relman's work [120] was selected due to the interesting treatment of two antibiotic intakes of the same antibiotic by three different subjects. The healthy period was considered to correspond only to those times before any antibiotic treatment, whereas the periods during and after antibiotic intake were considered as perturbations. The work by David *et al.* [59] was selected due to the comprehensive longitudinal data that it provides, plus its complete metadata and the interesting events experienced by both subjects (an infection and a trip abroad). The healthy period was taken from time points before or after each event. Finally, we also considered a study from our group carried out by Durban *et al.* [73] in which the healthy subjects were considered as those who did not suffer from irritable bowel syndrome, while the patients who had this disease were taken as perturbations. In the end, we had a total of 99 subjects that were analysed.

3.1.2 Sample collection in Chapter III

We enrolled a total of 16 children, 6 of them with a healthy condition, and the other 10 suffering from acute infectious diarrhoea. All children were from 3 to 4 years old; this age range was selected because the microbiota is finally stabilised during this stage [39], and so that the differences caused by age disparity between groups are reduced. To eliminate potential biases associated with the intestinal microbiota during infancy, we selected children who were normally delivered, breastfed for at least 6 months, and have not received antibiotic therapy since birth. From every child enrolled, we collected samples at five different time points, namely, at 0, 3, 5, 10, and 30 days, for a total of 80 different samples. More metadata can be found at Table A.6 in the Appendix section. Children with acute infectious diarrhoea were treated with a probiotic made of *Saccharomyces boulardii*. The sample at day 0 was taken before probiotic treatment in the group of children with infectious diarrhoea. All children belonging to the infectious diarrhoea group received oral lyophilised *S. boulardii* CNCM I-745 (250 mg twice daily, ReFlor, Biocodex) for the first five days, specifically days 3 and 5 (day 0 was with no intervention), in addition to ORS and/or intravenous therapy. Rehydration and electrolyte replacement were done using hypoosmolar ORS (glucose 20 g; sodium 60 mmol/L; potassium 20 mmol/L; bicarbonate 30 mmol/L). The frequency and consistency of the stools were recorded. Before the collection of the children's stool and

Table 3.1 List of the data sources of all the studies appeared in Chapter I

Study		Reference number(s)	Database
Moving Pictures of Human Microbiome	[47]	4457768.3-4459735.3	MG-RAST
The Long-Term Stability of the Human Gut Microbiota	[168]	Without number	Personal web-site gordonlab.wustl.edu
Gut microbiomes for malawian twins pairs discordant for Kwashiorkor	[169]	ERP001928 (16S rRNA) ERP001911 (shotgun)	European Bioinformatics Insitute (EBI)
Diet rapidly and reproducibly alters the human gut microbiome	[170]	6248	MG-RAST
Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation	[120]	SRA020961	NCBI Short Read Archive database
Instability of the faecal microbiota in diarrhoea-predominant IBS	[73]	Without number	Provided by the author
Host life affects human microbiota on daily basis	[59]	ERP006059	European Bioinformatics Insitute (EBI)

data, consent was given by their parents. This study was registered in Clinical Trials website with the NCT number *NCT00683982*.

3.1.3 Sample collection in Chapter IV

Unstimulated whole mouth saliva was collected from 26 volunteers (13 males and 13 females) recruited from students of the Comenius University in Bratislava, Slovakia. Exclusion criteria included active periodontitis, untreated caries, oral pain or any known oral diseases, but also smoking and major chronic systemic diseases. Subjects taking any medications including over the counter available antioxidants were excluded as well. At least 2 mL of saliva was taken by spitting into sterile tubes daily for 30 days at the same time between 06:00 and 08:00 at least 30 minutes after toothbrushing. The volunteers were instructed not to eat in the morning before saliva collection. All samples were immediately frozen after collection and stored at -20°C until further processing.

3.2 16S rRNA gene sequencing

In order to characterise the bacterial and archaeal composition of the microbiota, the 16S rRNA gene of the samples from the children in Chapter III and the volunteers in Chapter IV was amplified and sequenced by high-throughput technologies. Despite the limitations of the 16S rRNA gene sequencing are well known, especially the biases due to the PCR amplification step or the sequencing-preparation procedures [171], and the lack of specificity to lower taxonomical levels than genus, new methods are appearing in the recent years which can resolve differences at one nucleotide [172, 173]. Nevertheless, it is still an approach that has clear strengths such as its cost per sequence, the robustness of the methods available to treat the sequences, or the ability to capture broad temporal properties [171]. Also, it seems that it recovers the diversity information with the same confidence, at least, than WGS (whole genome sequencing) methods [174]. Nevertheless, the scientific community is becoming aware that the field should move from the 16S sequencing to the WGS [175].

3.2.1 Nucleic acid extraction and sequencing

For the children in Chapter III, the hypervariable V4 region of the 16S rRNA gene was selected to be amplified, while in the volunteers from Chapter IV it was selected the hypervariable V3-V4 regions. It has been shown that V4-V6 are the most reliable hypervariable regions to recover most of the phylogenetic resolution that can be obtained with the use of the total length of the 16S rRNA gene [176]. All faecal samples were stored at -80°C until the

DNA extraction. All samples were weighed to extract the total DNA using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The extracted DNA was shipped under cold conditions to FISABIO in Valencia, Spain, for further analysis. More information about the participants is available in Supplementary Table A.6. Then, the hypervariable V4 region of the 16S rRNA gene was selected to be studied, with a paired-ends strategy (as shown in Figure 3.1) to increase the final length of the sequence, and therefore, improve the taxonomical assignation in further steps. Sequencing was performed on the MiSeq Illumina platform according to the manufacturer's specifications. Sequencing was conducted at the Sequencing and Bioinformatics Service of FISABIO Foundation. The *Eskisehir Osmangazi University Local Ethics Committee* from Turkey reviewed and approved the protocol.

The protocol followed to obtain the DNA from the salivary samples, in Chapter IV, was described in Polgarova *et al.* [177]. The work from Polgarova shows that the most suitable way of DNA extraction from salivary samples is without any treatment, directly from fresh salivary samples. Additionally, they describe that the most efficient extraction protocol for DNA isolation is the benzyl-chloride protocol and Chelex-100 extraction. The use of benzyl-chloride is explained in the work of Zhu *et al.*, [178], where they observe that it can destroy cell walls of plants, fungi and bacteria through its ability to react with -OH residues in polysaccharides. Then, the hypervariable regions V3 and V4 of the 16S rRNA gene were amplified using primers including Illumina adapter sequences (forward primer 5' - TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG -3', reverse primer 5' - GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC - 3') with Kapa HiFi HotStart polymerase, and the samples were multiplexed by 96 index combination and sequenced on MiSeq Illumina platform.

3.2.2 Sequence analysis

After raw sequences acquisition from Illumina platforms, an essential step is required to ensure the reliability of the bioinformatic analyses. This step consists of the **quality filtering** of the sequences, to withdraw the ones that could induce misleading results. The basic idea consists on the use of the quality information that the High-throughput sequencers give along with the nucleotide sequence, which is encoded in a fastq type of file [179], the standard sequence filetype nowadays. These files have a standard structure, with a four-line pattern repetition for each different sequence. The structure is defined as follows:

quality) to the character 126 (~, the highest quality). The quality information stored in each ASCII character can be “translated” into a numerical scale that usually ranges from 0 (low quality) to 41 (highest quality). The numerical values are often used as the real thresholds used by the programs.

$$Q = -10 \cdot \log_{10} \cdot P \quad (3.1)$$

In order to carry out the quality filtering, two different software tools were used (*FastX toolkit* [180], and *PRINSEQ* [181]), both with the same principles. The user specifies a quality threshold, commonly between 20 and 25, and those sequences that do not match this requirement are withdrawn for further analyses. It is also used a *sliding window* approach to withdraw fairly good sequences. This method takes a window of a certain number of nucleotides, and the user specifies the percentage of those nucleotides that have to be above the quality threshold to continue the analysis. For example, if the user specify a window of 20 bases, with a certain quality threshold and a 75% as the percentage, there must be at least 15 bases with a quality equal or higher than the threshold in the window to continue the analysis. Another important quality filter is to discard all the sequences that do not match a minimum length. After the quality filtering, the remaining sequences are supposed to be fine candidates of analysis, but some other quality filters must be applied of a different nature than the sequence quality itself such as the chimaera checking.

3.2.3 OTU characterisation

The sequences obtained and filtered are then analysed to assign them to operational taxonomic units (OTUs). Robert Sokal and Peter Sneath first introduced this term as a part of their *Principles of Numerical Taxonomy* [182], referring to a group of organisms that were being studied. However, it was in 2005 that the term acquired the meaning it has today in the field of microbiome [183]. Briefly, an OTU is a group of sequences clustered by a similarity threshold, usually at 97% in 16S rRNA gene. This similarity threshold has been proved as a reliable *proxy* for a *species* in microbiology (chapter 12 of *General Microbiology* book by Brock, [184]), although the concept of species in microbiology is still discussed and debated in the present days [185].

One of the most used tools for the OTU characterisation right now in the field is the open source pipe-line QIIME, which stands for Quantitative Insights Into Microbial Ecology [186]. It is a collection of functions written mainly in Python programming language. At the time of writing this thesis, the second version of QIIME has outperformed and substituted the first version of the program, a clear statement of how rapidly evolves the field of bioinformatics.

QIIME version 1.8, which was used to analyse the data from Chapter I and Chapter III, is still a robust and reliable way of inspecting the properties of a human-related microbiota. In both chapters, the master script `pick_open_reference_otus.py` was used to analyse the whole dataset, and it involves many steps that will be explained in the following lines.

1. OTU creation, which is made with the script `pick_otus.py`. As has been explained, an OTU is a group of sequences that shares a certain similarity. There exist several algorithms to cluster sequence by its similarity, like *uclust* or *usearch* [187]. Both algorithms create “seeds” of sequences which generate clusters based on a certain similarity threshold.
2. Pick representative sequence for each OTU, with script `pick_rep_set.py`. For any cluster created, a sequence will be chosen to be the representative of the entire group, and thus, fastening the subsequent analyses. It will be chosen as the centroid sequence in *uclust* or *usearch*, for example.
3. Assign taxonomy to each sequence, with `assign_taxonomy.py`. The representative sequences from the last step will be assigned to a taxonomy against a reference database, such as Greengenes [188] or Silva [189].
4. Sequence alignment, with `align_seqs.py`. The representative sequences will be aligned to build the phylogenetic tree, and for chimera filtering.
5. Chimera filtering, with `identify_chimeric_seqs.py`. A chimeric sequence is a sequence that belongs to two different taxonomic groups. In order to filter these spurious sequences, the algorithm ChimeraSlayer in QIIME is usually applied. The representative sequences that have been aligned will be compared to the aligned representative set of the original database.
6. Tree building, with `make_phylogeny.py`. A phylogenetic tree will be built from the representative sequences, that can be used in other analyses as the calculation of β -diversity, explained in section 3.5
7. Make the final OTU table, with `make_otu_table.py`. This final step involves the creation of a table with all taxonomic groups annotated and their abundance.

The approach used in Chapter IV had the same ideas behind, except that OTU clusters were built with the purpose of enhancing some *Streptococcus* species of interest.

3.3 Metagenomic sequence analysis

As has been said in the previous section, 16S rRNA gene sequencing has some critical limitations, and therefore, the scientific community is moving toward the use of WGS sequencing [175]. WGS has several advantages such as the exploration of the functions encoded in a metagenome, or the ability to reach deeper taxonomical levels as strain [44, 190]. Although the computational requirements increase notably with this kind of studies, a considerable amount of software tools have been developed up to the present day to facilitate the analyses [191].

In the work of Chapter I, the samples from the study by Smith *et al.* [169] were also sequenced by shotgun metagenomic by HiSeq Illumina sequencing platform. We analysed that dataset to demonstrate that our methodology was robust independently of the origin of the metagenomic samples.

Raw sequences were analysed with LMAT (Livermore Metagenomics Analysis Toolkit) software package [192] (version 1.2.4, with Feb'15 release of database *LMAT-Grand*). We used the “Grand” database[193], release Feb'15, provided by the LMAT team, where “Grand” refers to a vast database that contains k-mers from all viral, prokaryote, fungal and protist genomes present in the NCBI database, plus Human reference genome (hg19), plus GenBank Human, plus the 1000 Human Genomes Project (HGP) (this represent about 31.75 billion k-mers occupying 457.62 GB) [193].

3.4 Clinical measurements

In Chapter IV, five different oxidative stress markers were measured. Oxidative stress can be measured by estimating oxidative damage to lipids (lipid peroxidation) and proteins, or by quantifying the capacity to resist oxidative damage (antioxidant capacity). Lipid peroxidation is a reaction of polyunsaturated lipids with reactive oxidative species (ROS), producing lipid hydroperoxides, that can react with DNA and other macromolecules [194]. Proteins are among the major targets of ROS due to their centrality in the cell, both for functionality importance and for number [195]. The oxidised proteins are either destroyed in proteasomal or lysosomal pathways, or accumulated in cellular compartments. The antioxidant status is a global measure that explains the total antioxidant capacity of a sample, and thus, is often desirable over the measurement of individual oxidative markers [196]. In the work of Chapter IV, lipid peroxidation was quantified by measuring thiobarbituric acid reacting substances (*TBARS*). Either carbonyl or oxidative stress can cause protein damage. Thus, advanced glycation end products (*AGEs*) and advanced oxidation protein products (*AOPP*) were herein

quantified as a carbonyl stress marker and as an oxidative stress marker, respectively. The capacity to resist oxidative damage was measured by total antioxidant capacity (TAC) and the ferric reducing ability of saliva (FRAS). The oxidative stress markers were quantified using plate reader Safire II (Tecan, Austria) as described by Banasova *et al.* [197].

More specifically, the marker of lipid peroxidation, thiobarbituric acid reactive substances (TBARS), was assessed according to Behuliak *et al.* [198]. Twenty μL of samples and standards (1,1,3,3 - tetraethoxypropane) were added into a 96 well plate. Thereafter, 30 μL of water, 20 μL of thiobarbituric acid together with 20 μL of glacial acetic acid were added and whole mixture was incubated at 95°C for 45 minutes. Afterwards, 100 μL of n-butanol was added, and plate was centrifuged at 2,000 x g, at 4°C for 10 minutes. Seventy μL of the upper organic phase was transferred into a new microtiter plate and, subsequently, the fluorescence was measured at $\text{ex} = 515 \text{ nm}$ and $\text{em} = 535 \text{ nm}$. Advanced oxidation protein products (AOPP) were measured as a marker of protein damage [199]. Chloramine T mixed with potassium iodide was used for preparation of the calibration curve. Two hundred μL of samples or standard were transferred onto a 96 well plate for analysis. Later, 20 μL glacial acetic acid was added to both standards and samples and incubated on a plate shaker (500 rpm) for 2 minutes. Then absorbance was measured at 340 nm. Advanced glycation end products (AGEs) were assessed as a marker of carbonyl stress [200]. Twenty μL of samples together with standards (AGE-BSA) were placed into a 96 well plate and diluted with phosphate buffer saline (pH = 7.2). The fluorescence was measured at $\text{ex.} = 370 \text{ nm}$ and $\text{em.} = 440 \text{ nm}$. The analysis of antioxidant status was represented by the assessment of the total antioxidant capacity (TAC) according to Erel *et al.* [196]. Twenty μL of samples and standards (Trolox) were mixed with 200 μL of acetate buffer (pH = 5.8) and measured at 660 nm as a blank. Thereafter, 20 μL of 2, 2 azino - bis (ethylbenzthiazoline - 6 - sulphonic acid) was added. Plate was incubated at room temperature for 5 minutes and the absorbance at 660 nm was measured. Ferric reducing antioxidant power of saliva (FRAS) was the next marker of antioxidant status. Ferrous sulphate was used as a standard for the construction of the calibration curve [201]. Two hundred μL of prewarmed (37 °C) FRAP reagent, composed of tripyridyl - s - triazine, $\text{FeCl}_3 \times 6\text{H}_2\text{O}$, acetate buffer (pH = 3.6) and water, was added to the assay plate. Initial absorbance at 593 nm was measured as a blank. Thereafter, the samples and standards were added and the absorbance was measured at 593 nm.

3.5 Diversity analysis

3.5.1 A primer of diversity in ecology

In the Ecology field, the diversity and richness are among the most basic and fundamental analyses that can be made to study essential properties of the ecosystem. There is a fundamental difference between the concept of **diversity** and **richness**, as the former only takes into account the number of different organisms in the study, and the latter also considers their abundance. It is usually the number of individuals per species which is studied, but other quantities can be treated in diversity analyses such as individuals per genera, haplotypes, functional types or biomass. There exists three main classes of diversity defined by Whittaker in his seminal work: α -diversity, β -diversity, and γ -diversity [202]. Shortly, α -diversity, explains the diversity of a single population. It is a direct measure of how diverse a sample is, increasing its value when we can find a higher number of species in the population. On the other hand, β -diversity indexes explain the difference between two populations. It can be managed as a similarity measure, or as a distance, and it says how close are two different samples between each other. And, lastly, γ -diversity, is the total diversity that it is found within a landscape, the total number of species, and it is the consideration of both α and β diversity, which can be either additive or multiplicative (eq. 3.2).

$$\gamma = \alpha \cdot \beta \quad (3.2)$$

In fact, one of the first definitions of β -diversity was deduced from the multiplicative expression from above (the original from Whittaker's work), with the form of $\beta = \frac{\gamma}{\alpha}$.

There are some fundamental properties that both α and β diversity indexes must have: α and β should be free to vary independently; γ must be determined entirely by α and β , and α can never be greater than γ [203]. Nevertheless, the concept of γ -diversity is seldom calculated in microbiome research, and thus, we have focused on the calculation of the values of α and β -diversity in our studies, as they give us valuable information about the composition shifts of the microbiotas.

3.5.2 α -diversity indexes and calculation

As has been previously described, α -diversity is the measure of the diversity in one single sample or ecosystem. Usually, this quantity is calculated as the Shannon entropy index, which has the mathematical form of the equation 3.3, where p_i is the proportion of the i species (of a total of R species). It measures the uncertainty of a system, being higher when

the uncertainty is higher. In the case that almost all the abundance concentrates on only one species, H' approaches 0 (there is not almost uncertainty); and in the case one have several species with similar abundances, H' approaches higher values, which will also depend on the number of elements we have in the population.

$$H' = \sum_{i=1}^R p_i \cdot \ln(p_i) \quad (3.3)$$

Pielou's evenness can be computed from the Shannon index (eq. 3.4), by dividing the current Shannon value by its maximum possible value (eq. 3.5), where R is the total number of species. It estimates the evenness of a population, ranging from 0 to 1. If there is a similar amount of species, the evenness will be near 1, but if there are significant differences between the abundances of the species, the evenness will approach to 0.

$$j' = \frac{H'}{H'_{max}} \quad (3.4)$$

$$H'_{max} = - \sum_{i=1}^R \frac{1}{R} \ln\left(\frac{1}{R}\right) = \ln(R) \quad (3.5)$$

A problem that arises from this concept is that there are several ways to calculate the α -diversity, such as the Shannon index, the Simpson index, the Gini-Simpson index or the Renyi entropy to give some examples. The problem with all these forms of α -diversity is that they are not comparable between them as non-linearities appear due to their mathematical expressions, or also because they give differential importance to the most or least abundant species. To solve this issue, Lou Jost proposed the use of the numbers equivalent, or the effective number of species [203]. The idea behind this concept is to convert the non-linear indexes to a linear estimator, and thus, helping to infer the real changes in diversity. Another interesting property of the numbers equivalent is that they can be calculated from all the diversity estimators mentioned above making them comparable. This concept has also been named as the *true diversity*. The expression that transforms between Shannon index and its true diversity is $H_t' = \exp(H')$.

In this thesis, it has been used both Shannon entropy and its true diversity estimator, and Pielou's evenness. These values have been calculated with the abundance tables obtained in Chapter III and Chapter IV, with R programming language and the function "diversity" in *vegan* package (v. 2.4-3) [204]. Differences in α -diversities were studied by performing permutations over the Wilcoxon signed-rank test with a significance level of 0.05, calculated with the *coin* R package [205]. In the case of multiple comparisons, we applied false discovery rate (FDR) in Chapter III, or Holm correction in Chapter IV to avoid type I errors.

3.5.3 β -diversity and ordination plots

Whitaker defined his β -diversity as “the extent of change in community composition, or degree of community differentiation, in relation to a complex-gradient of the environment, or a pattern of environments” in his seminal work in 1960 [202]. Despite de vague definition that Whitaker proposed, the concept of β -diversity is of great importance and has been extensively studied all over the past years, where many interpretations have arisen [203, 206–208]. In the practical review from Anderson *et al.*, [206], they give two primary definitions of β -diversity, one based on the directional turnover of a community, and the other based on the variation in community structure within a given spatial or temporal extent. The latter definition, implied in the first definition of Whitaker, is the one that is often used in microbiome studies, and its main objective is to define the differences between two communities. These differences are encoded in metrics that can be understood as the degree of similarity or a distance between the different samples, which commonly are normalised between 0 and 1.

In the projects of Chapter III and Chapter IV, β -diversity was estimated with a combination of QIIME version 1.8 with the R programming language and the package *vegan*. The UniFrac distance [209] was selected in Chapter III, and Bray-Curtis dissimilarity [210] was used in Chapter IV. UniFrac distances were calculated with the QIIME script `beta_diversity.py`, and Bray-Curtis dissimilarity was calculated in R with *vegan* package.

Once the distances or dissimilarities metrics have been computed, they can be used to place every sample in the space of 2 or 3 dimensions, observing if they cluster together or not depending on the explanatory variables of the problem. The main idea of these multivariate ordination techniques is to reduce the complexity of the problem to a limited set of dimensions, helping the interpretation of the results. It is the case, for example, of the Principal Components Analysis (PCA) or the Multidimensional Scaling (MDS), which take the general idea of extracting the eigenvalues and eigenvectors of a problem to see its principal directions. In this case, we have used the Canonical Correspondence Analysis (CCA) in Chapter III because it extracts the significant gradients among combinations of explanatory variables in a dataset. A more generalist approximation, the Nonmetric Multidimensional Scaling (NMDS), is used in Chapter IV, which attempts to represent, as closely as possible, the pairwise dissimilarity between objects in a low-dimensional space.

We applied the ADONIS test, within the *vegan* package, a multivariate ANOVA based on dissimilarity tests, to assess the variables that were statistically significant in the separation of the samples in two dimensions in Chapter III. The Procrustes test (*vegan*) was performed to test whether the datasets with different OTU numbers produced the same ordination of samples in the NMDS, in Chapter IV.

3.6 Biomarker discovery

In order to determine the features, like organisms, that most likely explain differences between classes, we used a Linear discriminant analysis (LDA) with effect size (LEfSE). This analysis has been made with the algorithm available from the Galaxy software package of Huttenhower laboratory [211] to identify the specific taxa that served as the biomarkers in Chapter III. LEfSe first determines the significant differences in taxa composition between groups by using non-parametric factorial Kruskal-Wallis sum-rank test. Then, LEfSe uses LDA to estimate the effect size of each differentially abundant feature. We fixed a cut-off α value of <0.05 for the Kruskal-Wallis test, and for the bacterial taxa that had significant differences between samples, we fixed a threshold of 4.0 in the logarithmic LDA score to consider it a discriminative feature.

3.7 Functional analysis

The single analysis of the taxonomic groups that are present in a given microbiota is an incomplete picture of the whole scene. A multi-omic approach is desirable, where it should include information about the functional capability of a microbiome together with other kinds of data as transcriptomics and proteomics. By this procedure, it can be captured not only who is there but also what are they doing [44, 212–214]. Another kind of analysis that can help is the metabolic reconstruction of the species of a specific microbiota, a demanding task that has been only possible in the recent times [215]. Nevertheless, to be able to study the microbiome in such depth, the use of whole genome sequencing is almost mandatory. In the case only sequences from 16S rRNA gene are available, there exist some useful algorithms that take advantage of databases and taxonomical relatedness. For the functional content prediction in Chapter III, we used the PICRUSt software from Huttenhower laboratory [216], a computational approach to predict the functional composition of a metagenome using marker gene data and a database of reference genomes. Using an algorithm for the reconstruction of the ancestral state, the software predicts the gene families present in a metagenome and, thus, helps to provide a functional perspective. We used default parameters and the KEGG database for inference [217–219]. We extracted the information at levels 2 and 3 in KEGG hierarchies, as we have more specificity when we increase the level of the hierarchy. To assess the differences in function abundances, we applied the Kruskal-Wallis test, with the α value fixed at 0.01.

3.8 Correlations and interactions inference

Correlations in Chapter II and Chapter IV have been plotted using R packages *corrplot* [220] and *beeswarm* [221]. Pearson's and Spearman's temporally direct correlations and Local Similarity Analysis for lagged correlations were used to assess correlations between bacterial OTUs in oral cavity of each volunteer individually taking into account the temporal dimension. Local Similarity Analysis checks the existence of a delay in the correlations between all possible combinations of variables. Pearson's, Spearman's, and LSA correlation coefficients were computed using eLSA Python package from Xia *et al.*, [145], with default parameters. This package computes permutation tests that are helpful when multiple hypothesis tests are being studied in this kind of problems. The correlation values >0.3 and <-0.3 were then filtered by their p-value at a level of 0.05 for statistical significance. In addition, interactions between the 15 most abundant OTUs per sample were calculated with LIMITS algorithm from Fisher and Mehta [114] in Wolfram Mathematica software (version 11.0) [222]

3.9 Temporal analysis

The temporal analysis of the microbiome is the central part of the present thesis, and thus, it will be explained here with detail. It is one of the main results of the Chapter I, and then it is used as a tool in Chapter II, Chapter III and Chapter IV. Taylor's law is an ubiquitous law that is present in many natural systems, as ecosystems, and it has the mathematical expression of a power law:

$$\sigma_i = V \cdot x_i^\beta, \quad (3.6)$$

where σ_i is the variance of the species i in the population, x_i is the mean abundance of the species i , V is the amplitude of variation, and β is the scaling index.

Our main goal was to fit this law to our data in a robust way, which is explained in the following sections.

3.9.1 X-weighted power-law fit

When fitting the power-law of std vs mean, we can take into account that every mean has uncertainty and estimate it for a sample size n by the SEM (*Standard Error of the Mean*). Here, the uncertainties affect the independent variable, so the fit is not so trivial as a Y-weighted fit, where the uncertainties affect the dependent variable. A standard approach to do this fit is: a) invert the variables before applying the weights, b) then perform the weighted

fit, and finally, c) revert the inversion. This method is deterministic, but the approximate solution worsens with smaller coefficients of determination. To overcome this limitation, we have developed a stochastic method by using a bootstrapping-like strategy that avoids the inversion and is applicable regardless of the coefficient of determination.

The basic idea of bootstrapping is that inference about a population from sample data (sample \rightarrow population) can be modelled by resampling the sample data and performing inference on (resample \rightarrow sample) [223]. To adapt this general idea to our problem, we resample the x-data array using its errors array. That is, for each replicate, a new x-data array is computed based on:

$$x_i^* = x_i + v_i \quad (3.7)$$

where v_i is a Gaussian random variable with mean $\mu_i = 0$ and standard deviation $\sigma_i = \text{SEM}_i$, as defined previously. For each replicate, a complete un-weighted power-law fit is performed, where to choose between fitting power laws ($y = Vx^\beta$) using linear regression on log-transformed (LLR) data versus non-linear regression (NLR) we mainly follow *General Guidelines for the Analysis of Biological Power Laws* [224]. The parameters of the X-weighted fit are then estimated by averaging through all the replicate fits performed, and their errors are estimated by computing the standard deviation also for all the fits. At the end of each step, the relative error is calculated by comparing the fit parameters estimation in the last step with the previous one. Finally, both the coefficient of determination of the fit and the coefficient of correlation between the fit parameters are estimated by averaging.

3.9.2 Standardisation

In order to properly show all the studies under common axes, we decided to standardize the Taylor parameters from section 3.9.1 using the group of healthy individuals for each study, in Chapter I and Chapter III per separate. With this approach, all the studies can be visualised in a shared plot with units of Taylor-parameters standard-deviation on their axes.

For a Taylor parameter, e.g. V , the estimate of the mean (\hat{V}) for the healthy subpopulation, composed of h individuals, is:

$$\hat{V} = \frac{1}{W_1} \sum_{i=1}^h V_i \omega_i = \sum_{i=1}^h V_i \omega_i \quad (3.8)$$

as $W_1 = \sum_i^h \omega_i = 1$, since ω_i are normalized weights calculated as:

$$\omega_i = \frac{\frac{1}{\sigma_{V_i}^2}}{\sum_i^h \frac{1}{\sigma_{V_i}^2}} \quad (3.9)$$

being σ_{V_i} the estimation of the uncertainty in V_i obtained together with V_i from the X-weighted power-law fit described in Section 3.9.1, for healthy individuals.

Likewise, the estimation of the standard deviation for the healthy population ($\hat{\sigma}_V$) is:

$$\hat{\sigma}_V = \sqrt{\frac{1}{W_1 - \frac{W_2}{W_1}} \sum_{i=1}^h \left[\omega_i (V_i - \hat{V})^2 \right]} \quad (3.10)$$

being $W_2 = \sum_i^h \omega_i^2$, which finally yields to:

$$\hat{\sigma}_V = \sqrt{\frac{1}{1 - \sum_i^h \omega_i^2} \sum_{i=1}^h \left[\omega_i (V_i - \hat{V})^2 \right]} \quad (3.11)$$

3.9.3 Langevin Model

One of the main parts of Chapter I is the model that governs the temporal variation of the microbiota, based on the Langevin equation. This mathematical model has been used as the solution of the Brownian motion [225], and its solution implies the existence of a phase diagram with important dynamical properties. We model the microbial abundances across time along the lines of Blumm *et al.* [162]. The dynamics of taxon relative abundances is described by the Langevin equation:

$$\dot{x}_i = F_i \cdot x_i^\alpha + V \cdot x_i^\beta \xi_i(t) - \phi(t) \cdot x_i, \quad (3.12)$$

where F_i captures the fitness of the taxon i , V corresponds to the noise amplitude and $\xi_i(t)$ is a Gaussian random noise with zero mean $\langle \xi_i(t) \rangle = 0$ and variance uncorrelated in time, $\langle \xi_i(t) \xi_i(t') \rangle = \delta(t' - t)$. The function $\phi(t)$ ensures the normalization at all times, $\sum x_i(t) = 1$, and corresponds to $\phi(t) = \sum F_i x_i^\alpha + \sum V x_i^\beta \xi_i(t)$. The temporal evolution of the probability that a taxon i has a relative abundance $x_i(t)$, $P(x_i, t)$, is determined by the Fokker-Planck equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x_i} [(F_i \cdot x_i^\alpha - \phi(t) \cdot x_i) \cdot P] + \frac{1}{2} \frac{\partial^2}{\partial x_i^2} (V^2 \cdot x_i^{2\beta} \cdot P). \quad (3.13)$$

The microbiota evolves towards a steady-state with a time-independent probability depending on the values of α , β , F_i and V . For $\alpha < 1$ (otherwise, systems are always unstable), the steady-state probability may be localized in a region around a preferred value or broadly distributed over a wide range, depending on whether the fitness F_i dominates or is overwhelmed by the noise amplitude V . The steady-state solution of the Fokker-Planck equation is given by:

$$P_0(x_i) = C_{ne}(\alpha, \beta, F_i, V) \cdot x_i^{-2\beta} \cdot \exp \left[\frac{2F_i}{V^2} \frac{x_i^{1+\alpha-2\beta}}{1+\alpha-2\beta} - \frac{\phi_0}{V^2} \frac{x_i^{2-2\beta}}{1-\beta} \right] \quad \text{if } 2\beta \neq 1 + \alpha,$$

$$P_0(x_i) = C_e(\alpha, \beta, F_i, V) \cdot x_i^{\frac{2F_i}{V^2}-2\beta} \cdot \exp \left[\frac{\phi_0}{V^2} \frac{x_i^{2-2\beta}}{1-\beta} \right] \quad \text{if } 2\beta = 1 + \alpha,$$

where $\phi_0 = (\sum_i F_i^{1/(1-\alpha)})^{1-\alpha}$ and C_{ne} and C_e are integrals that should be solved numerically for the parameters of interest. The ordered phase happens when the solution has a maximum in the physical interval ($0 < x_i < 1$). For larger V , the transition to a disordered phase happens when the maximum shifts to the unphysical region $x_i < 0$, which sets the phase transition region $V(\alpha, \beta, F_i)$. The phase transition region can be calculated analytically in particular cases:

$$F_i^2 = 4\beta\phi_0V^2 \quad \text{if } \beta = \alpha \neq 1,$$

$$F_i = \beta V^2 \quad \text{if } 2\beta = 1 + \alpha,$$

where the first case, simplifies to $F = 3V^2$ if $\beta = 0.75$ and the fitness of this taxon dominates in ϕ_0 . In many physical systems (Brownian motion is the classical example [226]), the two terms of the Langevin equation are related. The *Fluctuation Dissipation Theorem* states a general relationship between the response to an external disturbance and the internal fluctuations of the system [227]. The theorem can be used as the basic formula to derive the fitness from the analysis of fluctuations of the microbiota, assuming that it is in equilibrium (the ordered phase).

3.9.4 Rank Stability Index

The Rank Stability Index (RSI), used in Chapters I, III and IV, is shown as a percentage in a separate bar on the right of the rank matrix plot (as for example in Fig. 4.5). The RSI is strictly 1 for an element whose range never changes over time, and is strictly 0 for an element

whose rank oscillates between the extremes from time to time. So, RSI is calculated, per element, as 1 less the quotient of the number of true rank hops taken between the number of maximum possible rank hops, all powered to p :

$$\text{RSI} = \left(1 - \frac{\text{true rank hops}}{\text{possible rank hops}} \right)^p = \left(1 - \frac{D}{(N-1)(t-1)} \right)^p \quad (3.14)$$

where D is the total of rank hops taken by the studied element, N is the number of elements that have been ranked, and t is the number of time samples. The power index $p = 4$ is arbitrarily chosen to increase the resolution in the stable region.

3.9.5 Temporal analysis of V and β

In Chapter I, we have studied the time dependence of the variability V and power law index β by using a sliding window approach. For this purpose, a code was written in R programming language in order to generate the data. The total number of time points was divided into subsets of five points, where the following subset was defined by adding the next sampling time and eliminating the earliest one. Then, both Taylor's parameters were calculated for each subset.

3.10 Lotka-Volterra model with an stochastic component

We model the dynamics of our mock microbiome in Chapter II with the system of equations

$$\frac{d}{dt}x_i(t) = \mu_i x_i(t) + x_i(t) \sum_{j=1}^L M_{ij} x_j(t) + Bx_i^b \cdot dW, \quad (3.15)$$

based on the generalised Lotka-Volterra (gLTV) model extended with an stochastic component described as a Wiener process. It measures the temporal evolution of a species x_i , which depends on its growth rate (μ_i) and the interactions with the other species parameterised in the matrix M_{ij} . Ecological networks satisfy three properties: scale-free, small world and modularity [228]. We have built the interaction matrix with the algorithm created by Klemm and Eguíluz [229]. We used this algorithm, written in R package seqtime [230], where we have specified a connectivity of 0.1. The matrix M_{ij} is represented as Figure 5.1 in Chapter II. The amplitude of the stochastic component is based on the observation that microbial populations follow the Taylor's law [231], an empirical law, previously observed in ecology, that relates the variance in the number of individuals of a given species in a population with

its mean value by the power law ($\sigma_i = Bx_i^b$). This relationship is based on the results of Chapter I. In the equation, B is the amplitude of fluctuations and b is the scaling index.

We numerically solve the system of equations with MATLAB [232], using the SDETool package [233]. In the simulations, all μ_i are always equal to 1. We run several simulations with different values of B and b in the dynamical equations. We start all the simulations discussed in Chapter II with $B = 0.05$, and $b = 1.5$. The time duration was limited to 100 (in arbitrary units), with a δt equal to 10^{-2} , i.e., 10^4 time points per simulation. A non-negativity condition was imposed to correctly characterise abundances. The initial conditions were given by the steady state solution of the gLV equations.

For the data analysis, we use normalised abundances at all times, as usual in microbiome studies. We compute Pearson correlation matrices with 1000 points from the simulated time series described above. We plot correlation matrices with *corrplot* R package [220].

3.11 Graph inference and error measurement

For the interaction matrix inference, we use a general regression model by solving a least squares optimisation problem with the following expression:

$$\arg \min_{a_{ij}, \lambda} \left(\sum_k \sum_i (|x_i(t+k) - a_{ij}x_j(t+k-1)|^2 + \lambda |\sum_j a_{ij}|) \right) \quad (3.16)$$

Here, x_i is the vector of species, and a_{ij} is the inferred interaction matrix from this regression (different from the original interaction matrix M_{ij}). We generalise the process assuming that it can be applied at any k times. λ term is used as a Lagrange multiplier to force sparsity to our inferred interaction matrix. We assume that the diagonal of a_{ij} has to be negative, and that the other elements of the matrix range between -0.5 and 0.5 according to our original matrix. In the end, we get a_{ij} and λ that minimises the expression above by an optimisation problem encoded in MATLAB, assuming it has a convex solution. For this purpose, we use *CVX* modelling system for convex optimisation [234].

Given the real interaction matrix M_{ij} , the error can be computed as the following matrix norm:

$$\|M_{ij} - a_{ij}\|_2^2, \quad (3.17)$$

where a_{ij} is the matrix formed by the vectors obtained from the optimisation problem.

Then, the error for a specific species x_i , which is being modified by a noise-induced experiment or by a pulse, can be measured as the vector norm:

$$\|M_i - a_i\|_2^2 \quad (3.18)$$

Chapter I - Temporal stability of the human microbiome

4.1 Abstract

Animal microbiota (including human microbiota) plays an essential role in keeping the physiological status of the host healthy. Research seeks greater insight into whether changes in the composition and function of the microbiota are associated with disease. We analysed 16S rRNA, and shotgun metagenomic sequencing (SMS) published data pertaining to the gut microbiota of 99 subjects monitored over time. Temporal fluctuations in the microbial composition revealed significant differences due to factors such as dietary changes, antibiotic intake, age and disease. This work shows that a fluctuation scaling law can describe the temporal changes in the gut microbiota. This law estimates the temporal variability of the microbial population and quantitatively characterises the path toward disease via a noise-induced phase transition. Estimation of the systemic parameters may be of clinical utility in follow-up studies and have more general applications in fields where it is important to know whether a given community is stable or not.

Original publication: Martí JM*, Martínez-Martínez D*, Rubio T, Gracia C, Peña M, Latorre A, Moya A, P. Garay C. 2017. Health and Disease Imprinted in the Time Variability of the Human Microbiome. *mSystems* 2:e00144-16.

* Equal contribution

4.2 Material and Methods

A general view of the material and methods used in this Chapter has been explained in the Material and Methods Chapter. Nevertheless, some important points have been summarised here to facilitate the reading of this Chapter, and more concrete details have been added when needed.

4.2.1 Model

The mathematical expression that governs the microbial abundances across time is derived of the work of Blumm *et al.* [162], and it is based on the Langevin's equation:

$$\dot{x}_i = F_i \cdot x_i^\alpha + V \cdot x_i^\beta \xi_i(t) - \phi(t) \cdot x_i, \quad (4.1)$$

where F_i captures the fitness of the taxon i , V corresponds to the noise amplitude and $\xi_i(t)$ is a Gaussian random noise with zero mean $\langle \xi_i(t) \rangle = 0$ and variance uncorrelated in time, $\langle \xi_i(t) \xi_i(t') \rangle = \delta(t' - t)$. The function $\phi(t)$ ensures the normalisation at all times, $\sum x_i(t) = 1$, and corresponds to $\phi(t) = \sum F_i x_i^\alpha + \sum V x_i^\beta \xi_i(t)$. The temporal evolution of the probability that a taxon i has a relative abundance $x_i(t)$, $P(x_i, t)$, is determined by the Fokker-Planck equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x_i} [(F_i \cdot x_i^\alpha - \phi(t) \cdot x_i) \cdot P] + \frac{1}{2} \frac{\partial^2}{\partial x_i^2} (V^2 \cdot x_i^{2\beta} \cdot P). \quad (4.2)$$

The steady-state solution of the Fokker-Planck equation is given by:

$$P_0(x_i) = C_{ne}(\alpha, \beta, F_i, V) \cdot x_i^{-2\beta} \cdot \exp \left[\frac{2F_i}{V^2} \frac{x_i^{1+\alpha-2\beta}}{1+\alpha-2\beta} - \frac{\phi_0}{V^2} \frac{x_i^{2-2\beta}}{1-\beta} \right] \quad \text{if } 2\beta \neq 1 + \alpha,$$

$$P_0(x_i) = C_e(\alpha, \beta, F_i, V) \cdot x_i^{\frac{2F_i}{V^2}-2\beta} \cdot \exp \left[\frac{\phi_0}{V^2} \frac{x_i^{2-2\beta}}{1-\beta} \right] \quad \text{if } 2\beta = 1 + \alpha,$$

where $\phi_0 = (\sum_i F_i^{1/(1-\alpha)})^{1-\alpha}$ and C_{ne} and C_e are integrals that should be solved numerically for the parameters of interest. The ordered phase happens when the solution has a maximum in the physical interval ($0 < x_i < 1$). For larger V , the transition to a disordered phase happens when the maximum shifts to the unphysical region $x_i < 0$, which sets the phase transition region $V(\alpha, \beta, F_i)$. The phase transition region can be calculated analytically in particular cases:

$$\begin{aligned}
 F_i^2 &= 4\beta\phi_0V^2 \quad \text{if } \beta = \alpha \neq 1, \\
 F_i &= \beta V^2 \quad \text{if } 2\beta = 1 + \alpha,
 \end{aligned}$$

where the first case, simplifies to $F = 3V^2$ if $\beta = 0.75$ and the fitness of this taxon dominates in ϕ_0 .

4.2.2 Cohort

We did chose studies about relevant pathologies, containing metagenomic sequencing time data series of bacterial populations from humans in various healthy and non-healthy states. We selected healthy people from the study by Caporaso *et al.* [47]; people with different BMI from the study of Faith *et al.* [168]; discordant children twins which had kwashiorkor disease from the study of Smith *et al.* [169]; people that followed animal or plant-based diets from the work of David *et al.* [170]; three patients that had two antibiotic intakes, from Dethlefsen and Relman's work [120]; a long temporal data from two volunteers that had traveled abroad, from David *et al.*'s work [59]; and finally, people that had IBS, from Ana Durban *et al.*'s work [73]. In the end, we had a total of 99 subjects that were analysed. For more detailed information, please, see Section 3.1.1 and Table 3.1.

4.2.3 16S rRNA sequences processing

Reads from the selected studies were first quality filtered using the FastX toolkit [180], with the command `fastq_quality_filter`, allowing only those reads which had more than 25 of quality along the 75% of the complete sequence. 16S rRNA reads were then clustered at 97% nucleotide sequence identity (97% ID) into operational taxonomic units (OTUs) using QIIME package software [186] (version 1.8). We followed open reference OTU picking workflow in all cases. The clustering method used was *uclust*, and the OTUs were matched against Silva database [189] (version 111, July 2012), and were assigned to taxonomy with an *uclust*-based consensus taxonomy assigner. The parameters used in this step were: similarity 0.97, prefilter percent id 0.6, max accepts 20, max rejects 500.

4.2.4 Metagenomic sequences processing

Shotgun metagenomic sequences were analyzed with LMAT (Livermore Metagenomics Analysis Toolkit) software package [192] (version 1.2.4, with Feb' 15 release of the LMAT-Grand database). LMAT was run using a Bull shared-memory node belonging to the team's

HPC (high-performance computing) cluster. It was equipped with 32 cores (64 threads available using Intel Hyper-Threading Technology) as it has two Haswell-based Xeons (22 nm technology), the E5-2698v3@2.3 GHz, sharing half a terabyte of DRAM memory. This node is also provided with a PCIe SSD card as NVRAM, the Micron P420m HHHL, with 1.4 TB, and 750000 reading IOPS, 4 KB, achieving 3.3 GB/s. The computing node was supplied with a RAID-0 (striping) scratch disk area. We used the “Grand” database, released in Feb’15, provided by the LMAT team, where “Grand” refers to a huge database that contains k-mers from all the viral, prokaryote, fungal and protist genomes present in the NCBI database, plus the Human reference genome (hg19), plus GenBank Human, plus the 1000 Human Genomes Project (HGP) (this represents about 31.75 billion k-mers occupying 457.62 GB) [193]. Before any calculations were made, the entire database was loaded into the NVRAM. With this configuration, the observed LMAT sustained sequence classification rate was 20 kbp/s/core.

4.2.5 Taxa level robustness

We selected genus as the taxonomic level for the subsequent steps of our work. To ensure that there were no crucial differences between adjacent taxonomic levels which could still be of relevance after standardisation, we tested two different datasets. In the former, the antibiotics study [120] with 16S data, we tested the differences between genus and family levels. The latter dataset tested was the kwashiorkor discordant twins study [169] for both genus and species taxonomic levels. The Appendix Figures A.1 (overview) and A.2 (detail) plot the comparison between studies (and so, 16S and WGS) and between adjacent taxonomic levels.

4.3 Results

4.3.1 Gut microbiota follows the Taylor’s law

Microbiome temporal variability was analysed to extract the global properties of the system. As systematic errors plague fluctuations in total counts, we worked on the temporal variability of relative abundances for each taxon. The first finding was, without exception, that changes in the relative abundances of taxa followed a ubiquitous pattern, known as the fluctuation scaling law [235] or Taylor’s power law [236]. In other words, the microbiota of all detected taxa followed $\sigma_i = V \cdot x_i^\beta$, a power law dependence between the mean relative abundance x_i and the dispersion σ_i . The law seems to be ubiquitous, spanning even to six orders of magnitude in the observed relative abundances. As shown in Figure 4.1, where V corresponds to the y-intercept and β is the slope of the fit, the most abundant species were less volatile in

relative terms than the less abundant ones. The fit to the power law was always robust ($R^2 > 0.88$) and did not depend on microbiome condition. The power law (or scaling) index β and the variability V (hereafter Taylor's parameters) appear to be correlated with community stability. Accordingly, we assume that Taylor's parameters behave as proxies for stability. On the one hand, β is a scaling index that provides information about the statistical properties of the ecosystem. If it is $1/2$, the system behaves like a Poisson distribution. If β is 1, the system behaves like an exponential distribution. Generally speaking, metagenomes undergo time-course variations with β between these two universal classes. In our study, the fact that β was less than 1 indicates that the most abundant taxa in the microbial community were less susceptible to perturbations than the other taxa. On the other hand, the variability V is a direct estimator of the amplitude of fluctuations over time. V represents the maximum variability attainable by a hypothetically dominant genus (with relative abundance close to 1). It is an important parameter that characterises the type of system. If V is small, the ranking is stable. As a way of example, this would be the case for the number of diagnoses of a particular disease recorded in Medicare during a month [237]. If V is large, as occurs for metagenomic samples, the ranking might be unstable, as it would be for the number of hourly pages views of articles in Wikipedia [162, 235]. Interestingly, the Taylor parameters were related to the health status of the host, which is the main finding of this study.

Taylor's parameters describing the temporal variability of the gut microbiome in our sampled subjects are shown as supplementary tables in Appendix section A.1.1. The results are indicative of ubiquitous behaviour. Firstly, the variability (which corresponds to the maximum amplitude of fluctuations) was large, which suggests the resilient capacity of the microbiota. Secondly, the scaling index was always smaller than one, which means that more abundant taxa were less volatile than less abundant ones. In addition, Taylor's parameters for the microbiome of healthy individuals in different studies [47, 59, 73, 120, 168–170] were compatible with estimated errors. This enabled us to define an area in the Taylor parameter space that we called the *healthy zone*.

In order to jointly visualise and compare the results of subjects from the above-mentioned studies [47, 59, 73, 120, 168–170], their Taylor parameters were standardised, with standardisation meaning that each parameter was subtracted by the mean value and divided by the standard deviation of the group of healthy subjects for every study independently. Due to the different systematics in each study, we defined a healthy region from each of them, standardised to mean zero and variance one, then we computed mean and variance of unhealthy with this standardisation (for details of the procedure, please see Standardisation subsection in Material and Methods). Therefore, different studies were isolated so that subjects from a given study did not affect the results for the unhealthy subjects of the other studies. It

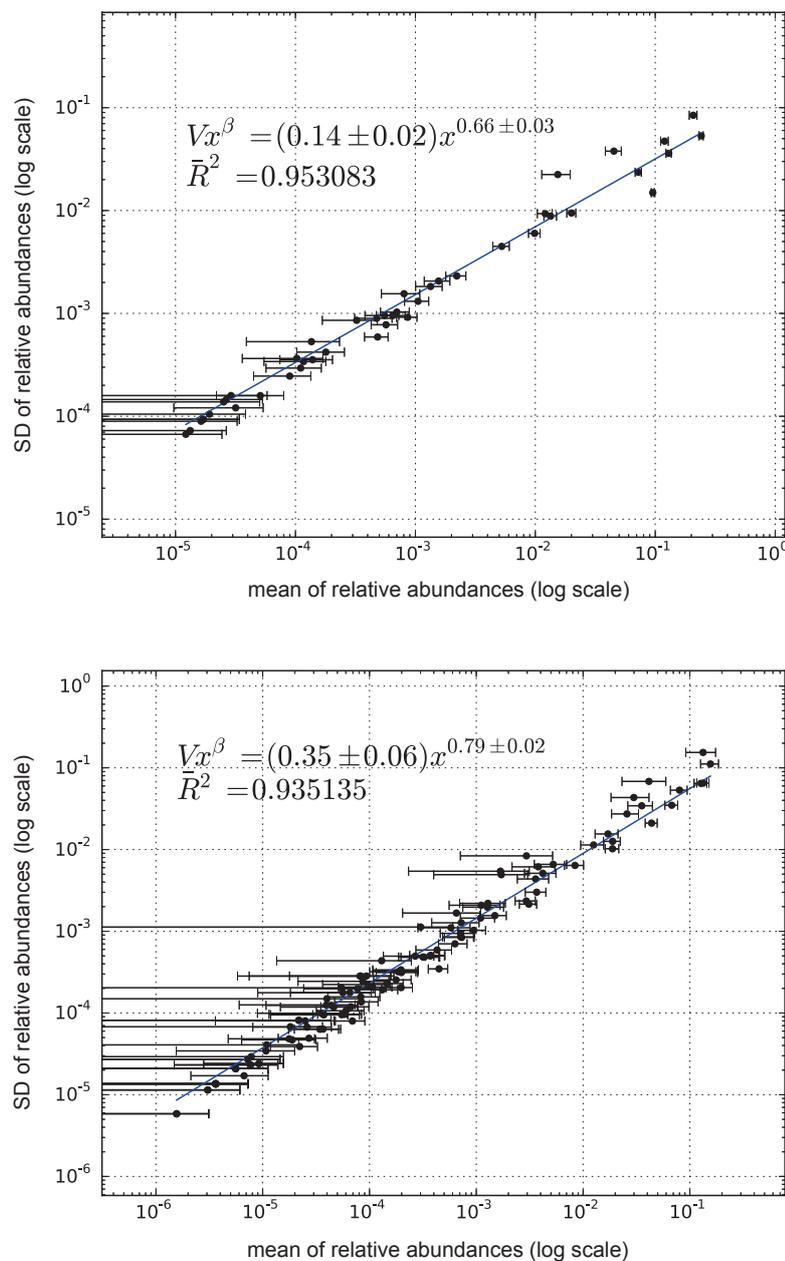


Fig. 4.1 X-weighted power-law fits of the standard deviations versus the mean values for each bacterial genus monitored over time. The fit is shown for samples from a healthy subject (top) and from a subject diagnosed with irritable bowel syndrome (bottom), studied in our lab [73]. Taylor's power law seems to be ubiquitous, spanning to six orders of magnitude. V corresponds to the y-intercept and β to the slope of the line. The error bars (*mean-axis*) are the SEM.

was assumed that this statistical approach was safer, as it was avoided combining data with very different systematic errors. The healthy zone and the standardised Taylor parameters

for the temporal variability of the gut microbiome in subjects whose gut microbiota was compromised (i.e., they were suffering from IBS, kwashiorkor, altered diet, antibiotic intake, a *Salmonella* infection, or had gone on a trip abroad) are shown in Figure 4.2. The variability in children with kwashiorkor was smaller than that of their healthy twins. A meat/fish-based diet significantly increased variability when compared to a plant-based diet. All other cases presented increased variability, which was mainly severe and statistically significant at over 95% confidence level (CL), for grade III obese patients on a diet, subjects taking antibiotics, the subject who had a *Salmonella* infection, the subject who had travelled abroad and the IBS-diagnosed patients. One global property emerged from these comprehensive data: Taylor's parameters characterised the statistical behaviour of microbiome changes. Furthermore, we verified that our conclusions were robust to systematic errors resulting from the taxonomic assignment (see Taxa-level robustness in Material and Methods of this chapter).

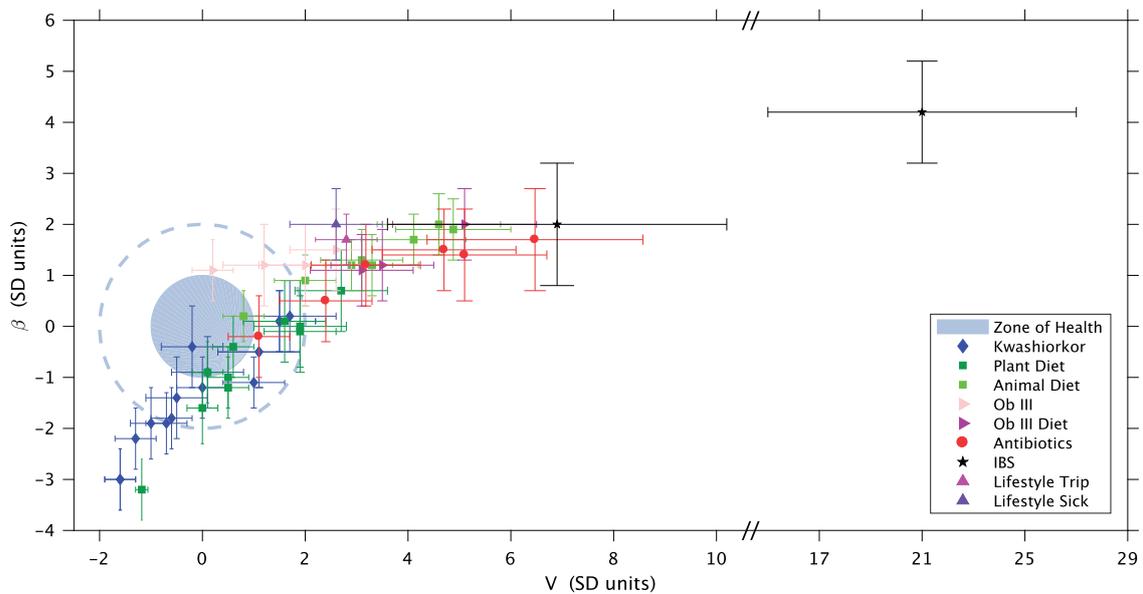


Fig. 4.2 Taylor's law parameter space. All the data studied in this work were compiled here. The colored circle corresponds to a 68% confidence level (CL) region of healthy subjects in the Taylor's parameter space, while the dashed line delimits the 95% CL region. Points with errors place gut microbiome in the Taylor's parameter space, for each subject whose microbiota was compromised. It should be noted that the parameters were standardized (standard deviation units) to the healthy group in each study for every single study independently, for demonstrative and comparative purposes.

4.3.2 The phase space of gut microbiota

Taylor's power law has been explained in terms of various effects, though without a consensus. It has its origin in mathematical convergence, which is similar to the central limit theorem, and thus virtually any statistical model designed to produce a Taylor law converges to a Tweedie distribution [238], providing a mechanistic explanation based on the statistical theory of errors [239–241]. To reveal the generic mechanisms that drive different scenarios in the $\beta - V$ space, we modelled the system by assuming that taxon relative abundance followed a Langevin equation with, on the one hand, a deterministic term that captured the fitness of each taxon and, on the other hand, a randomness term associated with Gaussian random noise [162]. Both terms were modelled by power laws, with coefficients that can be interpreted as the taxon fitness F_i and the variability V (see Model in Material and Methods). Fitness F_i captures the time scale that the system needs to reach equilibrium (the size of variability V may or may not allow equilibrium to be reached). F_i has dimensions of 1/time and roughly corresponds to the half-life of the system when decaying to the stable state. In fact, it is exactly the half-life if β is one and V is negligible. In this model, when V is sufficiently low, abundances are stable in time. Differences in the variability V can induce a noise-induced phase transition in the relative abundances of taxa. The Fokker-Planck equation governs the temporal changes in the likelihood that a given taxon has the abundance x_i , given its fitness. The results of solving this equation show that stability is best captured by a phase space determined by the fitness F and the amplitude of fluctuations V (see Figure 4.3).

The model predicted two phases for the gut microbiome: a stable phase with large variability that enabled some changes in the relative abundances of taxa; and an unstable phase with even larger variability, above the phase transition, where the order of abundant taxa varies significantly over time. The phase transition is continuous (of second order), as is the crossing of the boundary. The state variable is the composition. Any disturbance modifies the composition of the microbiota, with different compositions encoding different F and V . We have shown that effective perturbations significantly change V and lead the microbiota to a transition from the ordered phase to the noise-induced one. Our model can be solved analytically, which allows for a simple understanding of the different regimes and, in particular, to calculate the formula of the transition region. The order parameter is the composition x_M that maximises the probability distribution: $0 < x_M < 1$ defines the ordered phase, while $x_M > 1$ defines the disordered phase. If V is sufficiently small compared to F , the likelihood peaks in the physical region (relative compositions larger than zero and smaller than one), i.e., there is a best composition solution of the differential equation, which is the ordered solution. Conversely, if V is sufficiently large compared to F , the likelihood peaks outside the physical region, i.e., the best composition solution of the differential equation is

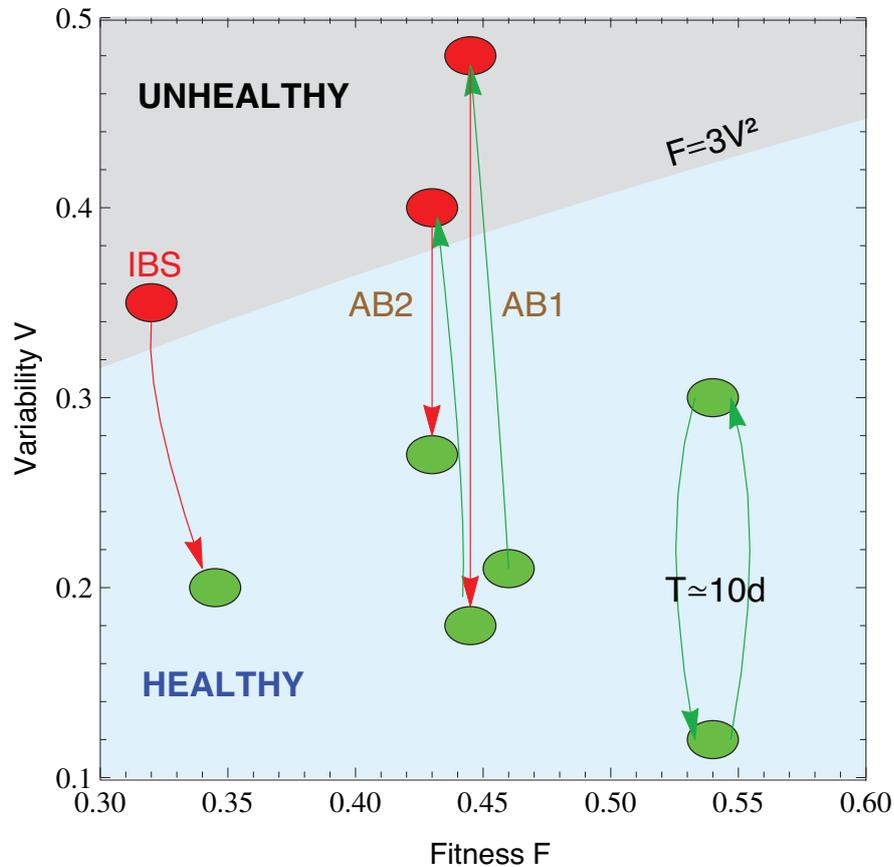


Fig. 4.3 Microbiota states can be placed in the phase space $F - V$. The light-blue shaded region corresponds to the stable phase, while the grey shaded region is the unstable phase (the phase transition line is calculated for $\alpha = \beta = 0.75$). We placed healthy subjects (green) and subjects whose gut microbiota is threatened (antibiotics, IBS) in the phase space fitness–variability. The gut microbiota of healthy subjects over a long term span show a quasi–periodical variability (central period is ten days). We show that taking antibiotics (AB1 and AB2 correspond to the first and second treatment, respectively) induces a phase transition in gut microbiota, which impacts on future changes. We also show an IBS–diagnosed patient transiting from the unstable to the stable phase.

at the boundaries (either zero or one), and all physical solutions have comparable likelihoods, i.e., the noise–induced phase. The microbiome of healthy subjects was found to be in the stable phase, while the microbiome of several other subjects was in the unstable phase. In particular, subjects taking antibiotics and the IBS–diagnosed patient $P2$ had the most severe symptoms. In this phase diagram, each microbiota state is represented by a point at its measured variability V and inferred fitness F . The model predicted high average fitness for all taxa, i.e., taxa were narrowly distributed in F . The fitness parameter was chosen with

different values for demonstrative purposes. Fitness was higher for the healthiest subjects and smaller for the IBS–diagnosed patients.

4.3.3 Rank stability of the taxa

The rank dynamics and stability plots in Figures 4.4 and 4.5 show the variations in rank over time for the most dominant taxa and their calculated Rank Stability Index (RSI, as discussed in Material and Methods) for the gut microbiome taxa of a healthy subject, namely subject A in the host lifestyle study [59]. The Figure 4.4 covers the period when the subject is travelling abroad and Figure 4.5 covers the subsequent period. The taxa are listed according to their accumulated frequency over the time series, with the y-axis being the overall dominant axis for each sample set. Generally speaking, we observed that the most dominant taxa had the highest rank stability.

For the trip abroad in Figure 4.4, beyond the differences in the dominance of particular taxa, we observed that the most dominant were the also most rank–stable. Moreover, the medium–ranked taxa were quite rank unstable, mostly due to transient (often one or two consecutive samples) albeit dramatic declines in their relative abundance, which usually occurred more than twice during their time series.

Nevertheless, in the particular case of the next period (Figure 4.5), the one after the trip, some taxa showed higher stability than other more dominant taxa, forming *rank stability islands* for medium–ranked taxa displaying a moderately stable index (RSI roughly over 70%). In particular, this was the case for the genera *Actinomyces*, *Leuconostoc*, *Lachnobacterium*, *Eggerthella*, *Clostridium* and *Collinsella*. For the genera mentioned above, both the overall rank and the RSI were lower during the trip (RSI under 70%). *Actinomyces* and *Lachnobacterium* are not shown in Figure 4.4 because they sank to positions 56 and 77, respectively. By contrast, *Leuconostoc* was the least sensitive to the lifestyle change. Interestingly, *Lachnobacterium* showed anti-correlation over time compared to the vast majority of the taxa classified in this study.

We also found those *rank stability islands* for maximum-ranked taxa in the other periods belonging to subject A in the host lifestyle study [59] (see Supplementary Figure A.3 and Supplementary Figure A.4 for the corresponding rank plots in Appendix A.1). See Supplementary Table A.5 for details about the rank and RSI for the above-mentioned taxa over the different periods considered.

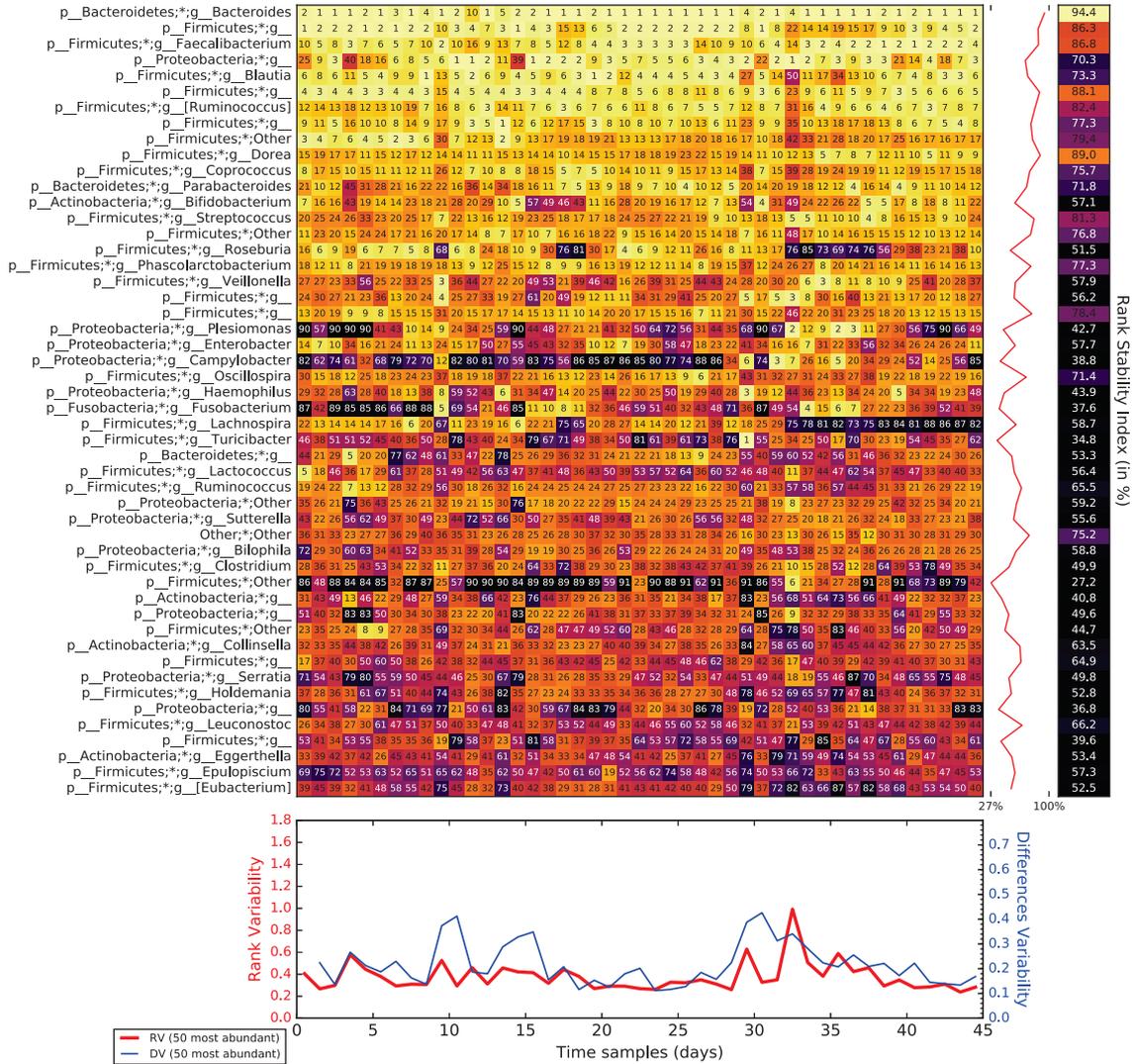


Fig. 4.4 Rank variation over time for the 50 most dominant elements (taxa) and their calculated RSI (Rank Stability Index), Rank Variability (RV) and Differences Variability (DV), as detailed in Rank stability and variability in Material and Methods, for a special period (days 72 to 122, travelling abroad) belonging to subject A in the host lifestyle study [59].

4.3.4 Time dependence of model parameters

Finally, we studied the time dependence of the variability V and power law index β (see Model in Material and Methods) by using a sliding window approach. The total number of time points was divided into subsets of five points, where the following subset was defined by adding the next sampling time and eliminating the earliest one. Both parameters were calculated for each subset against the average time lapse. Figure 4.6 shows the variability V

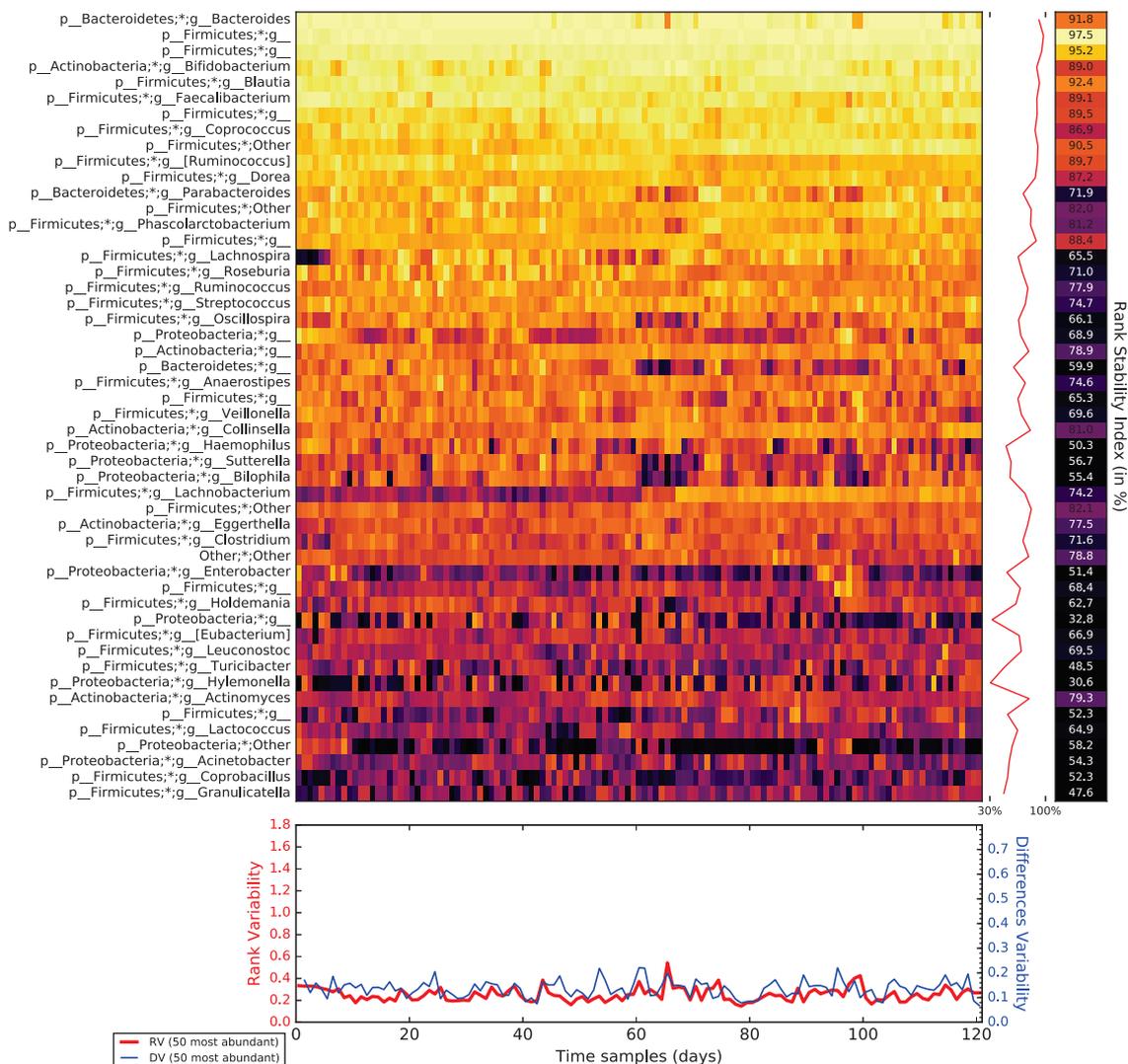


Fig. 4.5 Rank variation over time for the 50 most dominant elements (taxa) and their calculated RSI (Rank Stability Index), Rank Variability (RV) and Differences Variability (DV), as detailed in Rank stability and variability in Material and Methods, for an ordinary period (days 123 to 256, after the trip) belonging to subject *A* in the host lifestyle study [59].

as a function of time for the two subjects in Caporaso's study [47] corresponding to the gut microbiota of a male (upper plot) and a female (lower plot). Both samples showed changes in the variability V with quasi-periodic behaviour peaking at about ten days. Variability grew more for the gut microbiota of the male and shared a minimal value of around 0.1 with the gut microbiota of the female.

Figure 4.7 shows time-course changes in V for patient *P2* in the IBS study [73] (upper plot) and patient *D* in the antibiotics study [120] (lower plot). The variability of the gut

microbiota of $P2$ decreased from over 0.3 to below 0.2, showing a slow tendency to increase the order of the system. Antibiotic intake led to a quick increase in variability which lasted for a few days to recover ordering. The second antibiotic treatment showed some memory traits (a lower increase of variability) with a slower recovery.

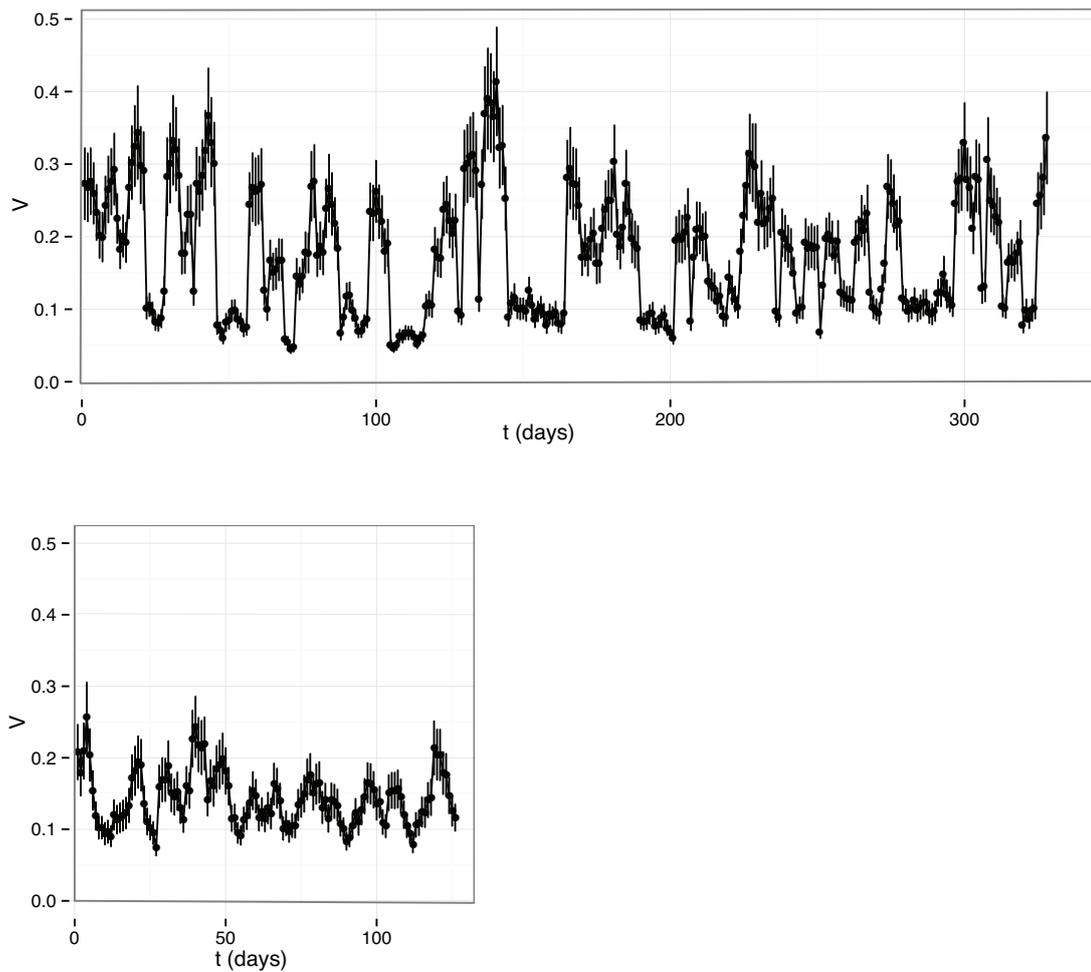


Fig. 4.6 V as a function of time for the two subjects in Caporaso's study [47]: samples of gut microbiome of a male (upper plot) and a female (lower plot).

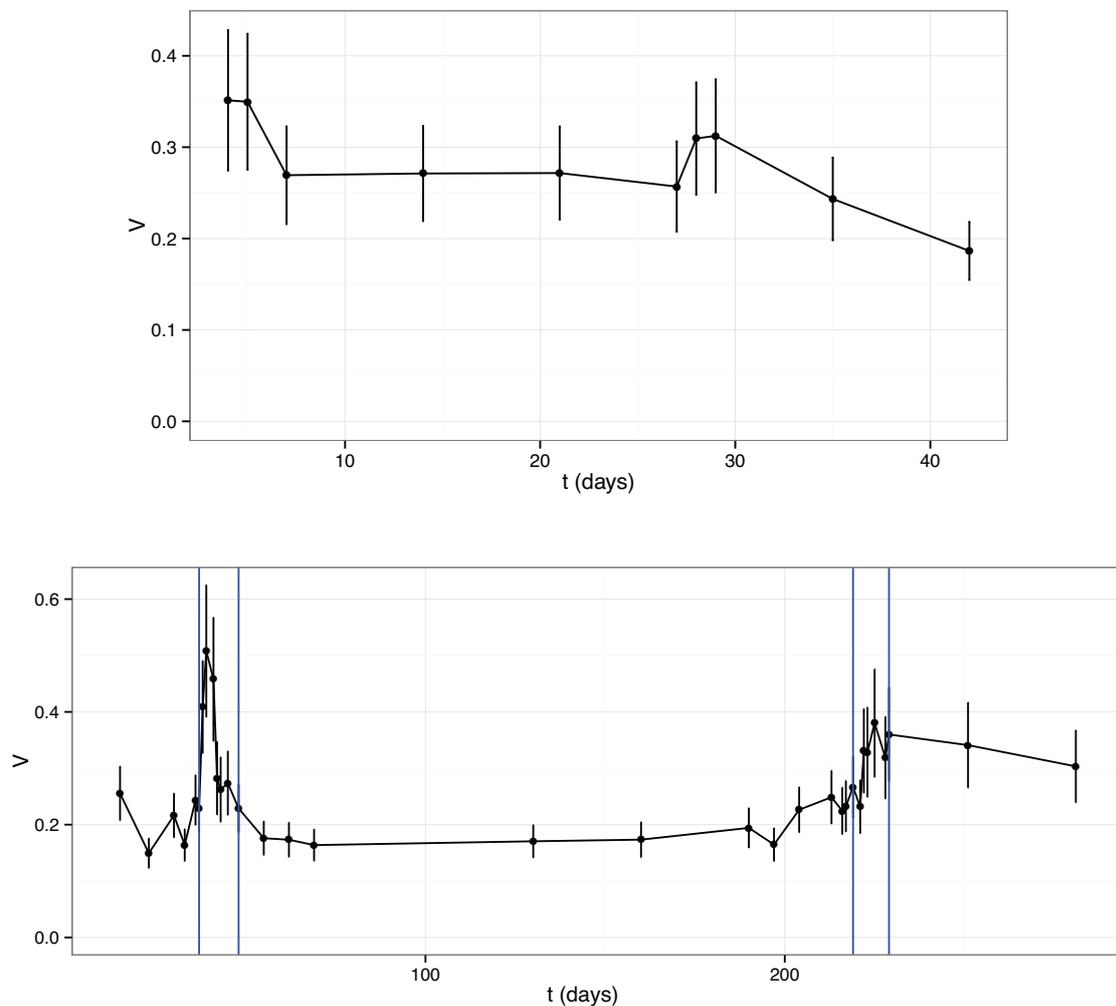


Fig. 4.7 V as a function of time for patient $P2$ in the IBS study [73] (upper plot) and patient D in the antibiotics study [120] (lower plot). The blue vertical lines in the lower plot show the periods of antibiotic treatment.

4.4 Chapter Summary

As a summary of the results presented here, which will further discussed in the General Discussion, we can observe that all the studied microbiomes follow the Taylor's law. What is more interesting is that there is a robust pattern in the parameter values of the ecological law, being β , the scaling factor, always less than 1. Briefly, this means that, in relative terms, the most abundant elements in the population are less volatile to perturbations than the less abundant ones. The other Taylor parameter, V , measures the variability amplitude of the whole population. Our results highlight the connection between V and the stability of the

microbiome, relating higher values with the perturbations under study, i.e., subjects suffering from IBS [73], children suffering from Kwashiorkor [169], different diets [170], obese patients [168], subjects receiving antibiotics [120], or abrupt changes in the host lifestyle [59], as it is represented in Figure 4.2. As the results imply, the microbiome variability is different in each condition, being more variable than the healthy subjects in adults with diseases or other kinds of perturbation, and less variable in children with Kwashiorkor compared to the healthy subjects from the study of Smith *et al.* [169].

Here, we propose the Langevin equation as a model to study the stable regimes of the microbiomes, derived from the work of Blumm *et al.* [162]. This equation has two components, one that measures the system variability, and the other component that measures the ‘fitness’ of the system. Variability can be directly measured from data with the Taylor’s law, and fitness still has to be modelled. The solution of this equation gives a phase space with two regimes: one where the microbiomes are stable against perturbations, and the other where they are unstable, as represented in Figure 4.3. Depending on the fitness, a microbiome can be more or less variable and still be in the stable region. Interestingly, studying the behaviour of V parameter in time, we can observe that microbiomes transitions from one state to the other, as represented in 4.3. Furthermore, with the study of the Rank Stability of the different subjects, we can easily observe changes in their composition due to perturbations, as happens with the two subjects in the work of David *et al.*, [59], represented in Figures 4.4 and 4.5. Also, in the rank stability matrices, we can observe the presence of *rank stability islands* among medium-ranked taxa. We find, for example, that *Lachnobacterium*, *Clostridium* and *Actinomyces* remain quite stable in health conditions, becoming more variable in perturbed states. The implications of this fact will be considered in the General Discussion.

Chapter II - Microbial interactions as a matter of noise

5.1 Summary

The behaviour of a microbial ecosystem is determined by the interactions among the participating species. Time series studies of the microbiome are widely used to determine temporal correlations among their members. However, it has been shown that correlations in microbiota do not directly imply biological interactions. We present a novel method able to extract microbiome interactions from metagenomes in time series experiments, where the experimental design must include alterations of the microbial ecosystem to uncover specific interactions. In this work, we simulate a microbial ecosystem with interactions based on a small-world, scale-free and modular ecological networks, with parameters that mimic the human gut microbiota. We demonstrate that microbial interactions can be inferred from noise induced experiments in which the variability of a given species is modified. We illustrate our findings with two examples where a four-fold increase in the variability of a given taxon uncovers some of its interactions with other taxa in the ecosystem.

Original work:

Martínez-Martínez D, Moya A, P. Garay C. Uncovering microbial interactions by noise induced experiments. *Work in progress*

5.2 Results

5.2.1 Species simulations

In this work, we aim to study microbial interactions from a bottom-up approach, by the modification of specific species in a community. In order to understand the potential of the method proposed here, we have simulated a realistic microbiome with N species in equilibrium [228], with typical variability observed in this system [242, 243], comparable to human gut microbiome [231]. We have modelled the dynamics of the system is described by the generalised Lotka-Volterra (gL_V) model extended with an stochastic component (sgL_V), so that the compositions are random variables, as shown in equation 5.1. The model is explained in Material and Methods section.

$$\frac{d}{dt}x_i(t) = \mu_i x_i(t) + x_i(t) \sum_{j=1}^L M_{ij} x_j(t) + Bx_i^b \cdot dW \quad (5.1)$$

For simplicity, we have assumed that all species have the same growth rate ($\mu_i = 1$). The interaction matrix was built with the Klemm-Eguíluz's algorithm. In particular, we have constructed a network of 30 different species, with small-world, scale-free and modularity properties (see Figure 5.1 and discussion in Methods). The network is non-symmetric, and the interaction coefficients range from -0.5 to 0.5. The small-world property leads to only a few members, species (sp) 14 or sp7, widely connected to other individuals, while most of the members are scarcely connected, for example sp10 or sp25. The interaction network passed all three stability tests in *testStability* function of the seqtime R package: Coyte's criteria on the return speed to the steady-state [153]; negative real part of the interaction matrix eigenvalues; and Ricker's stability criteria for explosions in simulations [230]. For the stochastic component, we have assumed that the noise amplitude for each component follows Taylor's law.

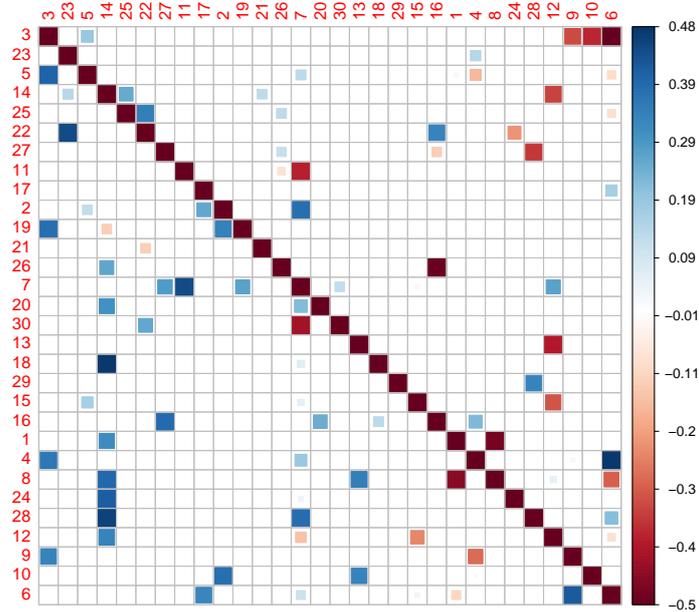


Fig. 5.1 Interaction matrix used in our simulations. The model includes thirty labeled as n , with $n = [1,30]$. The order of matrix rows (and columns) is given by the abundances. The interaction matrix coefficients are color coded from red (-0.5) to blue (+0.5), and following the standard convention of the adjacency matrix: outputs are encoded in rows and inputs are encoded in columns.

5.2.2 Noise-induced and pulse perturbation experiments

Firstly, we have evaluated how the amplitude of the time variability propagates from one species to others. For doing it, we follow the time evolution of the relative abundances and at fixed times, we modify the noise amplitude for a given species in the sgLV equation. We have performed this study for some species and we show the results for modifications of species sp3 and sp14. The top panel in Figure 5.2 shows the time variability of the relative abundances of the most abundant species under different initial conditions. We have evolved the system with initial conditions (default conditions) $b = 1.5$ and $B = 0.05$ for the noise amplitude. At a given time, in arbitrary units, $t = 20$, we modify the sgLV equation by increasing B_{sp3} fourfold, while keeping all others fixed to the default conditions. Later ($t = 40$), we have set the sgLV equation to default conditions. At $t = 60$, we modify the sgLV equation by increasing B_{sp14} fourfold, while keeping all others fixed to the default conditions. Finally ($t = 80$), we have set the sgLV equation to default conditions. The reader can easily see the species with increased variability in a given period, but data analysis is needed to

determine how the increased noise in one species is propagated to other species. Secondly, instead of modifying the default conditions of the sgLV model, we have instantaneously halved the relative abundance of a given species (sp3 at $t = 20$ and sp14 at $t = 60$), as shown in the lower panels in Figure 5.2. The abundance of the modified species quickly recovers in less than 1 unity of time, easily seen in the figures.

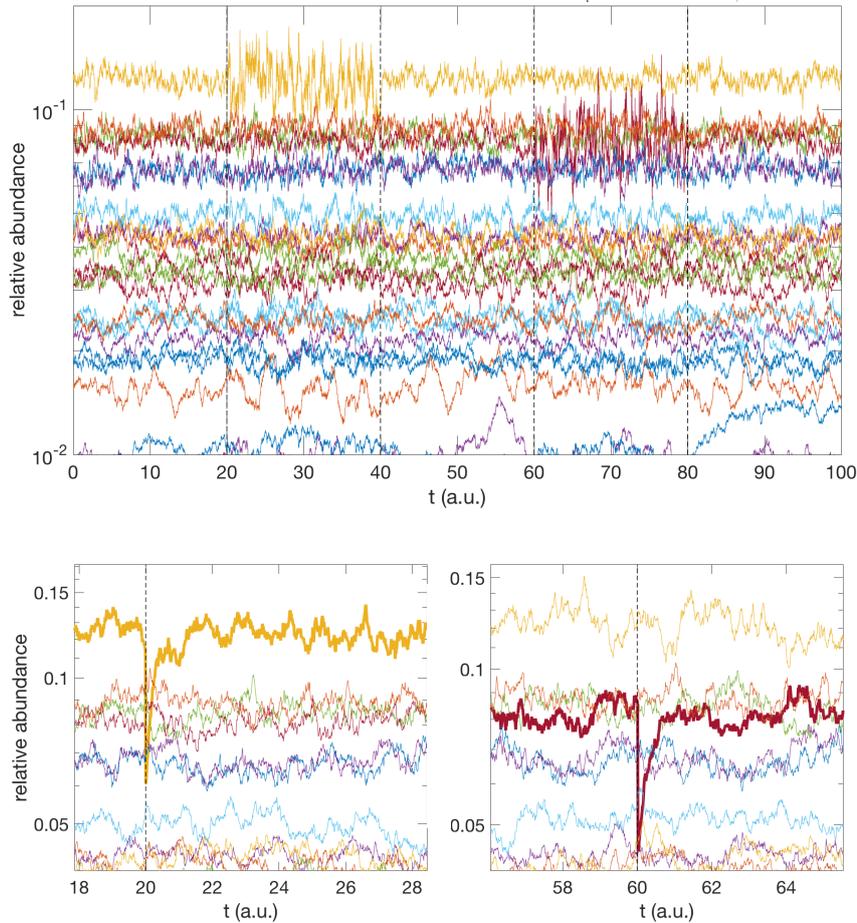


Fig. 5.2 Relative abundances in log scale of the most abundant species as a function of time in our simulations. On the upper part, it is represented the time evolution of our population. From $t = 0$ to $t = 20$, the population evolves with default conditions. At $t = 20$, the noise amplitude of sp3 is increased fourfold until $t = 40$, when we recover the default conditions. At $t = 60$, the noise amplitude of sp14 is increased fourfold until $t = 80$, when we recover the default conditions to the end of the simulation. We have shown abundances larger than 1%. On the bottom left, sp3 abundance is halved at $t = 20$, and on the bottom right, sp14 abundance is halved at $t = 60$. Both sp3 and sp4 have been highlighted on the bottom panels to facilitate the visualisation.

We computed Pearson's correlations over 1000 time points of every simulation aforementioned (Figure 5.3). To facilitate the visualisation, we only represented the 24 most abundant species ordered by overall abundance, because the 6 less abundant accounted only for $4.7 \cdot 10^{-6}$ % of the total abundance. We observe that the modification of an important member of the community, i.e., sp3 and sp14, enable changes in the correlation coefficients pattern compared to the scenario with default conditions. The changes in noise amplitude of sp3 and sp14 (middle row in Figure 5.3) produce clear changes in the correlation coefficient patterns, with special emphasis in the case of sp14, a highly connected member. On the other hand, when the species abundance is halved, we observe a different correlation pattern compared to the previous type of perturbation. Interestingly, both types of perturbation on sp3 and sp14 produced big correlation changes on the less abundant species, i.e., sp1, sp4 and sp8.

5.2.3 Graph Inference

Once we have calculated the correlation matrix associated to the time series, it is often diagonalised the matrix to know the principal directions in the 30-dimensional space of random functions (i.e., the time series for each individual species). Such a big dimensional space makes unpractical the visualisation, particularly for monitoring the changes due to the increase in noise or for a pulse in the relative composition. Nevertheless, it shows the potential of using linear combinations of the individual time series if we were able to define an optimisation condition that selects particular directions most sensitive to the changes introduced in the experimental design. One of these strategies, commonly used when we do not require any knowledge of the dynamical underneath the time series, is the *Granger causality* test, an statistical hypothesis test used for verifying whether one time series can predict future values of another one. Intuitively, the test verifies the hypothesis that the past values of a time series should be helpful to predict future values of the other. More formally, a time series X is said to 'Granger cause' another time series Y in terms of past values of Y and X if the accuracy of regressing Y in terms of the past values of Y and X is statistically better than just regressing with the past values of Y .

One of the fundamental parameters in the analysis is the maximum time lag allowed in past observations. The number of lags to be included is usually chosen using an information criterion. Any particular lagged value of one of the variables is retained in the regression if: (1) it is significant according to a t-test; and (2) it and the other lagged values of the variable jointly add explanatory power to the model according to an F-test. Then the null hypothesis of no Granger causality is not rejected if and only if no lagged values of an explanatory variable have been retained in the regression. The original definition of Granger causality

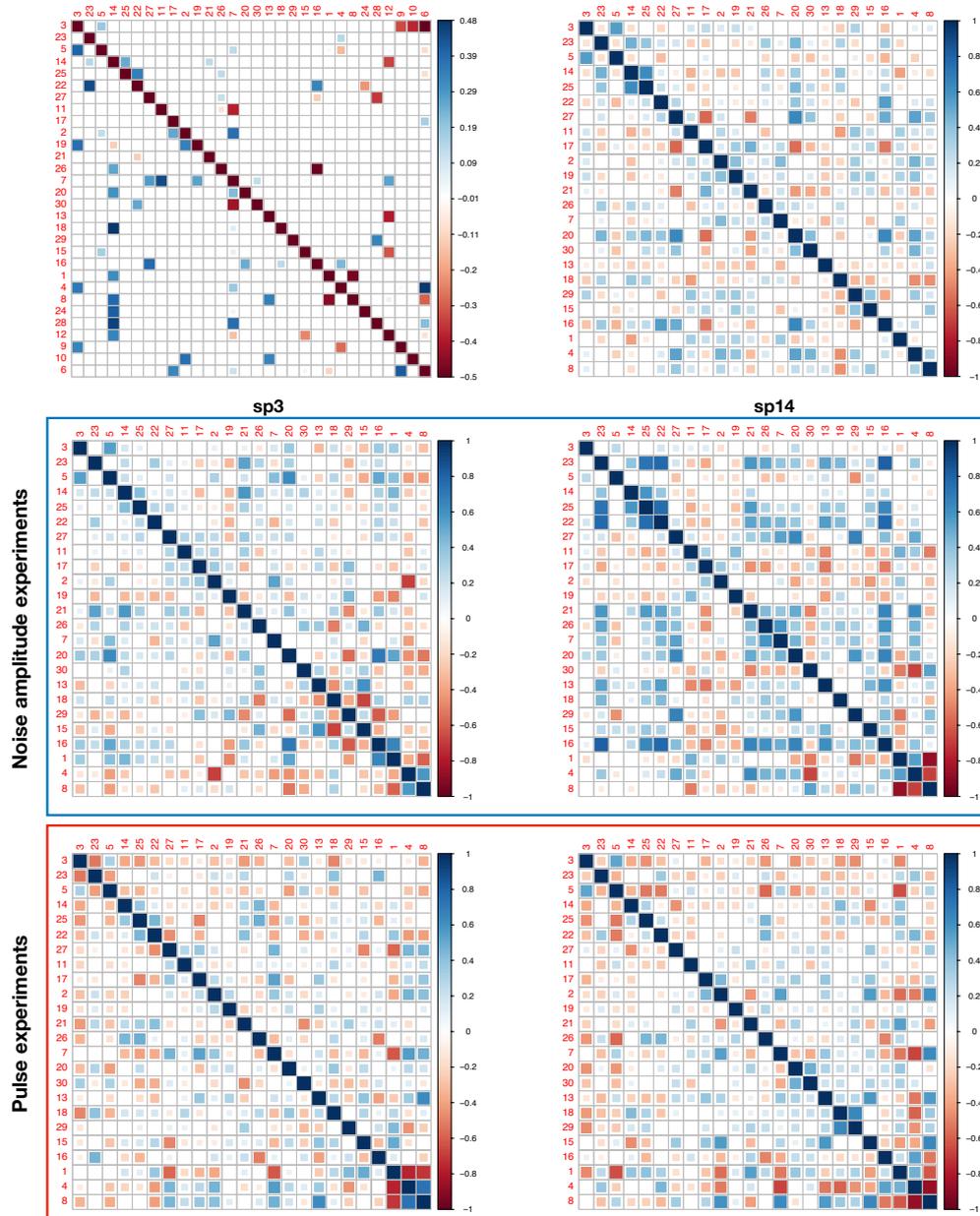


Fig. 5.3 Interaction and correlation matrices for the normalised counts of the simulations. On the upper left, it is represented the original interaction matrix. On the upper right, it is represented the correlation matrix for the default conditions. On the center left, it is represented the correlation matrix of sp3 with its noise amplitude increased fourfold. On the center right, it is represented the correlation matrix of sp14 with its noise amplitude increased fourfold. On the bottom left, it is represented the correlation matrix after halving sp3 abundance. On the down right, it is represented the correlation matrix after halving sp14 abundance. All matrices have been truncated at species higher than an average abundance of 10^{-3} . The average abundances give species order. Colour code is blue for positive interaction/correlation, and red for negative interaction/correlation.

does not capture instantaneous and non-linear causal relationships, though several extensions have been proposed to address these issues.

In practice, to perform the regression, we have solved the least squares problem to do the interaction matrix inference (see Methods section). In particular, we have solved the regression problem to all the aforementioned experiments: the ones in which the noise amplitude of sp3 and sp14 were increased, and the pulse in which we halved their abundances at a certain point. As can be seen in Figure 5.4, the inferred matrices highlighted the importance of the most abundant species as most of the inferred interactions are located in sp3, sp23, sp5 and sp14. On one hand, the increase in sp3 noise amplitude shows the most marked effects on the interactions inferred. Compared to the values per default, applying noise to sp3 seems to clean most of its connections with the exception of the interaction between sp3 and sp5. However, it also adds spurious interactions in sp23 and sp14. Interestingly, when we increase sp14 noise amplitude most of its spurious interactions disappear, remaining only the interactions between sp14 with sp23, sp25, and with itself. On the other hand, the pulse experiments do not seem to improve in a great deal our ability to infer the interactions.

To study whether the noise-induced and the pulse experiments improved our ability to uncover the real interactions, we calculated the error of the inference between the matrix a_{ij} with the original M_{ij} by calculating the norm between them (see Methods section). We applied different noise amplitudes for sp3 and sp14 and compared: first, the error between the entire matrices; and second, the specific rows of the modified species. In Figure 5.5, we show that the inference error for sp3 and sp14 rows improve as we increase the noise amplitude to values of $B = 0.20$, while the overall inference error does not necessarily decrease. We obtain the same pattern in the inference error when we vary the noise amplitude of other highly abundant species as sp23, improving the inference of sp23 interactions but not for the entire matrix (see Appendix Figure A.6). On the contrary, it seems that increasing the species noise amplitude do not work properly with low abundant species as sp7. We modified this species as it is the most connected element of our matrix and also with a low abundance, and we did not improved the inference error neither with the entire matrices, nor when comparing the specific row of sp7.

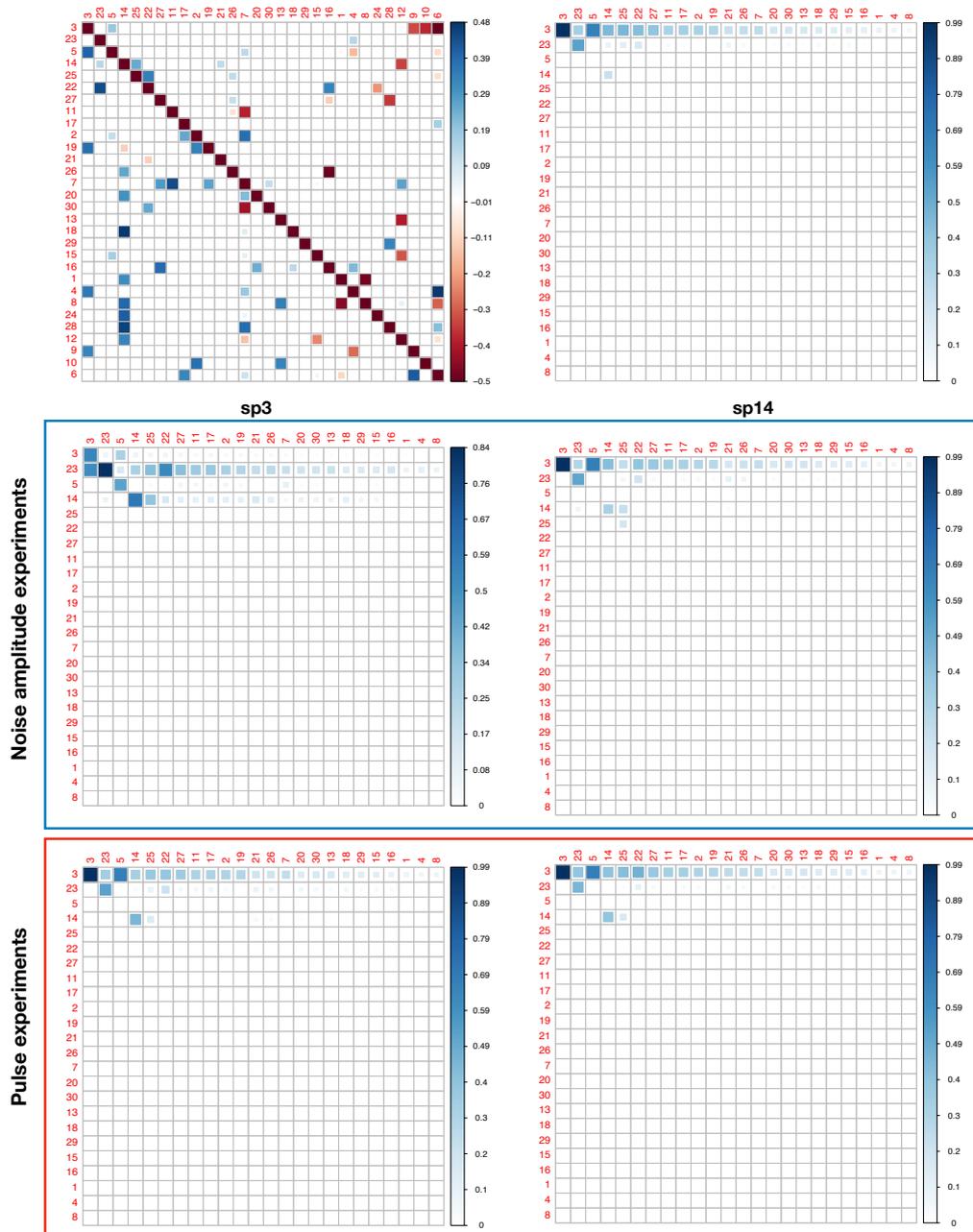


Fig. 5.4 Interaction matrices inferred from the regression model for the normalised counts.. On the upper left, it is represented the original interaction matrix. On the upper right, it is represented the interaction matrix for the default conditions. On the center left, it is represented the interaction matrix of *sp3* with its noise amplitude increased fourfold. On the center right, it is represented the interaction matrix of *sp14* with its noise amplitude increased fourfold. On the bottom left, it is represented the interaction matrix after halving *sp3* abundance. On the down right, it is represented the interaction matrix after halving *sp14* abundance. All matrices have been truncated at species higher than an average abundance of 10^{-3} . The average abundances give species order. Colour code is blue for positive interaction/correlation, and red for negative interaction/correlation.

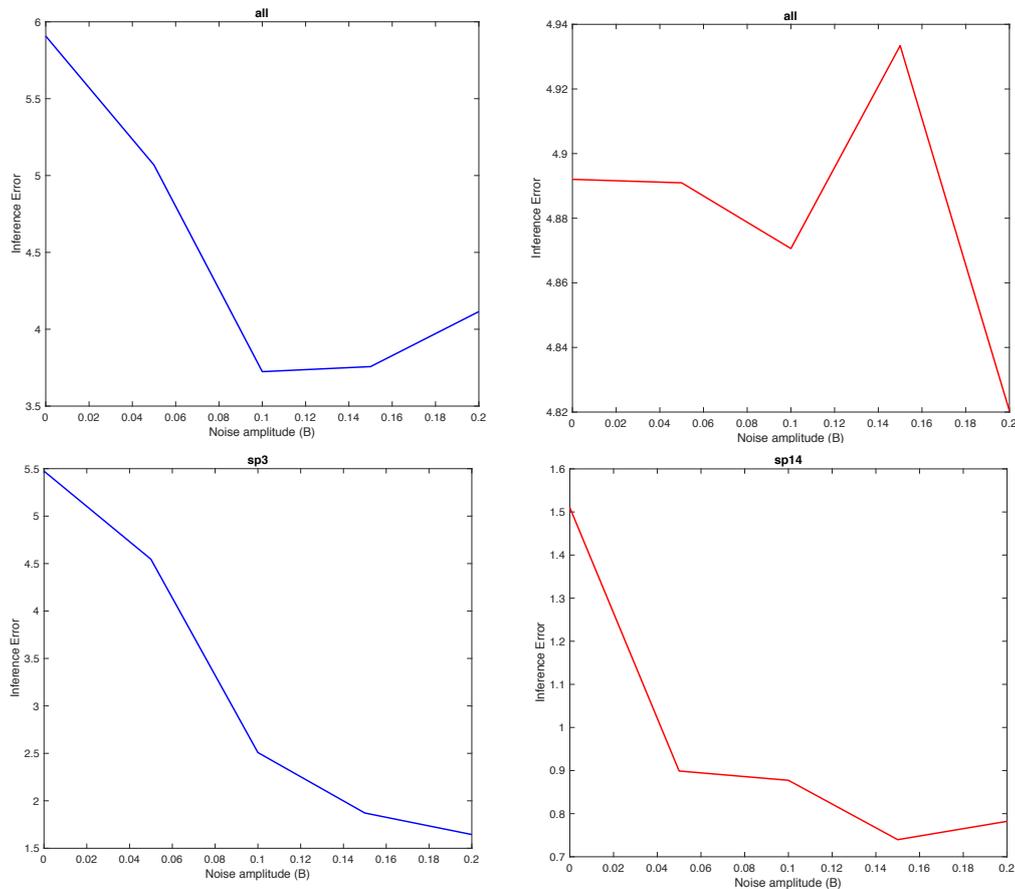


Fig. 5.5 Error between original and inferred interaction matrix. In the upper row, it is represented the error between the entire matrices. On the left, the case when sp3 is modified, and on the right, the case when sp14 is modified. In the lower row, it is represented the error between the specific rows of sp3 (left), and sp14 (right) compared to the original interactions.

5.3 Chapter Summary

As a summary of the results presented here, which will be further discussed in the General Discussion, we have seen that we can model the dynamics of an ecosystem, based on a modified version of the generalised Lotka-Volterra model, extended with a stochastic component so that the compositions are random variables. In this chapter, we propose to make use of the noise present in this biological system, and modify the behaviour of certain species to uncover their interactions with the community. For this purpose, we have modified the behaviour of sp3 and sp14 in two different ways: 1) by increasing their noise in time, and 2) by applying a pulse where we halved their abundances at a certain time point. These

modifications are intended to mimic controlled perturbations that can be actually applied to bacteria in laboratory conditions. The correlation matrices showed marked changes in the behaviour of the community, shedding some light to the real interactions. However, the predictive power of interaction coefficients is limited to use them as proxies for real interactions.

In order to examine whether the noise-induced experiments increases our ability to uncover the microbial interactions, we have solved an optimisation problem with a least-squares regression model, applying a Lagrange multiplier to force sparsity to the inferred interaction matrix. We have observed that the inferred matrices (a_{ij}) were similar, at some extent, to the real interaction matrix (M_{ij}), highlighting the importance of the most abundant species in our problem. Modifying sp3 and sp14 behaviours uncovered some of their most representative interactions. Furthermore, we can observe that the inference error of the modified species in our experiments decrease as their noise amplitude is increased, implying that these kind of experiments could help us to uncover real interaction in controlled experiments with real microorganisms. However, the results presented in this chapter are still preliminary, the inferred matrices still do not appear to be very similar to the original interaction matrix, and the error seems too high in some occasions. Thus, to take advantage of the presented experimental methodology, we should build better tools for the interaction matrix prediction, more robust to take into consideration any type of graph topology to further validate the utility of the proposed experiments.

Chapter III - Dynamical analysis of children's microbiota with acute diarrhoea

6.1 Abstract

Gut microbiota is closely related to acute infectious diarrhoea, one of the leading causes of mortality in children worldwide. Understanding the dynamics of the recovery from this disease is of clinical interest. This chapter aims to correlate the dynamics of gut microbiota with the evolution of children who were suffering from acute infectious diarrhoea caused by a rotavirus, and their recovery after the administration of a probiotic. The experiment involved 10 children with acute infectious diarrhoea caused by a rotavirus, and six healthy children, all aged between 3 and 4 years. The children who suffered the rotavirus infection received *Saccharomyces boulardii* CNCM I-745 twice daily for the first 5 days of the hospitalisation. Faecal samples were collected from each participant at 0, 3, 5, 10, and 30 days after hospitalisation. Microbial composition was characterised by 16S rRNA gene sequencing. Diversity indices were computed, along with dynamical analysis based on Taylor's law. We observed low alpha diversities in the first five days (p-value <0.05, Wilcoxon test), larger at 10 and 30 days after probiotic treatment. CCA showed differences in the gut microbiota of healthy children and of those who suffered from acute diarrhoea in the first days (p-value <0.05, ADONIS test), but not in the last days of the experiment. Temporal variability was larger in children with acute diarrhoea than in healthy ones. In particular, *Gammaproteobacteria* class was found to be abundant in children with acute diarrhoea.

Original publication:

Dinleyici EC*, Martínez-Martínez D*, *et al.*. 2018. Time series analysis of the microbiota of children suffering from acute infectious diarrhoea and their recovery after treatment. *Frontiers in Microbiology*. doi:10.3389/fmicb.2018.01230.

* Equal contribution

6.2 Material and Methods

A general view of the material and methods were explained in Material and Methods Chapter. Nevertheless, the most relevant information is described in the following subsections to facilitate the understanding of the results. Detailed information has been provided when necessary.

6.2.1 Sample selection

We enrolled a total of 16 children, 6 of them with a healthy condition, and the other 10 suffering from acute infectious diarrhoea. All children were from 3 to 4 years old. From each child enrolled, we collected samples at five different time points, namely, at 0, 3, 5, 10, and 30 days, for a total of 80 different samples. More information can be found at Table A.6, in the Appendix. The sample at day 0 was taken before probiotic treatment in the group of children with infectious diarrhoea. All children belonging to the infectious diarrhoea group received oral lyophilised *Saccharomyces boulardii* CNCM I-745 (250 mg twice daily, ReFlor, Biocodex) for the first five days, specifically days 3 and 5 (day 0 was with no intervention), in addition to ORS and/or intravenous therapy. Rehydration and electrolyte replacement were done using hypoosmolar ORS (glucose 20 g; sodium 60 mmol/L; potassium 20 mmol/L; bicarbonate 30 mmol/L).

6.2.2 16S rRNA gene sequencing and read processing

A region of the 16S rRNA gene (V4) was selected to be studied. Sequencing was performed on the Illumina MiSeq platform according to the manufacturer's specifications. 16S rRNA gene reads were quality filtered with PRINSEQ, lite version [181]. Reads with a low score (<20) and short read lengths (<50 nucleotides) were removed. Chimeric sequence filtering and taxonomic assignment of the 16S rRNA sequences were performed with the ChimeraSlayer algorithm and the open-reference workflow, respectively, in QIIME (v 1.8) [186], with default parameter values. OTUs were clustered at 97% similarity with the Uclust algorithm. The most representative sequences for each OTU were compared with those in the Greengenes database (representative sequences aligned at 97% similarity, August 2013). Only those annotations that had more than 0.6 in their bootstrap confidence estimation values were accepted for further analyses. For more information about these processes, see the sections 3.2.2 and 3.2.3 in Material and Methods Chapter.

6.2.3 Diversity and functional analyses

α -diversity estimators were computed with R programming language using the *vegan* package (v. 2.4-3) [204]. We applied two filters on the OTU table that resulted from QIIME: any OTU that had less than 80 counts in total was deleted, and any OTU that had a count of 0 in more than 65 samples was excluded from the analysis. Shannon entropy and Pielou evenness were computed. Then, the true diversity index based on Shannon entropy was also computed with the exponentiation of the Shannon values. Differences in α -diversities were studied by performing permutations over the Wilcoxon signed-rank test with a significance level of 0.05, and it was calculated using the *coin* package in R [205]. In the case of multiple comparisons, we applied false discovery rate (FDR) correction to avoid type I errors.

For β -diversity, canonical correspondence analysis (CCA) was performed by using the *vegan* package in R programming language together with weighted *UniFrac* distances [37] computed in QIIME. We applied the function *ADONIS* within the *vegan* package, a multivariate ANOVA based on dissimilarity tests, to assess the variables that were statistically significant in the separation of the samples in two dimensions. We used the linear discriminant analysis (LDA) effect size (LEfSe) algorithm from the Galaxy software package of Huttenhower lab to identify the specific taxa that served as the biomarkers for patients with probiotic treatment and the controls [211]. A fixed cut-off α value of 0.05 was specified for the Kruskal-Wallis test. For the bacterial taxa that had significant differences between samples, it was fixed a threshold of 4.0 in the logarithmic LDA score to consider it a discriminative feature. We split the dataset into two subsets: the first one containing the days 0, 3, and 5, and the second one grouping days 10 and 30. This separation was made to have, on the one hand, the time points of the probiotic treatment and, and on the other hand, the time points without the probiotic treatment. We performed both CCA and LEfSe analysis on the resulting subsets.

For functional content prediction, we used the PICRUSt software from Huttenhower laboratory [216]. To assess the differences in function abundances, we applied the Kruskal-Wallis test, with the α value fixed at 0.01.

Further information is available at sections 3.5 and 3.7 in Material and Methods Chapter.

6.2.4 Temporal analysis and stability

The temporal analysis was conducted according to the results of the Chapter I to measure the stability, over time, of all the individuals. Taylor's parameters were computed and standardised in the same manner as explained in section 3.9, with those samples of children with acute infectious diarrhoea compared with the ones of the healthy individuals in the

study. We also computed the rank stability of all samples, which shows time differences per participant. For this purpose, we calculated the rank stability index (RSI), which shows the temporal stability of a certain taxon, by considering its rank in the total population. We also calculated the *rank variability*, which is the absolute difference between each taxon rank and the overall rank, as well as the *differences variability*, which is the absolute difference between each taxon rank at a given time and the value it had in the previous time step, averaged over all taxa present.

The microbiome data from this study are available at NCBI SRA Database under Bioproject ID PRJNA416445.

6.3 Results

6.3.1 Bacterial composition and diversity

A total of 3,664,535 sequences of 16S rRNA gene amplicons from 80 samples passed the quality filters and were assigned to a taxonomy with QIIME. Each sample had an average number of 45,806 sequences with a standard deviation of 12,598 sequences. The 25 most abundant taxa are illustrated in Figure 6.1, and the remaining taxa were joined into a single group named as *Other*. The most abundant genus is *Bacteroides* (13.14%), followed by *Faecalibacterium* (9.6%) and two unclassified genera from the *Ruminococcaceae* (9.35%) and *Enterobacteriaceae* (8.02%) families respectively. The abundance profile of the children has a high heterogeneity. We can see that the *Bacteroides* genus is most abundant in patients C1, H1, H4, or H5, whereas *Faecalibacterium* is in higher proportions in patients C2, C8, C10, or H5, depending on the day we made the observation. C5 is dominated by the unclassified genera from *Enterobacteriaceae*, as this is the sample with the clearest dominance of one OTU over the others, in comparison with the rest of the participants. We can also observe that the *Others* group fluctuates with a higher degree in the unhealthy group.

The analysis of α -diversity with the Shannon diversity index, which compared both health states and considered all times together, reveals that healthy and non-healthy children are statistically different in the Wilcoxon test (p-value = 6.3e-05), with a higher diversity in the former group than in the latter (see Figure 6.2). When we compare the differences between states, but per time, we find statistical differences at day 0 (p-value = 0.02), day 3 (p-value = 0.034) and day 5 (p-value = 0.011). On the contrary, we did not find any statistical significance at day 10 (p-value = 0.42) nor day 30 (p-value = 0.21). All the tests were also applied to Pielou's Evenness index (Supplementary Figure A.7), and the results were almost

the same except that no significant difference was observed between the healthy and sick states at day 3 (p-value = 0.093).

We performed a canonical correspondence analysis (CCA) to the OTU community matrix, with an ADONIS test to assess the variables that were separating the samples with statistical significance. Not only we found statistical differences between the individuals of the experiment (p-value = 0.001), but also between the health status of the host (p-value = 0.009). Both variables separate the samples in two directions (see Figure 6.3). As can be observed, a clear separation exists between the community matrix belonging to healthy children compared with the children suffering from the rotavirus infection. By contrast, more overlapping is present between individuals when this variable is observed in the CCA (see Supplementary Figure A.8). This result points to a differential composition in the communities depending on the health state. To determine if this separation between health status was consistent with the results of α -diversity, we applied CCA and the ADONIS test to the dataset split into two different subsets: the first one corresponding to days 0 through 5, and the second one corresponding to days 10 and 30. Running the ADONIS test on these two subsets revealed statistical differences in the status of the first group (p-value = 0.019), but not in the second group (p-value = 0.289). In the first case, the region of the healthy children is more constrained and centred than that of the children with acute diarrhoea; in the second case, both regions overlap their area (see Figure 6.3 part B and C). These results reinforce the results obtained by alpha diversity, highlighting that we have differences in the first three time points, but not in the last two.

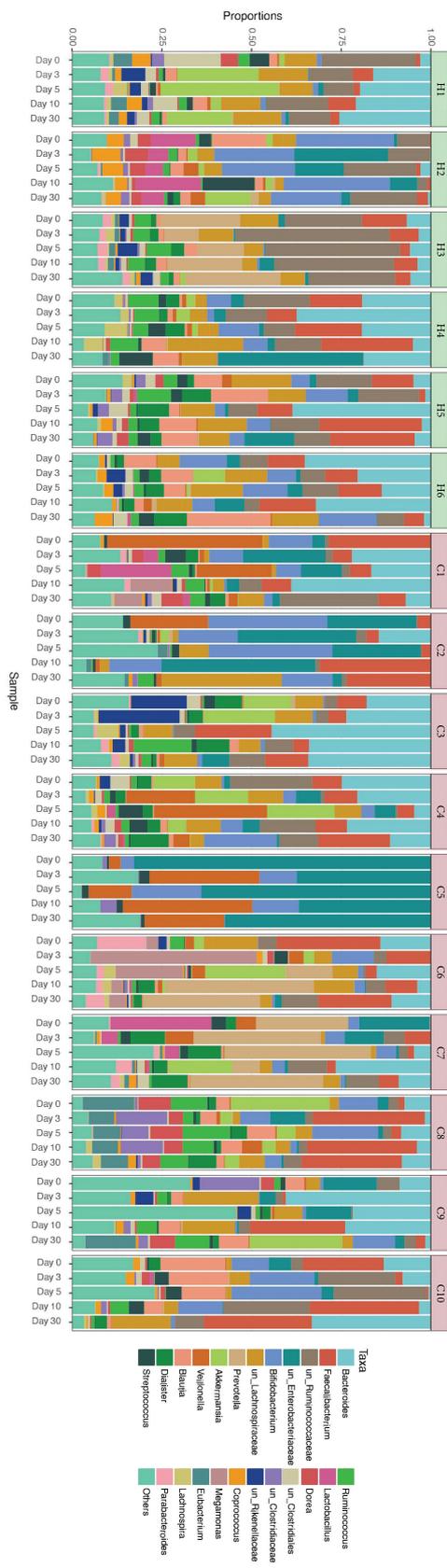


Fig. 6.1 Relative abundances of gut microbiota in every individual. Taxonomic relative abundances at genus level of each individual, separated per day of sampling. Healthy children (H1 to H6) are marked in green, and cases (C1 to C10) are marked in red. Only the 25 most abundant OTUs overall were represented. The less abundant OTUs were joined into the *Others* group.

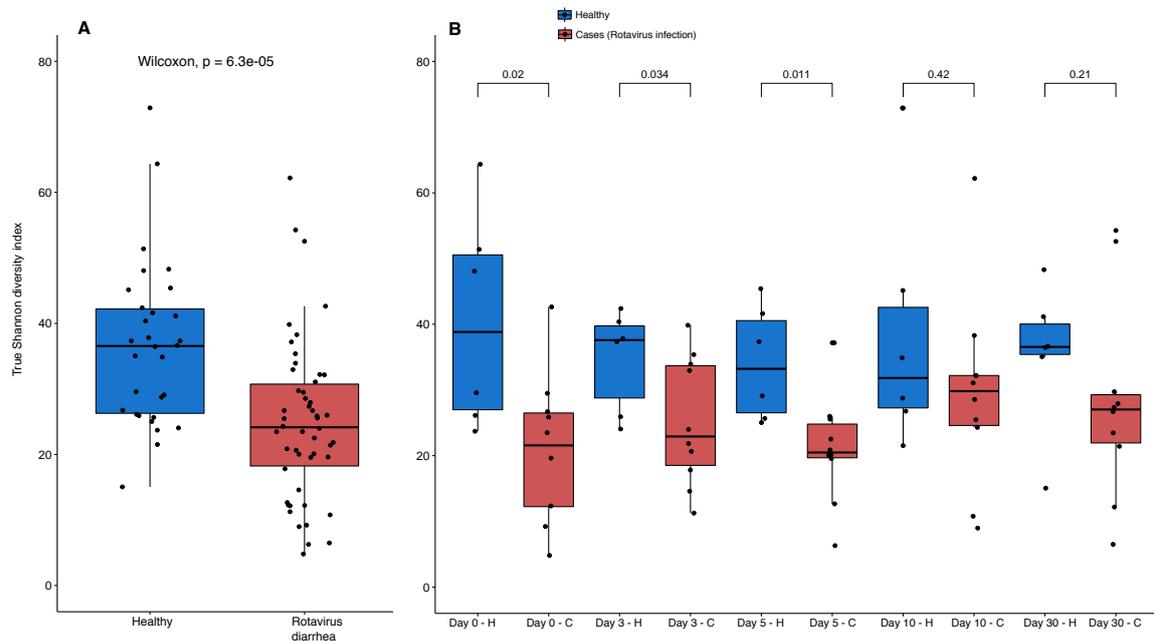


Fig. 6.2 Shannon diversity index per health status and per day Boxplots showing the Shannon diversity index between A) both health status with all times together; and B), separated by time of sampling. The colour blue represents healthy children, and red colour represents children with acute diarrhoea. P-value of Wilcoxon test is shown in the upper part of each comparison in both parts of the Figure, and the different health states are represented by H and C letters in part B for healthy and cases respectively. All data points were represented using the function jitter in R.

6.3.2 Taxonomic and functional enrichment

To determine which taxa were responsible for the differences between healthy status, we performed the LEfSe analysis test in the two subsets mentioned above (days 0 to 5, and days 10 to 30), which is represented in Figure 6.4 (parts A and B). We found that in the case of the subset of days 0, 3, and 5, there was an enrichment of bacteria belonging to phylum *Proteobacteria*, specifically the *Gammaproteobacteria* class. On the other hand, only phylum *Firmicutes* was abundant in the healthy children, in which we can find genera, such as *Blautia*, *Ruminococcus*, or other bacteria belonging to the *Lachnospiraceae* and *Ruminococcaceae* families. By contrast, the analysis of the second subset showed that only the genus *Blautia* was abundant in the healthy children. The results from PICRUSt revealed higher stability in functions, in general, concerning the variability observed at the taxonomic level (see Supplementary Figure A.9). The Kruskal-Wallis test showed the statistically significant differences in the metabolism of other amino acids at level 2 in KEGG hierarchy

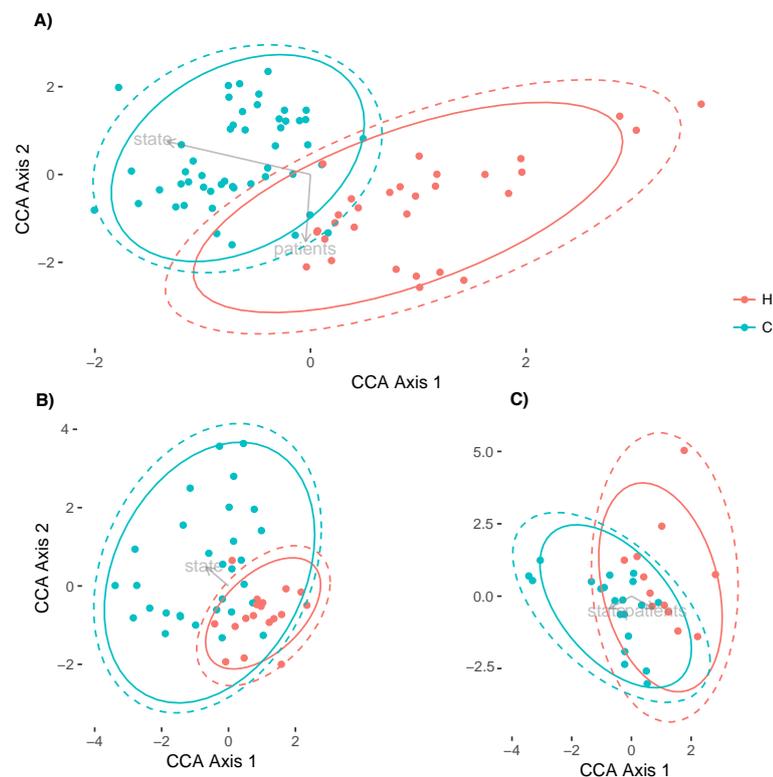


Fig. 6.3 Comparison of microbiotas between healthy children and children with acute diarrhoea. Three different representation of samples with Canonical Correspondence Analysis. In part **A**) is the global differences between both health status, Healthy (H, in red), and Cases (C, in green) for all times; in part **B**) we represented the CCA for the individuals from time 0 to time 5, and in part **C**) it is represented the CCA of samples belonging to days 10 and 30. Both health status were circled in their respective colours with no error and with 95% of Confidence Level in all Figure parts.

in the comparison of both health states. The same test was also significant for level 3 in KEGG, as we observed differences between four groups: the pentose phosphate pathway, the biosynthesis of ansamycins, ether lipid metabolism, and other ion-coupled transporters (see Figure 6.4, part C).

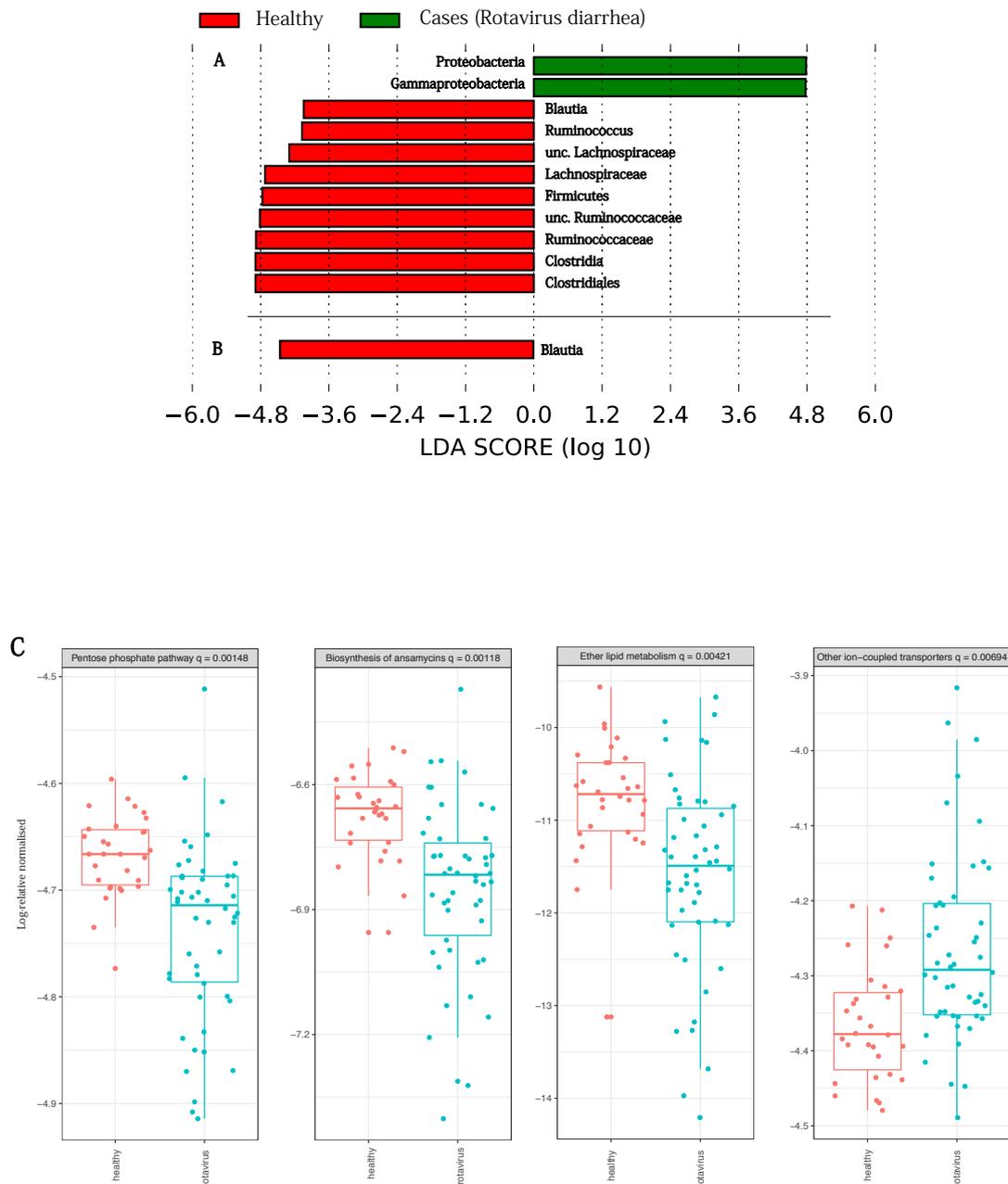


Fig. 6.4 Taxonomic biomarkers and functions enriched. Linear discriminative analysis (LDA) effect size (LEfSe) analysis between the healthy children (in red) and case children (in green), in **A**) from day 0 to 5, and in **B**) from day 10 to 30. LDA scores (log 10) for the most prevalent taxa in controls are represented on the negative side, whereas LDA-positive scores indicate enriched taxa in cases. In part **C**) are represented the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways at level 3 of hierarchy validated by Kruskal-Wallis test, specifically: pentose phosphate pathway, the biosynthesis of ansamycins, the ether lipid metabolism and other ion-coupled transporters.

6.3.3 Dynamic analysis of the microbiota

A temporal analysis was conducted to evaluate the temporal stability of the individuals over time. Of the two free parameters of the Taylor's Law fit, β was always less than 1 in all individuals, and we found pronounced differences between the values of the variability V . Figure 6.5 shows the standardisation of the parameters with respect to the healthy group. We can observe that all children suffering from acute diarrhoea (except child C8) are categorised out of the healthy zone, meaning they have a higher temporal variability than the healthy individuals of the study. The lower stability of children with rotavirus was in agreement with the other results, as their microbiomes have been suffering changes in time in respect to the healthy children, which have been stable through the sampling time. All the children who are outside of the *healthy zone* are, at least, two σ units distant from the centre in terms of V .

We also analysed the rank stability of the individuals. We represent the two samples with the least and the highest variability in the temporal analysis, which are H3 (a healthy child) and C6 (a non-healthy child), respectively (see Figure 6.6). We can see that the RSI has lower values in the non-healthy individual than in the healthy one. H3 has 33 taxa out of a total of 50, with an RSI above 70%, whereas C6 has only 8 taxa above that threshold. Some taxa also appear to be very stable despite their low proportions in the last 20 positions, another example of the *stability islands* found as a result in Chapter I. An example is the case of genus *Actinomyces*, which appears to be very stable at low proportions (around rank 40) in four of the six healthy children. There are other genera of interest, such as *Collinsela*, which also seems to be very stable in healthy children, but it alternates cases where it has low proportions with others in which is highly abundant in the affected children. Lastly, the case of *Blautia* is compelling; it is very stable and has high proportions in all healthy children, but it is absent in some of the children infected with the rotavirus and is found in low proportions in other cases. This finding is consistent with the LEfSe results. The general tendency of these less-abundant taxa is to belong to the phylum *Firmicutes*. Finally, as seen in Figure 6.6, we can observe that both the rank variability and the differences variability are significantly lower in H3 than in C6, an indicator that is also present in all healthy children compared with the non-healthy group. Finally, as seen in Figure 6.6, we can observe that both the rank variability and the differences variability are significantly lower in H3 than in C6, an indicator that is also present in all healthy children compared with the non-healthy group.

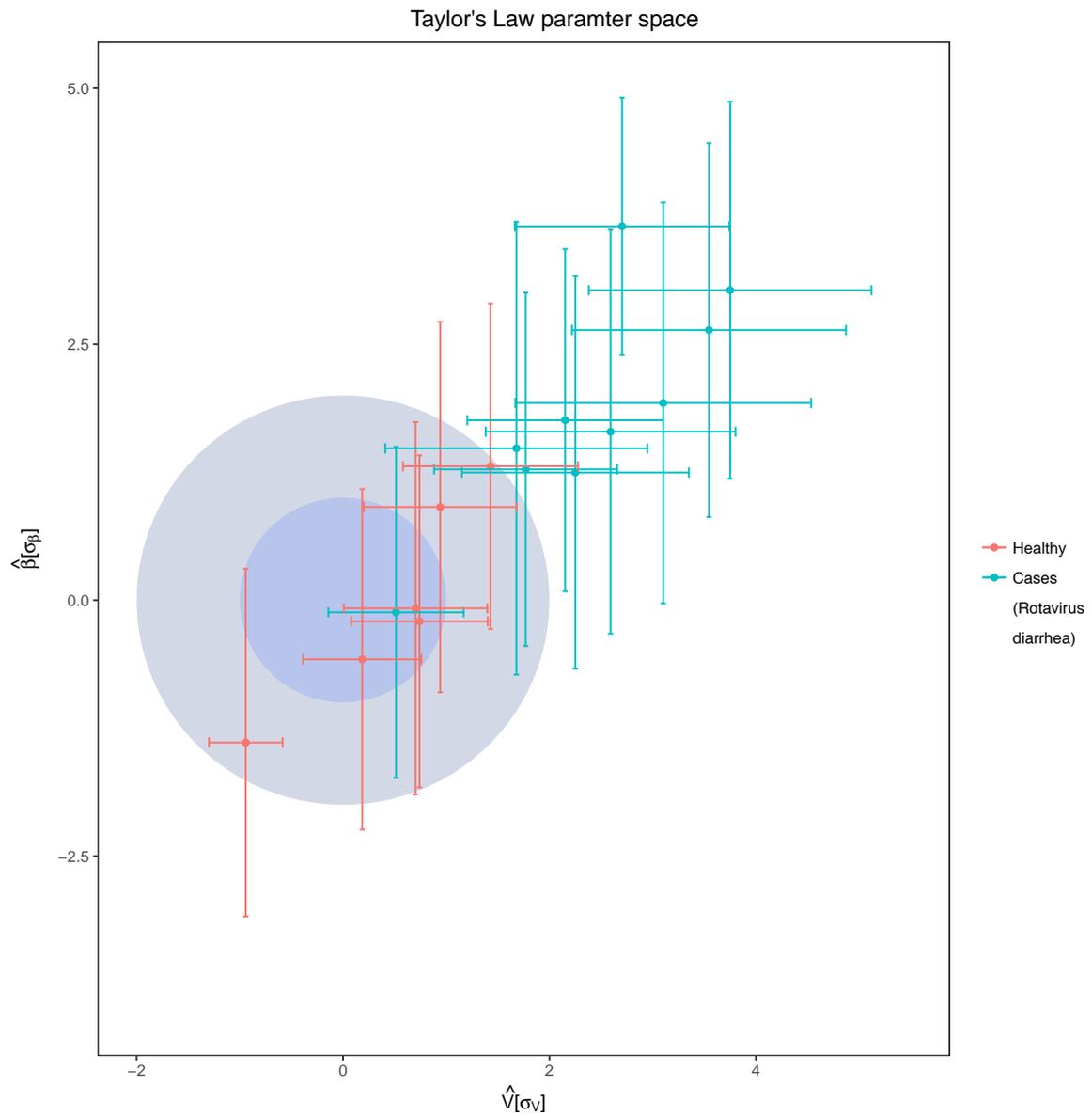


Fig. 6.5 Taylor's Law Parameter Space. The inner darker-blue circle corresponds to the 68% CL region of healthy children in Taylor's parameter space, while the bigger light-blue circle delimits the 95% CL region. In the Figure are represented both the healthy children (red points with error bars) and children with acute diarrhoea (green points with error bars). Taylor's parameters were standardised as mentioned in section 3.9.2 and they have standard deviation units.

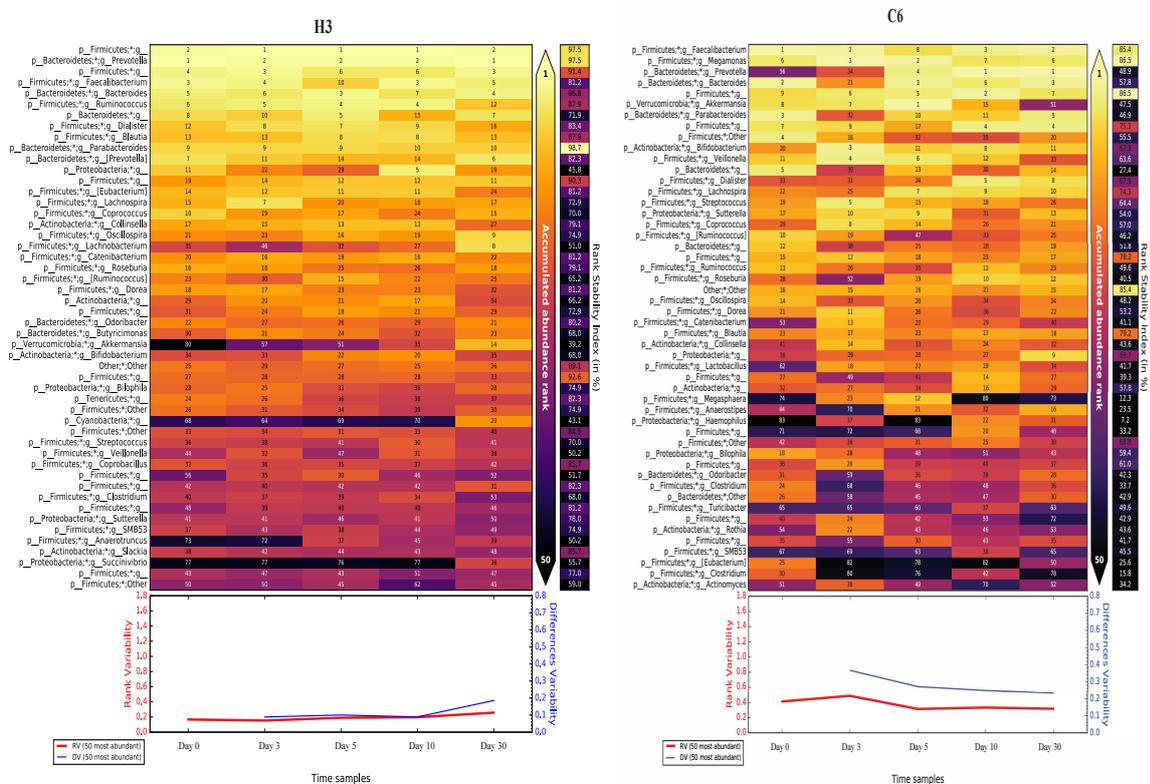


Fig. 6.6 Rank Stability Matrix. Rank Stability Matrix for the most variable subject (H3, part A), and the least variable subject (C6, part B). In both plots are represented the 50 most abundant genera of each case, and the numbers inside each cell represent the ranking of that specific genus at that specific time point. The colour inside each cell ranges from light-yellow for the rank 1, to black, representing very low ranks. At the right in each case it is shown the RSI, and below them, it is represented the Rank Variability (in red) and the Differences Variability (in blue).

6.4 Chapter Summary

As a summary of the results presented here, which will further discussed in the General Discussion, we can observe that the most abundant phyla in all children belonged to *Firmicutes* and *Bacteroidetes*, and the most abundant genus was *Bacteroides*. Concerning microbial diversity, we find marked differences in the α -diversity between healthy children and the ones that were suffering from the acute infectious diarrhoea in the first days of hospitalisation, from day 0 to day 5. Also, we find differences concerning β -diversity in the same period, implying that the microbial communities in both health states were different in composition

and abundance. Interestingly, both the α and β diversities recovered in the last days of hospitalisation, when the children who were suffering from the acute diarrhoea did not have symptoms anymore. These changes are represented in Figure 6.2 for α -diversity and Figure 6.3 for β -diversity. We can conclude that the affected microbiotas recovered their healthy status in time, as differences between both groups disappeared in the last days of hospitalisation.

The temporal analysis was based on the methodology developed in Chapter I, and as can be observed in Figure 6.5, there is a clear separation between healthy children and those who suffered from the acute infectious diarrhoea, being the unhealthy children more variable than the healthy group, except the case of participant C8. Exploring the rank stability from the children, we observe, again, the existence of *rank stability islands* as was also observed in Chapter I. As can be seen in Figure 6.6, we still find some genus as *Actinomyces* that are quite stable even their low overall abundance.

We found differences in the taxa enriched in each condition, in the first 5 days from intervention. We found that phylum *Proteobacteria* was more abundant in children with acute diarrhoea, and phylum *Firmicutes* was more abundant in healthy children. As happens with the diversities and dynamical analyses, we also found that these differences disappeared in the last two time points of the experiment, when the children stopped having symptoms and recovered their health status. The analysis of the functions encoded in the microbiomes shown differences between both groups, where we can find some metabolic pathways differentially expressed in both groups as the pentose phosphate pathway, the biosynthesis of ansamycins, the ether lipid metabolism and other ion-coupled transporters. All these analyses show that the microbial communities were different at the beginning of the intervention in both groups, but that the microbiomes from the children suffering from the acute diarrhoea recovered a healthy composition and status when those children improved their healthy condition.

Chapter IV: Bacterial individuality in oxidative stress profiles in the oral cavity

7.1 Abstract

Recent metagenomic studies suggest that oral diseases are caused by bacterial dysbiosis. However, the identification of the exact disease-causing bacterial consortium is complicated because the healthy individuals and patients share the same bacterial taxa. In the present chapter, we aimed to identify bacteria with pathogenic potential by correlation analysis with oxidative stress markers that reflect the inflammatory processes in the oral cavity. Saliva samples from 26 volunteers were collected daily during one month (551 samples) to quantify five oxidative stress markers and to profile temporal variability of the oral microbiome. The direct and lagged correlations between the markers and the bacterial taxa were computed individually for each volunteer. The results indicate that a particular marker-taxa pair can have negative correlations in some volunteers while positive in others. *Streptococcus mutans* and certain species of *Actinomyces*, *Rothia* and *Cardiobacterium* genera had exclusively either positive or negative correlations with some markers. However, these correlations were not confirmed in all volunteers. The distinct intra-individual correlation patterns suggest that different bacterial consortia are involved in the oxidative stress induction in each volunteer. These results emphasise the importance of longitudinal samples collection for correlation studies, and suggest that the oral diseases treatment should be more individual-oriented.

Original work:

Dzunkova M*, Martínez-Martínez D*, Gardlik R, Behuliak M, Jansakova K, Jiménez N, Vázquez-Castellanos JF, Martí JM, D'Auria G, Bandara HMHN, Latorre A, Celec P, Moya A. Oxidative stress in the oral cavity is driven by individual-specific bacterial communities.

In revision.

*Equal contribution

7.2 Material and Methods

A general view of the material and methods were explained in Material and Methods Chapter. Nevertheless, the most relevant information is described in the following subsections to facilitate the understanding of the results. Detailed information has been provided when necessary.

7.2.1 Saliva samples and Oxidative stress markers

Twenty-six volunteers (13 women and 13 men) of age between 21-28 years participated in this study. On average 21 non-stimulated saliva samples (2 ml) were collected from each participant during the period of 30 days. Five different oxidative stress markers were measured: lipid peroxidation was quantified by measuring thiobarbituric acid reacting substances (*TBARS*); advanced glycation end products (*AGEs*) and advanced oxidation protein products (*AOPP*) were herein quantified as a carbonyl stress marker and as an oxidative stress marker, respectively; the capacity to resist oxidative damage was measured by total antioxidant capacity (*TAC*) and ferric reducing ability of saliva (*FRAS*). The oxidative stress markers were quantified using plate reader Safire II (Tecan, Austria) as described by Banasova *et al.* [197]. More information is available in sections 3.1.3 and 3.4.

7.2.2 16S rRNA gene sequencing

The regions V3 and V4 of the 16S rRNA gene were amplified using primers including Illumina adapter sequences with Kapa HiFi HotStart polymerase (Ref. B4kk2602) and the samples were multiplexed by 96 index combination and sequenced on MiSeq Illumina platform (Ref. MS-102-3001). The Illumina paired-end reads were merged and chimeric amplicons were removed by *usearch* program [187]. Sequences shorter than 50 bp were removed. In addition, the samples with less than 2,500 sequences were removed from the analysis. The obtained sequences were deposited to the EBI database with the study accession number: PRJEB20213 (ERP022351).

In the first step, we determined which are the most suitable settings for clustering of sequences into OTUs [244] for obtaining separate sequence clusters of *Streptococcus mutans* and *Streptococcus parasanguinis*. This testing was crucial for distinguishing between caries-associated *S. mutans* [99] and commensal *Streptococcus* species not associated with caries [245]. For this purpose, all the 16S rDNA sequences belonging to the genus *Streptococcus* were downloaded from the Ribosomal Database Project website (RDP, January 2016), clipped for the region V3 and V4 (regions which are used for sequencing in this study) and the

clustering on different similarity levels (0.95, 0.96, 0.97, 0.98, 0.99) was tested using the *usearch* program. The similarity level 0.96 produced homogeneous clusters (OTUs) corresponding to separate *Streptococcus* species. Afterwards, all the obtained sequences from the 26 volunteers have been clustered together on 0.96 similarity level by the *usearch* program and the reference sequences of each OTU cluster have been taxonomically assigned by RDP classifier [246]; the annotation was accepted if the bootstrap confidence estimation value was over 0.8. The final OTU names used in this study consisted from the assigned genus of the reference sequence and the number of the cluster.

7.2.3 Analysis of the bacterial composition

The resulting bacterial composition was analysed in the R programming environment using packages *vegan* [204] and *ade4* [247]. As the clustering of large datasets into OTUs often results in a formation of artificial low abundant OTUs corresponding to sequencing errors [248], the dataset containing all detected OTUs was compared to the datasets containing only OTUs with average proportion $>0.001\%$, $>0.01\%$ and $>0.1\%$. In the datasets with reduced OTUs numbers, the low prevalent OTUs were discarded, while the proportions of the OTUs with higher prevalence were not altered. The Procrustes test (R library *vegan*, function “protest”) was performed to test whether the datasets with different OTU numbers produce the same ordination of samples in the nonmetric multidimensional scaling (NMDS) with Bray-Curtis dissimilarity.

Shannon diversity index, Pielou’s evenness index values were compared among volunteers using t-test with Holm correction of p-values using the R package *vegan*. In the next step, the “envfit” function from *vegan* R package was used for testing of the bacterial composition (OTUs with average proportion $>0.1\%$) for fitting on variable “volunteer” in the canonical correspondence analysis (CCA). This analysis was used to test whether each volunteer has its characteristic microbiome and whether the microbiomes differ significantly between volunteers. More information about diversity can be found in section 3.5.

7.2.4 Temporal variations of the microbiome

The temporal analysis was conducted according to the results of the Chapter I to measure the stability, over time, of all individuals. Taylor’s parameters were computed in the same manner as explained in section 3.9. We also computed the rank stability of all volunteers. For this purpose, we calculated the rank stability index (RSI), which shows the temporal stability of a particular taxon, by considering its rank in the total population. We also calculated the *rank variability* as well as the *differences variability*, discussed in Methods.

7.2.5 Correlations between salivary markers and microbiota

Pearson's and Spearman's temporally direct correlations and Local Similarity Analysis for lagged correlations were used to assess correlations between bacterial OTUs in oral cavity of each volunteer individually taking into account the temporal dimension. Local Similarity Analysis checks the existence of a delay in the correlations between all possible combinations of variables. Pearson's, Spearman's, and LSA correlation coefficients were computed using eLSA Python package from Xia *et al.*, [145], with default parameters. This package computes permutation tests that are helpful when multiple hypothesis tests are being studied in this kind of problems. The correlation values >0.3 and <-0.3 were then filtered by their p-value at a level of 0.05 for statistical significance. Correlations in this study have been plotted using R packages *corrplot* [220] and *beeswarm* [221]. In addition, interactions between the 15 most abundant OTUs per sample were calculated with LIMITS algorithm from Fisher and Mehta work [114] in Wolfram Mathematica software (version 11.0) [222].

7.2.6 Longitudinal sample collection vs. single sample collection

In most of the biomedical studies, only one sample per volunteer is collected and tested for correlations with its corresponding biomedical data. Such an approach ignores the temporal intra-individual variability of the microbiome composition and may lead to contradictory results obtained by different research groups [249]. To test our data, we simulated such a one-sample-one-volunteer approach and compared it with the result of the intra-individual correlation analysis based on our longitudinal sample collection described above. For that purpose, only one sample per each volunteer was selected randomly for a total 10,000 combinations and then tested for Pearson's correlations between salivary markers and OTUs proportions. The correlation values >0.3 and <-0.3 were also filtered by their p-value at a level of 0.05 for statistical significance.

7.3 Results

7.3.1 Bacterial composition

In total, 27,072,746 sequences of 16S rDNA amplicons from 551 samples collected from 26 volunteers have passed the quality filters and have been clustered into 18,146 operational taxonomic units (OTU). The microbiome of all volunteers was dominated by the *Streptococcus*-OTU0 (highest similarity with *Streptococcus parasanguinis*) forming on average $37.51 \pm 10.14\%$ of the whole microbiome (Supplementary Figure A.10) followed by taxon

Rothia-OTU1 ($20.91 \pm 11.31\%$). The ratios of *Streptococcus*-OTU0 and *Rothia*-OTU1 differed significantly ($p < 0.001$, t-test) among the 15% of volunteers. For example, the proportion of *Rothia*-OTU1 could be as high as $35.30 \pm 10.51\%$ in volunteer F06 or as low as $5.43 \pm 5.24\%$ in volunteer M12.

The overall microbiome composition of the volunteers differed significantly due to the less prevalent (average proportion 0.1 - 5.4%) individual-specific taxa ($p < 0.001$, “envfit” test, Supplementary Figure A.11). Some volunteers (e.g. F18, F20 and M20) were characterised by high proportions of *Granulicatella*-OTU2 ($8.66 \pm 4.13\%$) and *Atopobium*-OTU3 ($6.88 \pm 3.72\%$), while other volunteers (e.g. M12, M21) had high proportions of *Gemella*-OTU6 ($4.17 \pm 3.19\%$). In contrast, in some volunteers (e.g. F16), *Rothia*-OTU1 was very prevalent ($31.85 \pm 14.42\%$), while *Saccharibacteria*-OTU4 and *Gemella*-OTU6 were in very low proportions ($0.25 \pm 0.30\%$ and $0.37 \pm 0.25\%$). The volunteer F12 had the highest proportion of *Saccharibacteria*-OTU4 ($8.93 \pm 4.57\%$). *Saccharibacteria*-OTU4 and *Atopobium*-OTU3 were very prevalent ($4.82 \pm 3.79\%$ and $6.69 \pm 4.35\%$, respectively) in the volunteers F21, M08 and M19. Some volunteers did not possess any extreme proportions of the most detected prevalent OTUs (central part of the CCA plot in the Supplementary Figure A.11).

The Shannon index (mean of 2.46, and standard deviation of 0.38) and the evenness index (mean of 0.48, and standard deviation of 0.06) also possessed substantial inter-individual differences (Supplementary Figure A.12). The volunteer M04 had the highest Shannon and evenness indexes (2.80 ± 0.35 and 0.56 ± 0.06 , respectively), what differed significantly ($p < 0.001$, t-test) from the volunteer M20 with the lowest average Shannon and evenness indexes (2.24 ± 0.53 and 0.43 ± 0.08 , respectively). The Procrust test showed that the inclusion of OTUs with proportions of above 0.1% (50 OTUs) provide the same result as the inclusion of all detected OTUs (Supplementary Figure A.13). Thus, the following correlation analyses were performed with the 50 most abundant OTUs.

7.3.2 Temporal stability of the microbiome

All the values of the Taylor’s parameter β were between 0.5 and 1 which means that the most prevalent OTUs in all samples showed less relative variability over time than the less abundant OTUs, as explained in section 4.3. The microbiomes of the 26 volunteers had an average of the Taylor’s parameter V values of 0.27 ± 0.06 . Significant differences ($p < 0.001$, t-test) of V between volunteers existed (Figure 7.1) in 26% of pairwise comparisons. In terms of V values, the volunteer F11 had the least stable microbiome of our cohort ($V = 0.47 \pm 0.09$), while the microbiome of F18 was the most stable ($V = 0.20 \pm 0.03$).

The temporal variability of a microbiome in time can be expressed as an absolute difference between every taxon’s rank (proportion) at a specific time point compared to the

previous time point; it is expressed as difference variability (DV). DV had some peaks in all volunteers (Supplementary S5). Figure 7.2 shows that the DV was generally higher in the least stable volunteer F11 when compared to the most stable volunteer F18 (Figure 7.2). This is expected because there are more rank differences in subjects with higher Taylor's parameter V (F11) than in subjects with lower fluctuations (F18).

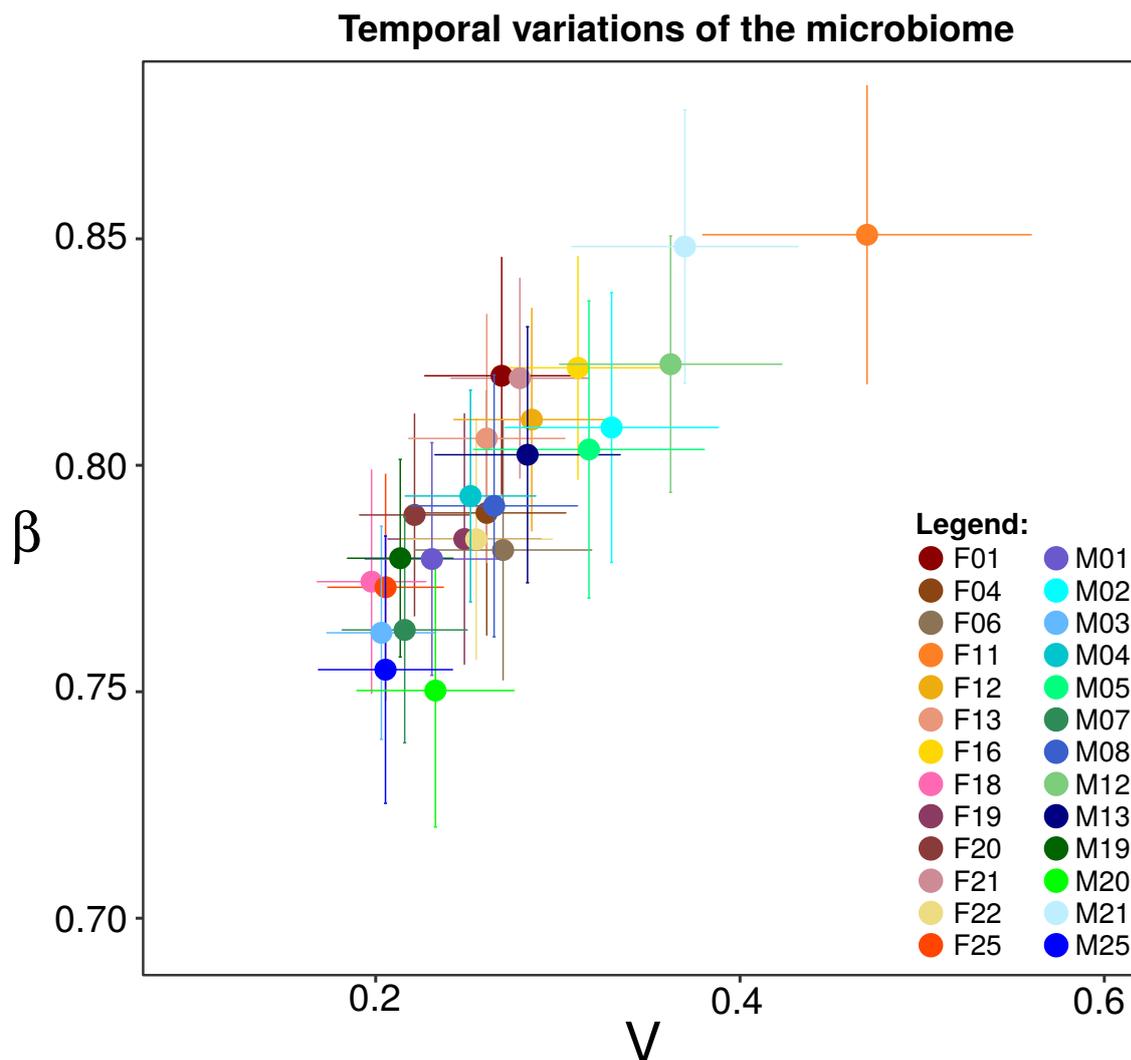


Fig. 7.1 Temporal variability of the microbiome. Taylor's Parameter space for the 26 volunteers of the study. V represents the y-intercept of the linear fit, and β to the slope of the line. Each individual has been placed in this plot according to its V and β value, where the error bars correspond to the SEM.

The most prevalent OTUs, *Rothia*-OTU1 and *Streptococcus*-OTU0, had the rank stability index (RSI) over 96% in both volunteers F11 and F18 (Figure 7.2). It indicated that the most prevalent OTUs were very stable in the volunteer with the most stable microbiome as well as in the volunteer with the least stable microbiome. The microbiome temporal variability of the volunteers F11 and F18 differed due to changing proportions of the less prevalent OTUs. The average RSI of all 50 OTUs in volunteers F11 (the least stable) and F18 (the most stable) were $52.35 \pm 20.11\%$ and $67.71 \pm 19.20\%$, respectively. The volunteer F11 had 26 OTUs with RSI below 50%, while the volunteer F18 had only seven such highly unstable OTUs. When the OTUs were sorted according to their prevalence, the last OTU with an RSI above 70% in F11 was *Streptococcus*-OTU730 placed in the 16th position, and the last OTU above 70% in F18 was *Streptococcus*-OTU56148, placed in the 49th position (Figure 7.2). The microbiome stability profiles of some selected volunteers in the Supplementary Figure A.14 show that the OTUs found in lower proportions in the microbiome were, in general, less stable than the highly abundant OTUs. However, the most prevalent OTUs were not always the most stable. For example, *Saccharibacteria*-OTU15 and *Rothia*-OTU15862 were in high proportion in F11 and F18, but it possessed a low RSI in both cases (Figure 7.2).

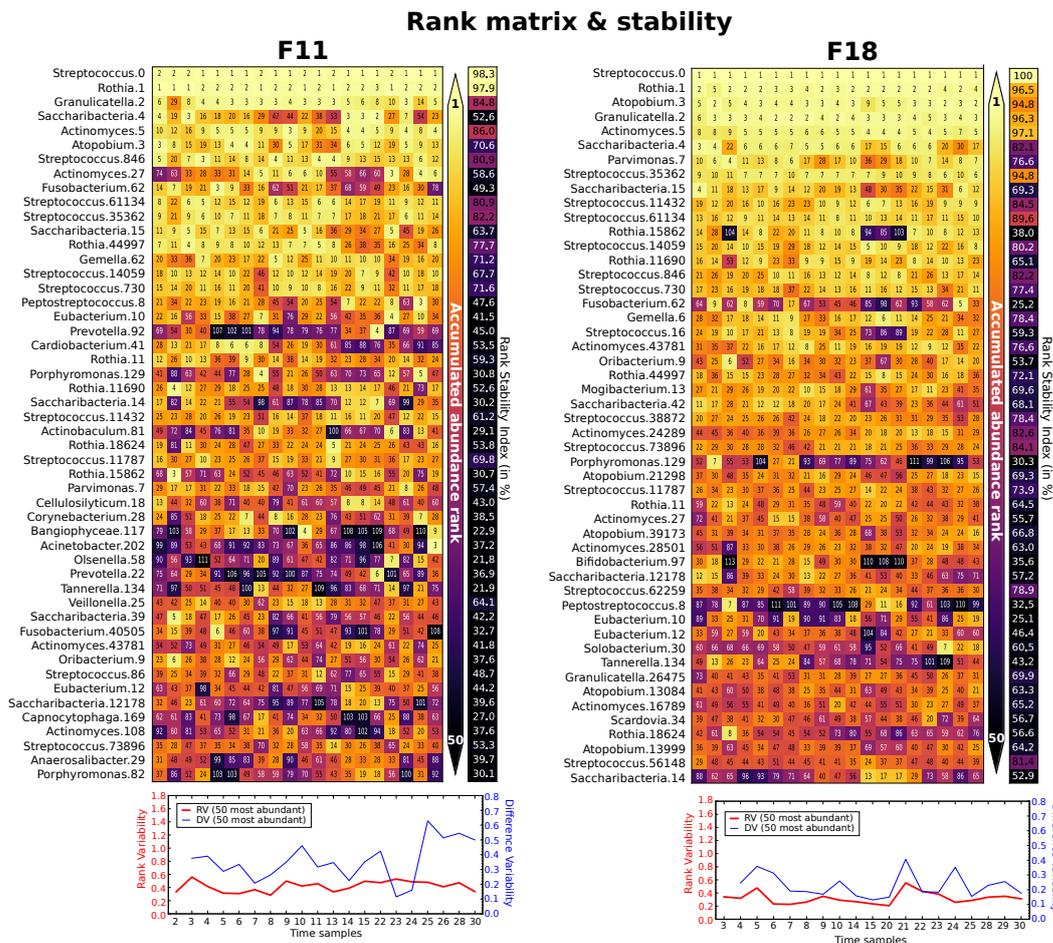


Fig. 7.2 Rank matrix for the 50 most abundant OTUs for F11 and F18. Rank matrix corresponding to the most and less time-variable volunteers, F11 and F18 respectively. Heat-map colours corresponds to the abundance of each OTU at each time. Alongside, the Rank Stability Index (RSI) is coloured by the percentage of rank stability. Below, both Rank and Difference Variability is plotted in red and blue colours for each time point.

7.3.3 Large inter- and intra-individual difference of the oxidative stress markers levels

Oxidative stress can be measured by estimating oxidative damage to lipids (lipid peroxidation) and proteins (protein oxidation), or by quantifying the capacity to resist oxidative damage (antioxidant capacity). The lipid peroxidation was quantified by measuring thiobarbituric acid reacting substances (TBARS, average $0.10 \pm 0.18 \mu\text{mol/L}$, Supplementary Figure A.15). Advanced glycation end products (AGEs, $0.27 \pm 0.19 \text{ g/L}$) and advanced oxidation products (AOPP, $37.6 \pm 21.8 \mu\text{mol/L}$) were herein quantified as carbonyl and oxidative stress

markers respectively, to express oxidative protein damage. The capacity to resist oxidative damage was measured by total antioxidant capacity (TAC, $578.6 \pm 149.4 \mu\text{mol/l}$) and ferric reducing ability of saliva (FRAS, $396.7 \pm 183.6 \mu\text{mol/L}$). The values of the five markers possessed intra-individual temporal variations, and they also differed significantly ($p < 0.001$, t-test) among volunteers, in 44 - 73% of volunteers pairwise combinations (Supplementary Figure A.15).

In addition, the five oxidative stress markers were tested for their pairwise correlations on intra-individual level. FRAS correlated positively with TAC in 14 out of the 26 volunteers, but the majority of the salivary markers pairwise combinations resulted in non-uniform correlation patterns on intra-individual level among the 26 volunteers (Figure 7.3). Correlations lagged by 1-3 days accounted for 36.7% of all significant correlations detected, while the remaining 63.3% were temporarily direct correlations.

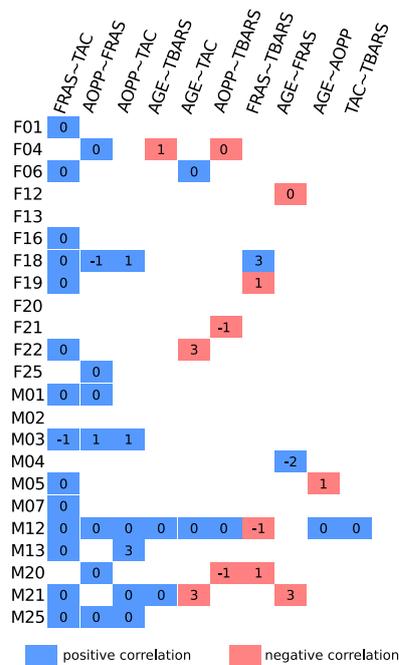


Fig. 7.3 Correlations between the oxidative stress markers. The summary of intra-individual correlations that can also be found in correlation plots in the Supplementary Figure A.16. The colour of square indicates either a positive or a negative correlation (blue or red) of a pair of oxidative stress markers in a volunteer. The number of the squares indicates the days by which the correlation was lagged. Empty white squares indicate no significant correlation. The majority of volunteers possessed a positive correlation between FRAS and TAC. Several marker pairs had positive correlations in some volunteers while negative correlation in others.

7.3.4 Correlations between bacterial taxa and oxidative stress markers are individual-specific

Correlations of oxidative stress markers and OTUs were analysed on the intra-individual level: five markers were correlated with 50 OTUs in 26 volunteers (Figure 7.4, details in Supplementary Figure A.16). The analysis showed that 94% of the 250 marker – OTU pairs (5 markers x 50 OTUs) had a significant correlation in some volunteers (1-6 volunteers). These correlations were either temporarily direct or lagged by 1-3 days. The remaining 6% of the marker – OTU pairs did not show any significant correlation in any of the volunteers.

The results of the Pearson's correlations showed that a particular marker-OTU pair can have positive correlations in some volunteers, while negative correlations in others (Figure 7.4, details in the Supplementary Figure A.16). This result is consistent with Spearman's correlation coefficients, a non-parametric correlation (Supplementary Figure A.16). The volunteers with contradictory correlation results did not possess any extreme values of a given marker-OTU pair.

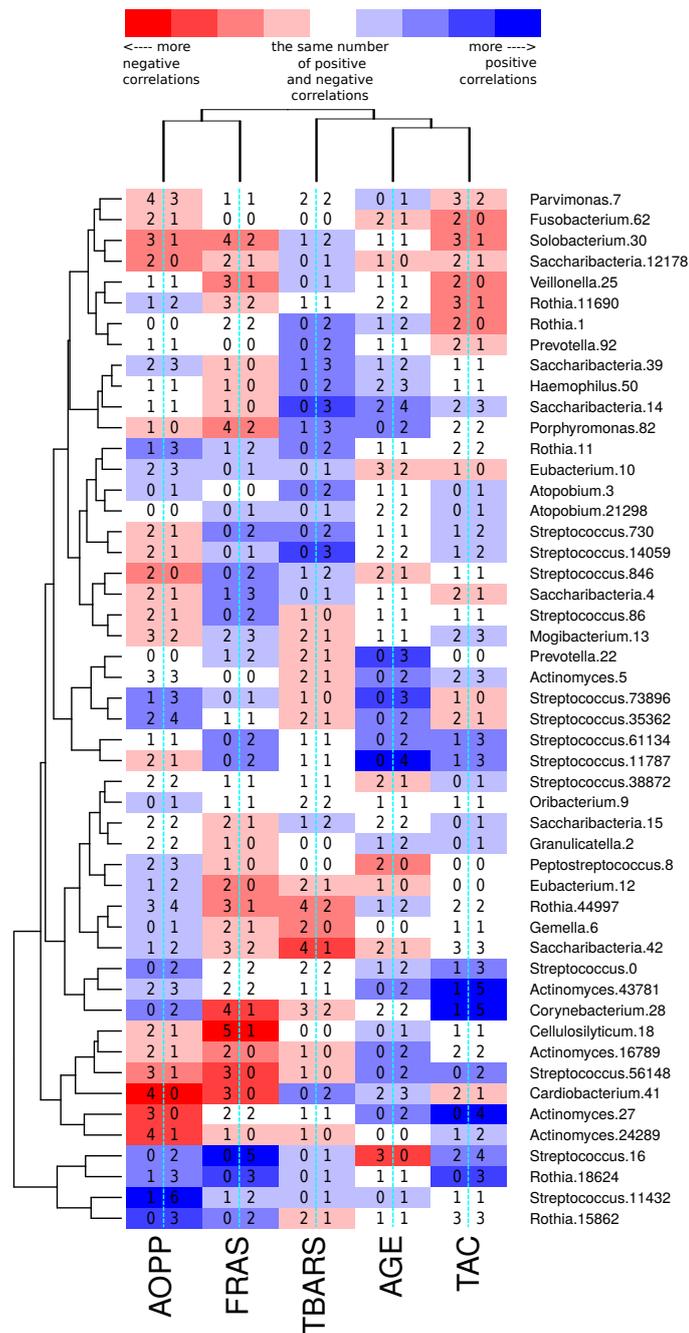


Fig. 7.4 Marker-OTU correlations on the intra-individual level. The summary of the correlation plots in the Supplementary Figure A.16. The graduated colour in tones from blue to white to red indicates whether the correlations found in the volunteers on intra-individual levels were mostly positive or negative. The number on the left in each rectangle indicates the number of volunteers with a negative correlation, while the number on the right in each rectangle indicates the number of volunteers with a positive correlation. The oxidative stress markers and the OTUs have been ordered according to their correlation tendency by hierarchical clustering using Euclidean distances. The OTUs with the most robust tendency of either positive or negative correlations with some of the oxidative stress markers were, e.g. *Streptococcus*-OTU16, *Actinomyces*-OTU27 and *Cardiobacterium*-OTU41.

However, there were also some marker – OTU pairs that showed only positive or only negative correlations throughout the cohort (Figure 7.4, details in Supplementary Figure A.16). The most prominent exclusive correlations with oxidative stress markers had *Streptococcus*-OTU16 (*S. mutans*), which has been previously reported to be a causative agent of dental caries [99]. *S. mutans* correlated positively with AOPP (in volunteers F18 and F25), with FRAS (F16, F18, F20, F25, M02) and with TBARS (F13). Furthermore, *S. mutans* exhibited an exclusively negative correlation with AGE (in volunteers F01, F12, M21). Correlations of *S. mutans* with the fifth marker TAC were ambiguous: it had a negative correlation with TAC in two volunteers (M05, M25) and a positive correlation in four volunteers (F18, F25, M04, M12).

S. mutans formed as few as $0.19 \pm 0.47\%$ on average in the studied cohort. The volunteer M04 had the highest average proportion of *S. mutans* ($0.71 \pm 0.90\%$), followed by F16, F25, F18 and F06. The lowest proportion of *S. mutans* (average $<0.01\%$) was found in volunteers F12, F19, F20, M07, M19 and M25. It suggests that significant correlations of *S. mutans* with the oxidative stress markers were independent of its proportion in the microbiome. Among other marker-OTU pairs with exclusively negative correlations found in higher number of volunteers (3-6) were e.g. AOPP - *Cardiobacterium*-OTU41, AOPP - *Actinomyces*-OTU27, FRAS - *Streptococcus*-OTU56148 and FRAS - *Cardiobacterium*-OTU41. The marker-OTU pairs resulting in exclusive positive correlations were e.g. AOPP - *Rothia*-OTU15862, FRAS - *Rothia*-OTU18624, AGE - *Streptococcus*-OTU11787 and TAC - *Actinomyces*-OTU27 (Figure 7.5).

7.3.5 Correlations between the bacterial OTUs

Similar to the marker - OTU pairs, the significant intra-individual correlations found among the OTU - OTU pairs did not show unusual generalised patterns (Supplementary Figure A.15). For example, the negative correlation between *Streptococcus*-OTU0 and *Rothia*-OTU1 was found in 17 out of 26 volunteers, while in the remaining nine volunteers no significant correlation was detected for this OTU - OTU pair. Furthermore, the analysis of interactions based on Lotka-Volterra system of differential equations was used to infer the relationships of the 15 most prevalent OTUs in each volunteer. However, they did not show any generalised correlation pattern either (Supplementary Figure S8).

Also, we analysed the interactions of the 15 most abundant OTUs of each volunteer using the generalised Lotka-Volterra model, a system of equations that has been often used for the inference of bacterial interactions in complex ecosystems such as the human microbiome [114]. These models allow us, when a temporal series is available, to make better predictions of bacterial interactions rather than the use of classic correlation coefficients that could be

filled by false positives and false negatives due to compositional effects. The results from the Lotka-Volterra model are more accurate, but it did not yield any generalised interaction pattern either (Supplementary Figure A.17).

7.3.6 Simulation of single sample collection

From each of the sample collections belonging to the 26 volunteers, single samples were selected and combined to form a set of 26 samples in 1,000 repetitions (1,000 different samples combination), and each set was tested for correlations of oxidative stress markers with OTUs. The majority (82.4%) of the 250 marker-OTU pairs (5 markers x 50 OTUs) produced contradictory results. For example, a negative correlation was found in 22.2% of these single sample combinations for the TBARS - *Rothia*-OTU15862 pair, but still 0.1% of combinations for this marker - OTU pair resulted in a positive correlation. Meanwhile, 77.7% of the combinations did not yield any significant correlation (Figure 7.5).

The results of the correlation analyses performed by this one-sample-one-volunteer approach simulation (Figure 7.5) and the correlation results obtained on the intra-individual level (Figure 7.4) were quite inconsistent. For example, *Rothia*-OTU11690 had negative correlations with TBARS in 31.2% of the 1,000 one-sample-one-volunteer combinations, however, on intra-individual level this negative correlation was detected only in the volunteer M07, while the volunteer M13 had a positive correlation in this marker - OTU pair. Another example is AGE - *Actinomyces*-OTU27 pair that exhibited exclusively negative correlation in 26.6% of the one-sample-one-volunteer combinations, however, it showed a positive correlations in two volunteers (F04, M01) on intra-individual level. Furthermore, the exclusive positive correlation of *S. mutans* (*Streptococcus*-OTU16) with FRAS noted in five volunteers on intra-individual level, was not confirmed by this one-sample-one-volunteer approach (only 2.2% of positive correlations and 0.3% of negative correlations in the 1,000 combinations). In contrast, in some cases the correlation results obtained by the two approaches were quite consistent; e.g. negative correlations between AGE - *S. mutans*, AOPP - *Actinomyces*-OTU27, AOPP - *Cardiobacterium*-OTU41 and FRAS - *Cardiobacterium*-OTU41.

7.4 Chapter Summary

As a summary of the results presented here, which will further discussed in the General Discussion, we can observe that OTUs belonging to *Streptococcus* and *Rothia* genera were the most abundant taxa overall, although we also found volunteers with high levels of *Granulicatella* or *Atopobium*. The α -diversity, measured as the Shannon index, was variable

between volunteers. From the temporal stability, measured by the methodology developed in Chapter I, we found that β was always less than one in agreement with the results from Chapter I and Chapter III. Volunteers F11 and F18 had the higher and lower variability respectively, and from the rank stability analysis, we found that *Streptococcus* and *Rothia* were the most rank stable taxa in both volunteers. However, both volunteers differed in terms of the number of stable OTUs. As a general result, the variability shown in all volunteers was usually lower than the variability shown by the non-healthy microbiomes in Chapter I.

Concerning the oxidative stress markers, we applied a methodology for detecting correlations that could be lagged in time, as the production of these components could be delayed for several unknown reasons. We did not find any general pattern between oxidative stress markers (see Figure 7.3) nor between stress markers and OTUs (see Figure 7.4). In fact, in Figure 7.4 we can see that in many of the marker-OTU pairs, there were cases in which the correlation was negative in some volunteers while positive in others. In any case, the OTUs with the most robust tendency of either positive or negative correlations with some of the oxidative stress markers belonged to *Streptococcus*, *Actinomyces* and *Cardiobacterium* genera. Also, the interactions inferred with the Lotka-Volterra model shown individuality in each volunteer, adding support to a lack of a general pattern of correlations/interactions. Finally, we provided an interesting methodology to highlight the importance of measuring time series in microbiome studies. Figure 7.5 shows that, in a microbiome study where we only measured one time per volunteer, we would have missed much information and we would have had false positives and false negatives.

General Discussion

In this section, it will be discussed all the results obtained through this thesis, with the aim to integrate the information of every chapter in an organised manner.

The **Chapter I** was focused on the study of the variability in the human gut microbiome, and its relationship with the health status of the host. One of the highlights of this study, and of the entire thesis, is that, independently of its condition, *every microbiota studied here follows Taylor's law*. We have seen that in each case the value of the scaling index, the parameter β , is always less than the unity, which provides us with information about the community structure. This means that, in relative terms, the most abundant elements in the population are less volatile to perturbations than the less abundant ones. The explanation for this universal pattern is not clear although some hypotheses have been tested in other studies, such as the presence of negative interactions in the population [250], and a demonstration that this may depend on reproductive correlation [251]. Nevertheless, none of these explanations are sufficient when it comes to microbiota, as the term reproduction is diffuse, and the interactions between its components are not only based on competition [79, 121, 114]. Moreover, even such negative interaction may not effectively yield values less than the unity when referring to a bacterial species. Nonetheless, the values obtained in all cases were very similar to each other, which may suggest that the community structure is preserved throughout the different scenarios studied herein.

The parameter V provides information about noise and can be directly linked to the variability or fluctuation amplitude of the population over time. It is a direct estimator of the stability of the system under study. As we have shown in Chapter I and Chapter III, the healthy subset of each study has lower variability than the non-healthy subset, when dealing with adult subjects. Interestingly, the variability parameter was higher in the healthy subset in the study of discordant twins suffering from kwashiorkor disease [169]. In this respect, research has shown that infant microbiota needs to develop toward a definite, adult state [252]. This implies that temporal variability is higher in children than in a healthy adult, which should be temporally stable. Thus, our results could indicate this variability is

necessary to reach that adult state. Furthermore, as we wanted to see how this variability shifted over time, we calculated the changes in V for the samples which had enough time sampling Chapter I. As shown in Figure 4.6, the variability of microbiota fluctuated over time. Interestingly, Figure 4.7 shows how this parameter reflected the two antibiotic intakes in one of the patients in the study by Dethlefsen and Relman [120] particularly the apparent resilience of the microbiota due to the reduced increase in variability during the second antibiotic intake.

The primary hypothesis of Chapter I is that, in adults, having a healthy microbiota means that the microbial population is stable over time. This stability means the microbiota does not shift and become susceptible to external or internal perturbations causing dysbiosis. In order to use the valuable information provided by the empirical law of Taylor's work, herein we have proposed the use of Langevin's equation to model how stability ranking changed over time. While the system noise component can be directly measured by its variability, the other main term needs to be inferred from the model. This term, which we have named "fitness", enables the system to remain stable when confronted with potential perturbations. In ecological terms, this could represent the nature of interactions present among bacteria, between bacteria and other minority populations, such as fungi or archaea, between bacteria and the viral component in the microbiota, and interactions between the host and the whole microbiota. As this is the first step to model the temporal stability of microbiota, and given its complex nature, we calculated fitness using the Fluctuation–Dissipation Theorem as a first approximation for demonstrative purposes [227]. Thus, future works are required to model the fitness of microbiota to provide a more accurate model with higher predictive power.

By solving Langevin's differential equation, we obtain a phase diagram where each microbiota sample can be placed on its fitness and variability into one of two phases, according to the stability ranking of the system. As shown by the phase–space in Figure 4.3, three different conditions can occur. The first is a healthy microbiota with some fluctuations, as shown by one of the subjects in Caporaso *et al.* study [47]. Because this case would have good fitness, its temporal variability would not place the microbiota in the unstable phase of the diagram. Secondly, we have a subject from the study by Dethlefsen and Relman [120] whose microbiota was perturbed twice by an antibiotic intake, undergoing sufficient change to lose its stability, and hence be placed in the unstable part. In this location, it is more sensitive to potential perturbations such as opportunistic infections. In the third and last condition, the subject was already in the unstable phase due to a health issue, i.e., IBS. This can be observed in one of the patients in Durban *et al.* [73]. Besides, it was shown that this subject's health status improved during the experiment, implying that his/her microbiota also recovered stability. Interestingly, in the study made by David *et al.* [59] the subject who

had a *Salmonella* infection during the experiment underwent a significant shift in variability with eventual recovery from the perturbed state (see Supplementary Figure SA.5).

Specifically, in the host lifestyle study [59], the presence of *rank stability islands* among medium-ranked taxa is an interesting feature revealed by the analysis of rank stability at different time periods in subject A. Interestingly, this stability was compromised when the period was not an ordinary one, suggesting that those taxa were sensitive to changes in lifestyle. Among the genera identified as *rank stability islands*, *Lachnobacterium* and *Clostridium* were catalogued as genera predictive of dysbiosis in the work of Larsen and Dai [253], which analysed the same dataset [59]. Furthermore, research has recently confirmed a clear relationship between *Actinomyces* and conventional adenoma [76], one of the two main precursors of colorectal cancer. Finally, *Eggerthella* is an opportunistic pathogen that is often associated with serious gastrointestinal pathologies [254]. As we will discuss later, we also found *rank stability islands* in the data analysed from the other chapters.

One might question the role of these taxa as key-players in the phase transition of the microbiota and wonder whether they are more susceptible to perturbations than the most abundant taxa. The types of interactions that could sustain this particular behaviour are unclear, as these non-abundant taxa are usually excluded from dynamic studies. Further experiments and data analysis are needed to clarify whether *rank stability islands* are a widespread feature of microbiotas and whether they appear at lower taxonomic levels too.

Notwithstanding the above, we should be aware that the above hypothesis is too simplistic to apply to reality directly. Indeed, the situation is more complicated than the idea that healthy people can be distinguished from non-healthy people in solely compositional terms, as highlighted by Moya and Ferrer in their recent review [71]. There are several feasible scenarios in which we can consider microbiota to be stable, irrespective of its compositional shifts over time. For example, it may depend on the ability of the microbiota to recover its initial composition (resilience), or its ability to recover its original function despite its composition (functional redundancy). What we have shown in this work could be explained as the transition of stable microbiota into a state of dysbiosis.

The results and the model proposed in Chapter I are the first steps towards understanding microbiota stability, although the model presents some limitations and thus further research is required. From a biological perspective, many questions arise from this work. We have observed the same pattern in Taylor's parameters under all the conditions studied, but a pertinent question is whether this is a universal feature in the hugely diverse microbial niches. Furthermore, another relevant question relates to mechanisms involved in maintaining the population structure. Undoubtedly, the nature of the interactions between community components has a significant bearing on this issue, and this is related to community fitness,

as mentioned above. How we should address community fitness remains unclear, but the interaction between microbes, and between host and microbiome will point us in the right direction and help us to unravel the complexity of microbiota and its relationship to host health.

Once established a definition of stability in the first chapter, we wanted to measure its utility and validity in an experiment with a different condition. For this reason, **Chapter III** was focused on the study of microbiomes belonging to children affected by an acute infectious diarrhoea caused by a rotavirus, and their comparison to healthy children. The affected children recovered in time, as we saw in the results, and here we will comment on the biological relevance of such changes. The heterogeneity observed in the abundance profile of every individual in this study is expected because of the inherent diversity that has been proven to exist in the human gut microbiota [166]. Several factors can cause this heterogeneity, such as host genetics [255], microbial assembly [129, 256] and type of birth [39], although the latter is controversial [48]. Nevertheless, the most abundant taxa belonged to *Firmicutes* and *Bacteroidetes*, which are the dominant phyla in the adult microbiota, in accordance with the literature [37, 257]. Specifically, the finding that *Bacteroides* is one of the most abundant genera is also consistent with literature results [258]. One interesting observation that can be made is the higher proportion and variability of the *Others* group in the children with acute diarrhoea. This finding reveals two interesting features: first, these unhealthy microbiotas showed a higher instability than the healthy subset; and second, they had a higher representation of some taxa that were rather a minority in healthy individuals.

The results of the α -diversity analysis showed a transition from a low-diversity to a high-diversity condition in the children with acute diarrhoea. This finding is correlated with the health status recovery of the children in this group because no differences exist between the healthy status at 10 and 30 days. The relationship between low diversity and diseases in the gut microbiota has been extensively demonstrated [166], and this seems to agree with the results of our experiment. Overall, the low α -diversity at diarrhoea stages and its recovery to a normal diversity are in accordance with the findings of previous works [259, 260], but it is important to remark that some of these studies were developed in low-income countries and malnourished infants. Moreover, the β -diversity analyses with the CCA and the ADONIS test revealed that we had different communities at the individual and condition levels. This means that not only were the microbiotas of the unhealthy children less diverse, but they had a different configuration of microbes, compared with those of the healthy children. In agreement with the α -diversity results, these differences seemed to appear only in the first time points of the experiment during the treatment. At days 10 and 30, the statistical differences disappeared, which means that the communities in both groups of children were

similar. Differences found at the individual level are also in agreement with the heterogeneity observed in Figure 6.1.

From the temporal analysis, a fine distinction between both groups of children was found regarding population dynamics. We demonstrated, in Chapter I, that temporal stability has been proved to be related to the health status of the host [231], and, in this study, we observed that almost every child infected with the rotavirus had lower temporal stability than the healthy children. It is important to note that a higher temporal variability is, to a certain extent, caused by the transition from the diseased state to a healthy one. This transition is marked by changes in the microbes that live in the gut, along with their relative abundance. Nevertheless, a high temporal variability during the diarrhoea process is expected because of different reasons, such as the rapid succession of microbes resulting from continuous and rapid faecal movement. The reason why participant C8 remained in the stable part in Figure 6.5 is intriguing, but it is worth noting that C8 is the child that had the fastest recovery compared to the other patients (see Table A.6 in the Appendix). He was the only affected child in which diarrhoea disappeared 24 h later. With this quick recovery, C8 microbiome would have been stabilised earlier than the other affected children. The rank stability plots presented some interesting trends, as well, because of the presence of some groups of bacterial genera that are stable even at a low abundance. The rank stability plots presented some interesting trends, as well, because of the presence of some groups of bacterial genera that are stable even at a low abundance. This result adds support to the *rank stable islands* found in Chapter I. The case of *Actinomyces* is particularly compelling, as it seems to inhabit these low-abundance regimes in a very stable manner [231] in the healthy children. Some species of this genus are observed to be in higher proportions in other cases of diarrhoea [258]. Because of their opportunistic nature, limiting these species to live in low abundance could be beneficial for the host.

The microbial communities of children suffering from diarrhoea seem less mature than those of healthy children [260]. It could be that some of the differences we see might be due to this effect, along with the microbial assembly processes that occur in the first stages of the human life. Suffering diseases as the acute infectious diarrhoea could delay the maturation of the gut microbiota [261]. This high variability in non-healthy children agrees with the other results of this study, as it seems that their microbiotas improve from a perturbed state to a healthy one with time. This is one of the essential points of this experiment, as it proves the power of time series analysis in the microbiota, especially in cases in which a disease is being examined. Besides, the importance of measuring time series in microbiome studies is one of the main general conclusions of this thesis.

As we stated, the microbiota has a different configuration of microbes in both health states, and the differences between them could be of clinical interest. LEfSe analysis showed a clear differentiation concerning the phyla enriched, with a high amount of *Proteobacteria* observed during the first few days in the children suffering from acute diarrhoea. This result agrees with those of extant literature, as an enrichment of bacteria, such as *Escherichia coli* (a *Gammaproteobacteria*), in children suffering diarrhoea was found [259, 262]. On the contrary, *Firmicutes* phylum is usually lowered in children with diarrhoea, so an enrichment of this group was expected in healthy children [40]. Interestingly, we did not see much differences regarding functions enriched. The most intriguing case is the pathway involved in the biosynthesis of ansamycins, a secondary bacterial metabolite that has antimicrobial properties against a broad range of bacteria [263]. Some analogues of ansamycins interact with Hsp90, which helps inhibit the severity of the infection by rotavirus [264]. Although being only an in vitro study, this could be of future interest when the study is expanded to new directions.

In summary, we have shown that microbiota composition seems to correlate with clinical improvement. Children stopped experiencing most of the symptoms three days after probiotic administration, and alongside, microbiota recovered both its diversity and healthy configuration within this period. Therefore, gut microbiota seems to play an essential role in diarrhoea-related processes.

Another system of interest, the oral microbiome, was the subject under study in **Chapter IV**. Here, we wanted to see the possible connections between oxidative stress markers and changes in the volunteer's microbiome. *Streptococcus mutans* used to be associated with caries before the metagenomics era, while recent metagenomics studies have assumed that the oral diseases are of polymicrobial origin [98, 265–267]. Nevertheless, metagenomic studies based on comparing oral microbiome composition of infected and healthy subjects often lead to inconsistent results [231, 249, 268].

We aimed to identify bacteria with pathogenic potential on individual-specific level by correlation analysis with oxidative stress markers that may reflect the inflammatory processes in the oral cavity sampled on daily basis. Salivary markers of oxidative stress were found to be associated with several oral diseases including periodontitis, caries and oral precancerose [194, 269]. However, their usage for diagnostics of oral diseases is limited due to their large temporal variability in healthy subjects [198]. The factors that determine this variability are largely unknown although microbes have been postulated as their modulators several years ago [270]. The association of microbes with oxidative stress can only be tested in interventional experiments including long temporal sampling, because both, the microbes and immune cells, can induce the production of ROS and oxidative stress [194, 271]. In

addition, potential systemic causes for the variability of both, oxidative stress markers in saliva and oral microbiome should not be omitted [272].

In order to assess association of bacterial taxa with oxidative stress markers, we needed to detect temporal variability of the oral microbiome composition on the intra-individual level. In general, the temporal variability of the microbiome detected in our cohort of healthy volunteers (as measured by the Taylor's parameter V) was lower than the variability detected in the perturbed microbiomes previously studied in Chapter I. However, even slight temporal variations in the oral microbiome composition provided important data for intra-individual correlation analysis in this study. The design of the vast majority of microbiome studies does not take into account intra-individual temporal variability, which can be in fact very informative on the health status of the host [231]. Our results are in accordance to the study of Gonze *et al.* [72] which demonstrated that a microbiome may adopt one or another distinct state in the same environmental conditions, meaning that intra-individual temporal variability does not have to correlate necessarily with the changes of the environmental conditions. Interestingly, the intra-individual variability of the microbiome has been identified as the major cause of contradictory results reported by different biomedical studies [249, 268]. As shown in this study, an individual can have an unstable microbiome composition (expressed by the variability parameter V , Figure 7.1), even being considered a healthy subject, thus a realistic picture of subject's microbiome cannot be captured by collecting only one sample per individual. The computational simulation revealed that contradictory correlation analysis results may be obtained, if only one sample per volunteer is collected. For example, if single samples from the 26 volunteers are collected in 1,000 different studies and analysed for correlations between *Rothia*.OTU15862 and TBARS, 22.2% of the studies would report a negative correlation, while 0.1% would report a positive correlation, and 77.7% would report no significant correlation. In comparison, our intra-individual correlation analysis found negative correlations in two volunteers and a positive correlation in one volunteer, while in the remaining volunteers the *Rothia*.OTU15862-TBARS pair resulted in no significant correlations. Another advantage of using correlation analysis on intra-individual level is in the possibility of identifying lagged correlations. The identification of lagged correlations was very important in the present study, as increased production of ROS may not necessarily start at the moment of overgrowth of a pathobiont, but it may be delayed by several days.

The detected correlation patterns between oxidative stress markers, between bacterial species and between markers and bacteria were unique for each volunteer. However, we also found some correlations which were highly consistent on intra-individual level in our cohort, e.g. negative correlation between *Streptococcus*-OTU0 and *Rothia*-OTU1, positive correlation between FRAS and TAC, etc. In addition, the Lotka-Volterra set of equations has

proved the uniqueness of each volunteer interaction matrix. Interaction-based models, such as Lotka-Volterra equations, are more robust to extract biologically relevant interactions in the ecosystem than correlation models [114].

The set of low-abundance species (0.1-5.4% average proportion), which were mostly individual-specific, contributed to the most of the microbiome temporal variability and correlated with the oxidative stress markers more often than the more prevalent species (such as *Rothia* and *Streptotococcus parasanguinis*). For example, *S. mutans* was one of the less prevalent species and it had significant correlations with the oxidative stress markers in the highest number of volunteers in our cohort. However, these correlations were not confirmed in all volunteers. Though, *S. mutans* is capable of producing ROS in vitro [273], its activity may be hampered by other species in the oral microbial community. It is possible that other bacterial species may inhibit *S. mutans* ROS production, particularly in the samples in which *S. mutans* did not correlate with oxidative stress markers. In addition, these volunteers are likely to contain other bacterial species that stimulate ROS production or benefit from ROS production to increase their biological niche in anoxic regions [274, 275]. Nevertheless, *S. mutans* was the bacterial species with one of the most pronounced correlations with the oxidative stress markers which supports its important role in caries pathogenesis.

Correlation patterns of bacterial species with oxidative stress markers differed among volunteers which may be explained by numerous factors. First of all, cells belonging to the same bacterial species distributed in different compartments of the oral cavity differ by their activity - some cells are actively metabolising nutrients, while others are waiting for optimal growth conditions in a dormant stage [276, 277]. However, differential activity of bacterial cells is not taken into account when total DNA from saliva is sequenced which might explain why intra-individual correlation patterns of bacterial OTUs with oxidative stress markers were non-uniform in this study. Laboratory experiments performed with pure cultures containing uniformly growing bacterial cells might not fully mimic the real in vivo ROS production due to the differential activity of bacterial cells in the oral cavity. Furthermore, it was demonstrated that only a portion of bacterial cells belonging to the same bacterial species stimulate human immune reactions [98]. Therefore, only a proportion of each bacterial species is involved in the induction of the ROS production and these proportions are likely inconsistent among different individuals. Although highly active bacterial cells and those which interact with the human immune system could be quantified using flow cytometry [278], it is very difficult to determine the exact number of variables that may influence the actual proportion of bacterial cells stimulating the ROS production in vivo. Having absolute number of cells belonging to each bacterial species, rather than compositional data, would help to address correlation patterns in each volunteer with more details.

In our study, many bacterial OTUs correlated either positively or negatively with some of the measured oxidative stress markers in some of the volunteers, which suggests that each volunteer has different homeostatic mechanisms. Therefore, there is no universal answer to which specific bacterial species are associated to the ROS production in the oral cavity in all humans. Integrating microbiome composition data with proteomics and metabolomics may help in determining which bacteria are associated to the ROS production in the oral cavity, as different bacterial species are likely to be responsible for the same metabolic function in different individuals [71, 279]. In such studies, information on microbiome temporal variability would be also of very high importance. In addition, future experimental studies should focus on the origin and the consequences of oxidative stress in the oral cavity in relation to the microbiome composition and its modulation or transplantation.

Finally, in this thesis we wanted to explore the importance of microbial interactions, and, in **Chapter II**, we proposed a new methodology based on noise-induced experiments to uncover bacterial interactions. Any biological system has uncertainties that often have been put aside from models for simplicity [280]. However, we can take advantage of uncertainties in microbial systems to propose new approaches for the study of species interactions. As discussed in Chapter I, the temporal variability of a microbiome can be modelled as Taylor's law. Thus, we modified the generalised Lotka-Volterra model by adding a stochastic component in which the noise amplitude is based on Taylor's law (see sgLV equation 5.1). With this model, we can specifically control the behaviour of any species in our community, introducing perturbations that can be of help to infer the interactions [79, 114].

Studies with an emphasis in theory that study synthetic populations are being more prominent in the microbiome field [72, 281], highlighting the importance of theory in biology [282]. For many years, several authors made impressive attempts to understand microbial interactions in different environments, evaluating them in different conditions by broad perturbations as antibiotic administration [79, 114, 121]. However, here we attempted to apply perturbations in a precise manner, namely, to modify specific types of microorganisms instead of modifying a wide range of them. We think that the technology to perform this kind of experimentation designs already exists, as we find high-specificity methods like the use of phages along with the CRISPR-Cas system [283, 284], a technique that has been proven valid as antimicrobial against some virulent members of the *Staphylococcus* genus [285]. We proposed two different types of experiments that introduces specific perturbations to precise community members, one that modifies the noise level of the target species, and another that reduces the abundance of the target at an specific time point.

We explored this methodology applied to two interesting species in our community: sp3, the most abundant species; and sp14, highly abundant and one of the most connected members in our community. We calculated Pearson's correlation coefficients of the generated time series because they are among the most used statistical tools to explore microbial communities. Some general behaviour can be extracted from these matrices, as the modification of sp14 showed marked effects on the community, most probably for its high connectivity. However, although we did see some changes in the community behaviour, correlation matrices alone are not sufficient to make any assumption of the real interactions that act in our community, as has been demonstrated [114]. For this purpose, we solved a least squares regression problem to infer the interaction matrix, under no assumption of any specific model behind the dynamic of our community. The results show that we increase our ability to uncover some of the original interactions through the noise-induced experiments, but only for the most abundant species. Interestingly, we can observe that applying noise to sp3 and sp14 'cleaned' the spurious interactions that appeared in the experiments with default parameters. This result highlights the utility of this kind of approach to study microbial interactions. On the contrary, the pulse did not seem to be of much help to explore these interactions, so we should think about other possible ways to deal with this method in order to improve it. These are still preliminary results, and more work has to be done in building better regression models, and modifying other parameters of our equations as the growth rates or the interaction parameters in time.

What we present here is both a tool and an idea to further comprehend the complexity of this subject. With the creation of an *in-silico* population governed by a simple model, we have been able to demonstrate that noise-induced experiments are of critical importance in the study of the relationship between the microorganisms in microbiomes. Like in particle physics, where one need to collide different particles between them to understand their fundamental properties, we propose that experimental studies in microbiome will take advantage of the strategy of 'colliding' the microbial species with controlled perturbations. This approach will enable new ways to understand the essential and necessary properties of microbial ecosystems that will end up in direct applications on microbial systems of interest.

In conclusion, through this thesis, it has been shown that temporal series is critical to gain fundamental insight about the complex problem of the interaction between humans and their microbiomes. By the use of general ecological laws as Taylor's law, we have been able to understand the notion of stability in different scenarios, and that Taylor's parameters characterised the statistical behaviour of microbiome changes. We proved this system to be of utility when studying the dynamics of children's microbiomes that were affected by an acute infectious diarrhoea, observing that the unhealthy patients were unstable with respect

to the healthy children. Furthermore, we also characterised the lack of general patterns of relations between the oral microbiota and oxidative stress markers, except in few, marked, examples. Finally, we used this notion of variability based on Taylor's law to implement it as the stochastic part of the generalised Lotka-Volterra model, to further study the microbial behaviour with experiments based on specific perturbations.

General Conclusions

The conclusions of the thesis are:

1. All human microbiomes under study in this thesis follow the Taylor's law. The scaling parameter β is always less than 1, implying that OTUs with the lesser abundance are more vulnerable to perturbations than the most abundant. Furthermore, the variability amplitude parameter V can help us to differentiate between healthy microbiotas and perturbed ones.
2. We propose the Langevin equation as a model to measure the stability of a given microbiome. Its solution describes a phase space where microbiomes can be placed depending on their variability, measured as a Taylor's law; and their fitness, dependent on many ecological factors. This phase space is divided into two different regimes, one *unstable* where microbiomes are susceptible to perturbations, and another *stable* where microbiomes show robustness against system perturbations.
3. We observed that microbiomes transitions from the stable phase to the unstable when the system is perturbed; microbiomes that can also travel to the stable phase if the subject improves its health status; and other microbiomes that show a quasi-periodical variability within the stable phase due to their high fitness.
4. Rank Stability matrices are powerful tools for stability insight. We have observed the existence of *rank stability islands* of low abundant OTUs that remain quite stable within the microbial community in all the experimental data from this thesis. The existence of such stable groups could be of interest to evaluate the stability of a microbiome.
5. Taylor's law can be used as the noise amplitude, in the stochastic parameter, in the generalised Lotka-Volterra model to mimic the behaviour of realistic microbial systems.

6. With noise-induced experiments, we can specifically modify the noise amplitude of certain species in our community to analyse its interactions. Furthermore, we can also apply a pulse that reduces the abundance of that species, as would happen in reality with the administration of antibiotics.
7. By using these types of experiments, we can see significant changes in the other species' behaviour that can be analysed to infer the interaction matrix.
8. The solution of the optimisation of a regression based on least-squares problem with a Lagrange multiplier show promising preliminary results, highlighting the reduction of the inference error of the interactions in the species modified by our experiments.
9. The gut microbiome composition from Turkish children was dominated by *Firmicutes* and *Bacteroidetes*, with a high representation of *Bacteroides* and *Faecalibacterium*. However, the variability of the less abundant OTUs was higher in children with acute infectious diarrhoea.
10. We observed that the bacterial diversity between healthy children, and those with the acute infectious diarrhoea, recovered in time. The α -diversity was lower in children infected by the rotavirus in the first 5 days, but it recovered to a higher diversity when they were clinically healthy. The β -diversity showed differences between both groups regarding their compositions also in the first 5 days, but the differences disappeared when the children stopped having symptoms.
11. The temporal analysis showed a clear differentiation between both health groups, being the microbiomes of children suffering from the acute infectious diarrhoea more variable than the healthy ones. Patient C8 was the only case of the non-healthy children that was comparable to the healthy group, but it was linked to be the patient with the fastest recovery of the whole group. The Rank Stability matrices reveal the presence of *rank stability islands* in healthy children.
12. We observed an enrichment of bacteria belonging to *Proteobacteria* in unhealthy children, and an enrichment of *Firmicutes* group in healthy children. In the functional analysis, among the metabolic pathways inferred that were enriched, the biosynthesis of ansamycins could be of clinical interest in the future for its relationship with rotavirus infection resistance.
13. The oral microbiomes studied in Chapter IV showed a some variability even belonging to healthy subjects. Nevertheless, they were less variable than the non-healthy microbiomes studied in Chapter I and Chapter III.

14. With the simulation of the one-sample-one-volunteer analysis, we highlighted the importance of measuring time series in microbiome studies. Taking only one sample per voluntary and computing the correlations between all OTU-marker pairs could yield false positives or false negatives, an issue that time series can help to solve.
15. There was not a general pattern in the correlation analyses, neither between oxidative stress markers nor between OTUs and markers. The only exceptions worth mentioning were the correlation between FRAS and TAC, and between *Streptococcus mutans* and several markers. This result points to a polymicrobial theory of oral diseases, and for multi-stability within microbial communities. The set of low-abundance species, which were mostly individual-specific, correlated with the oxidative stress markers more often than the more abundant species, as it is the case of *Streptococcus mutans*.

Resumen en castellano

Estabilidad del microbioma humano: salud y enfermedad relacionados con la variabilidad temporal de la microbiota humana.

El éxito evolutivo de los microorganismos ha hecho que éstos estén íntimamente ligados a organismos más complejos. El microbioma humano es la colección de toda la vida microscópica que habita con nosotros en las diferentes partes del cuerpo, desde el intestino, pasando por la cavidad oral, hasta incluso toda la superficie de nuestra piel. Esta inmensa comunidad, por tanto, tiene efectos directos sobre la homeostasis del hospedador, participando en las condiciones de salud y enfermedad debidas tanto a desajustes de la propia microbiota como a otros factores externos. Al ser un sistema tan dinámico, la microbiota puede variar en gran medida con el tiempo, siendo esta variabilidad un aspecto que puede ser de utilidad para comprobar el estado de salud o enfermedad de un hospedador. Sin embargo, hasta la fecha no se tiene una definición concreta y útil de estabilidad en términos de microbiota.

El objetivo central de esta tesis ha sido establecer una definición formal de qué significa tener una microbiota estable, y cuál es su relación con el estado de salud y enfermedad del hospedador. Para ello, la tesis se ha dividido en cuatro objetivos en los que se ha intentado dilucidar esta relación entre variabilidad de la microbiota y salud.

El *primer estudio* consistió en proponer un modelo matemático para cuantificar el nivel de estabilidad de la microbiota. Para ello, se tomaron datos de otros estudios ya publicados que consistieran en series temporales de microbiota, en varias condiciones diferentes, y con un mínimo de tres tiempos. Entre ellas utilizamos: datos de gente sana; pacientes con síndrome de colon irritable; pacientes con diferentes grados de obesidad; parejas de niños gemelos discordantes para la enfermedad de Kwashiorkor; voluntarios que siguieron dietas basadas en vegetales o en carnes; pacientes con toma de antibióticos; y una serie temporal para ver la influencia del estilo de vida de dos voluntarios. Todos los datos estaban en forma

de secuencias del gen ribosómico del 16S, salvo el de los niños gemelos discordantes, cuyos datos estaban en forma de SMS (shotgun metagenomic sequencing).

Para ver la variabilidad temporal, propusimos el uso de la ecuación diferencial de Langevin que tiene dos parámetros bien diferenciados: uno determinista, en nuestro caso correspondiente con el 'fitness' de la microbiota; y otro estocástico directamente relacionado con la cantidad de variación temporal de una microbiota. Este segundo componente lo modelamos siguiendo una ley ecológica presente en diversos sistemas naturales, la ley de Taylor. Esta ley relaciona la variabilidad de una población con su media, mediante una ley de potencias (eq. 10.1). Así pues, comprobamos que todas las microbiotas medidas, en total 99 series temporales, seguían esta ley. Además, vimos que el parámetro V de la ley estaba directamente relacionado con la magnitud de la variabilidad, y, por lo tanto, estaba directamente relacionada con el estado de salud o enfermedad del hospedador. Al resolver la ecuación de Langevin, obtenemos un diagrama de fases con los ejes del fitness y de la variabilidad, con una fase estable y otra fase inestable. Al modelar el fitness de algunos ejemplos, y medir su variabilidad en el tiempo, comprobamos que las microbiotas podían realizar una transición desde una fase a otra dependiendo de su estado de salud, y la variabilidad. Por lo tanto, concluimos que los cambios en la microbiota son un reflejo del estado de salud del hospedador, y que, con la debida modelización, puede ser un elemento importante para su uso en diversos ámbitos más aplicados de la ciencia.

$$\sigma = V \cdot x_i^\beta \quad (10.1)$$

El *segundo estudio* tenía como objetivo el estudio temporal de la microbiota en niños que padecían una diarrea aguda infecciosa en relación a niños sanos. Esta diarrea, causada por rotavirus, es la causa principal de mortandad y mortalidad en niños, con una alta incidencia en países del tercer mundo o en vías de desarrollo. Por lo tanto, estudiar cómo se comportaba la microbiota durante el transcurso de la enfermedad es importante para añadir conocimiento a posibles tratamientos que maximicen la eficiencia de la recuperación de la comunidad microbiana. Para este estudio, fruto de una colaboración con hospitales y médicos turcos, se enrolaron 16 niños de entre 3 y 4 años de edad, de los cuales 6 estaban sanos, y 10 sufrían la diarrea infecciosa. Se tomaron muestras fecales de los enfermos a la entrada del hospital (día 0), y se volvieron muestras en el día 3, día 5, día 10 y día 30. En los niños sanos también se muestreó de la misma forma. En total, tenemos 5 puntos temporales por cada uno de los 16 niños. De esas muestras fecales, se secuenciaron los fragmentos correspondientes al gen ribosómico del 16S, y se comprobó a qué géneros correspondían.

Al analizar las diversidades de las microbiotas, hemos podido comprobar que la diversidad de tipo alfa es inferior en los niños enfermos respecto a los niños sanos durante los tres

primeros puntos temporales, mientras que en los días 10 y 30 no hay diferencias en diversidad. Además, la diversidad de tipo beta nos da a entender que las comunidades microbianas de los días 0, 3 y 5 son diferentes entre niños sanos y niños enfermos, no encontrando estas diferencias en los días 10 y 30. Además, al analizar los taxones que estaban diferencialmente representados en ambos estados, hemos podido comprobar que, durante los primeros cinco días encontramos diferencias marcadas como la abundancia del filo *Proteobacteria* en los niños enfermos, y abundancia del filo *Firmicutes* en los niños sanos. En los días 10 y 30, sin embargo, estas diferencias desaparecieron salvo la abundancia del género *Blautia* en niños sanos. Todo esto concuerda con el análisis de la variabilidad temporal, en la que vemos que todos los niños enfermos están a varias sigmas de diferencia de los niños sanos en cuestión de variabilidad temporal, salvo el caso especial de un niño que dejó de tener síntomas el primer día de estar hospitalizado. Por último, hemos visto que hay algunas funciones enriquecidas tales como la biosíntesis de ansamicinas o la ruta de las pentosas fosfato.

Es importante decir que el estado de salud de los niños mejoró de tal forma que dejaron de tener diarrea a las 72 h de haber sido hospitalizados, algunos incluso antes. Este dato concuerda con la mejoría en las diversidades estudiadas, ya que, aunque hay diferencias al principio del periodo de muestreo, deja de haberlas al final. Estos resultados nos indican que la microbiota de los niños con la diarrea aguda infecciosa estaba afectada por la enfermedad, y que, de alguna forma, mejoraba con el tiempo. Este hecho se ve reforzado por la transición de un estado de enfermedad a uno de salud visto por los resultados de la variabilidad temporal. Por último, parece ser que algunas de las funciones enriquecidas en este experimento podrían tener relación con los mecanismos bioquímicos del transcurso de la infección por rotavirus.

El *tercer estudio* tuvo como objetivo el estudio temporal de la microbiota oral de 26 voluntarios, donde a la vez se midieron cinco marcadores de estrés oxidativo habituales en la búsqueda de enfermedades bucales como las caries. Hoy en día se sigue pensando que especies como *Streptococcus mutans* son responsables de las caries en personas, por lo que un estudio de la microbiota en el tiempo, junto con marcadores de estrés oxidativo, podría arrojar luz sobre la causalidad de este y otros problemas. Para ello, contamos con la ayuda de 26 voluntarios, 13 hombres y 13 mujeres, de entre 20 y 30 años de edad, de los que se midieron una media de 21 tiempos por persona. Una vez más, secuenciamos el gen ribosómico del 16S y obtuvimos las secuencias que luego usamos para establecer su taxonomía. También se midieron cinco marcadores de estrés oxidativo.

Con todos los datos temporales recogidos, medimos la variabilidad temporal de todos los sujetos, obteniendo un rango de valores de variabilidad que, a pesar de ser ciertamente amplio, estaba por debajo de los valores de V de individuos enfermos de otros estudios. A pesar de su variabilidad, los microbiomas de los voluntarios compartían la abundancia de

ciertos OTUs muy abundantes como *Streptococcus paransanguinis* o bacterias pertenecientes al género *Rothia*. Sin embargo, los menos abundantes fluctuaban más que los más abundantes, algo que también concuerda con los resultados de variabilidad temporal. Los análisis de correlaciones, tanto clásicas como con la dimensión temporal añadida al análisis, no dieron un patrón universal en la relación marcador-OTU, ni tampoco entre los diferentes OTUs de las muestras. Asimismo, en un análisis de las posibles interacciones por medio del modelo generalizado de Lotka-Volterra, las matrices obtenidas tampoco apoyaban la existencia de una matriz universal de interacciones. Todos estos datos apoyan la hipótesis de que las enfermedades bucodentales puedan estar causadas por varios microorganismos al mismo tiempo, y que sea la disbiosis causada por ellos la causante de ellas. Además, al haber obtenido varias posibilidades de correlación e interacciones en los voluntarios, también se apoya la hipótesis de la existencia de multi-estabilidad en microbiota, ya que parece ser que hay varios estados estables diferentes. Por último, también se hizo un estudio para demostrar la utilidad de las series temporales. Se escogió una sola muestra aleatoria por cada voluntario en 1000 iteraciones, y en cada una de ellas se midieron las correlaciones entre los 50 OTUs más abundantes y los marcadores. Los resultados nos permiten ver que, como máximo, solo conseguimos un 30% de iteraciones en las que la correlación (tanto positiva como negativa) existía, dando lugar a falsos positivos y negativos.

Finalmente, el *cuarto estudio* tuvo como objetivo estudiar la posibilidad de mejorar nuestras predicciones de las interacciones en una microbiota gracias a la perturbación específica de miembros de interés de la comunidad. Para ello partimos de la base de usar el modelo generalizado de Lotka-Volterra, ya que su aplicación es sencilla y bien entendida, al que implementamos un elemento estocástico basado en la ley de Taylor (eq. 10.2). Con ello conseguimos imitar la variabilidad que hemos observado en el *primer estudio*, existente en todas las microbiotas estudiadas a lo largo de esta tesis.

$$\frac{d}{dt}x_i(t) = \mu_i x_i(t) + x_i(t) \sum_{j=1}^L M_{ij} x_j(t) + Bx_i^b \cdot dW \quad (10.2)$$

Hemos propuesto el uso de dos experimentos diferentes, fácilmente estudiables en nuestro modelo, en los cuales se modifica de forma fina a miembros específicos de nuestra comunidad. El primer experimento consiste en aumentar la amplitud del ruido de esa especie durante un tiempo definido, mientras que el segundo experimento consistiría en disminuir bruscamente la abundancia de esa determinada especie en un tiempo específico. Al estudiar las correlaciones (con coeficientes de Pearson) de los diferentes escenarios, hemos comprobado que, aunque no se asemejan a la matriz de interacción original, el comportamiento global de todos los miembros se ve diferencialmente modificado según la especie modificada o el experimento

aplicado. El siguiente paso fue la implementación de un sencillo modelo de regresión, el cual se resolvió como un problema de optimización de mínimos cuadrados para inferir la matriz de interacciones. Se ha comprobado que, aunque perturbar determinadas especies no mejora en gran medida la inferencia de la matriz total, sí que hemos observado que la inferencia de las relaciones específicas de la especie modificada en los experimentos sí que mejora conforme aumentamos el ruido.

Como conclusiones finales, podemos decir que todas las microbiotas analizadas a lo largo de esta tesis siguen la denominada Ley de Taylor, y que los parámetros de esta ley nos permiten inferir el estado de salud de un hospedador. Asimismo, el uso de series temporales en microbiota es esencial si queremos conseguir una profundidad analítica que solo da el estudio de la dinámica de los sistemas. El uso de modelos matemáticos para abordar este estudio, y modelizar el comportamiento de las microbiotas, será esencial para desenredar la complejidad de este problema.

References

- [1] Takayuki Tashiro, Akizumi Ishida, Masako Hori, Motoko Igisu, Mizuho Koike, Pauline Méjean, Naoto Takahata, Yuji Sano, and Tsuyoshi Komiya. Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature*, 549(7673):516–518, 2017.
- [2] Eric Smith and Harold J Morowitz. *The origin and nature of life on Earth: the emergence of the fourth geosphere*. Cambridge University Press, 2016.
- [3] John D. Sutherland. Opinion: Studies on the origin of life — the end of the beginning. *Nature Reviews Chemistry*, 1:0012, 2017.
- [4] John D. Sutherland. The Origin of Life-Out of the Blue. *Angewandte Chemie International Edition*, 55(1):104–121, jan 2016.
- [5] J. Baz Jackson. Natural pH gradients in hydrothermal alkali vents were unlikely to have played a role in the origin of life. *Journal of Molecular Evolution*, 83(1-2):1–11, aug 2016.
- [6] José Aguilar-Rodríguez, Joshua L. Payne, and Andreas Wagner. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology & Evolution*, 1(2):0045, jan 2017.
- [7] R. Gil, F. J. Silva, J. Pereto, and A. Moya. Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews*, 68(3):518–537, sep 2004.
- [8] J. I. Glass, N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, H. O. Smith, and J. C. Venter. Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences*, 103(2):425–430, jan 2006.
- [9] C. A. Hutchison, R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, and J. C. Venter. Design and synthesis of a minimal bacterial genome. *Science*, 351(6280):aad6253–aad6253, mar 2016.
- [10] Esteban Martínez-García and Víctor de Lorenzo. The quest for the minimal bacterial genome. *Current Opinion in Biotechnology*, 42:216–224, dec 2016.

- [11] Andrés Moya, Juli Peretó, Rosario Gil, and Amparo Latorre. Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nature Reviews Genetics*, 9(3):218–229, mar 2008.
- [12] C. R. Woese and G. E. Fox. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11):5088–5090, nov 1977.
- [13] Herbert Copeland. *The classification of lower organisms*. Palo Alto, Calif., Pacific Books, 1956.
- [14] R. H. Whittaker. On the broad classification of organisms. *The Quarterly Review of Biology*, 34(3):210–226, sep 1959.
- [15] L Margulis and M J Chapman. *Kingdoms & Domains an Illustrated Guide to the Phyla of Life on Earth Preface*. Elsevier, 2009.
- [16] Cindy J. Castelle and Jillian F. Banfield. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell*, 172(6):1181–1197, 2018.
- [17] Eugene V Koonin. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biology*, 11(5):209, 2010.
- [18] John M. Archibald. Endosymbiosis and eukaryotic cell evolution. *Current Biology*, 25(19):R911–R921, oct 2015.
- [19] Joran Martijn, Julian Vosseberg, Lionel Guy, Pierre Offre, and Thijs J. G. Ettema. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*, 557(7703):101–105, 2018.
- [20] Katarzyna Zaremba-Niedzwiedzka, Eva F. Caceres, Jimmy H. Saw, DIsa Bäckström, Lina Juzokaite, Emmelien Vancaester, Kiley W. Seitz, Karthik Anantharaman, Piotr Starnawski, Kasper U. Kjeldsen, Matthew B. Stott, Takuro Nunoura, Jillian F. Banfield, Andreas Schramm, Brett J. Baker, Anja Spang, and Thijs J.G. Ettema. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541(7637):353–358, 2017.
- [21] G. I. McFadden. Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 6(4):a016105–a016105, apr 2014.
- [22] C. W. Morden and A. R. Sherwood. Continued evolutionary surprises among dinoflagellates. *Proceedings of the National Academy of Sciences*, 99(18):11558–11560, aug 2002.
- [23] M. J. Gosalbes, A. Lamelas, A. Moya, and A. Latorre. The striking case of tryptophan provision in the cedar aphid *cinara cedri*. *Journal of Bacteriology*, 190(17):6026–6029, jun 2008.

- [24] R. C. H. J. van Ham, J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernandez, L. Jimenez, M. Postigo, F. J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Moran, and A. Moya. Reductive genome evolution in *buchnera aphidicola*. *Proceedings of the National Academy of Sciences*, 100(2):581–586, jan 2003.
- [25] Rafael Patiño-Navarrete, Andrés Moya, Amparo Latorre, and Juli Peretó. Comparative genomics of *blattabacterium cuenoti*: The frozen legacy of an ancient endosymbiont genome. *Genome Biology and Evolution*, 5(2):351–361, jan 2013.
- [26] T. Cavalier-Smith, M. Brasier, and T. M. Embley. Introduction: how and when did microbes change the world? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1470):845–850, 2006.
- [27] Michael Gross. How life shaped earth. *Current Biology*, 25(19):R847–R850, oct 2015.
- [28] S. L. Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529, may 1953.
- [29] Timothy W. Lyons, Christopher T. Reinhard, and Noah J. Planavsky. The rise of oxygen in Earth’s early ocean and atmosphere. *Nature*, 506(7488):307–315, 2014.
- [30] Martin F. Hohmann-Marriott and Robert E. Blankenship. Evolution of Photosynthesis. *Annual Review of Plant Biology*, 62(1):515–548, jun 2011.
- [31] Rochelle M. Soo, James Hemp, Donovan H. Parks, Woodward W. Fischer, and Philip Hugenholtz. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, 355(6332):1436–1440, 2017.
- [32] Xi Chen, Hong-Fei Ling, Derek Vance, Graham A. Shields-Zhou, Maoyan Zhu, Simon W. Poulton, Lawrence M. Och, Shao-Yong Jiang, Da Li, Lorenzo Cremonese, and Corey Archer. Rise to modern levels of ocean oxygenation coincided with the cambrian radiation of animals. *Nature Communications*, 6(1), may 2015.
- [33] Jorge Valdés, Inti Pedroso, Raquel Quatrini, Robert J Dodson, Herve Tettelin, Robert Blake, Jonathan A Eisen, and David S Holmes. *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics*, 9(1):597, 2008.
- [34] C. Schleper, G. Puhler, H.P. Klenk, and W. Zillig. *Picrophilus oshimae* and *picrophilus torridus* fam. nov., gen. nov., sp. nov., two species of hyperacidophilic, thermophilic, heterotrophic, aerobic archaea. *International Journal of Systematic Bacteriology*, 46(3):814–816, jul 1996.
- [35] Elena V. Pikuta, Richard B. Hoover, and Jane Tang. Microbial extremophiles at the limits of life. *Critical Reviews in Microbiology*, 33(3):183–209, 2007.
- [36] Nathanael Lampe, Vincent Breton, David Sarramia, Télesphore Sime-Ngando, and David G. Biron. Understanding low radiation background biology through controlled evolution experiments. *Evolutionary Applications*, 10(7):658–666, jun 2017.

- [37] Peter J. Turnbaugh, Micah Hamady, Tanya Yatsunenko, Brandi L. Cantarel, Alexis Duncan, Ruth E. Ley, Mitchell L. Sogin, William J. Jones, Bruce A. Roe, Jason P. Affourtit, Michael Egholm, Bernard Henrissat, Andrew C. Heath, Rob Knight, and Jeffrey I. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, 2009.
- [38] Daniel A. Leffler and J. Thomas Lamont. Clostridium difficile infection. *New England Journal of Medicine*, 372(16):1539–1548, apr 2015.
- [39] Juan Miguel Rodríguez, Kiera Murphy, Catherine Stanton, R Paul Ross, Olivia I Kober, Nathalie Juge, Ekaterina Avershina, Knut Rudi, Arjan Narbad, Maria C Jenmalm, Julian R Marchesi, and Maria Carmen Collado. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health & Disease*, 26:26050, feb 2015.
- [40] Shirajum Monira, Shota Nakamura, Kazuyoshi Gotoh, Kaori Izutsu, Haruo Watanabe, Nur Haque Alam, Takaaki Nakaya, Toshihiro Horii, Sk Imran Ali, Tetsuya Iida, and Munirul Alam. Metagenomic profile of gut microbiota in children during cholera and recovery. *Gut Pathogens*, 5(1):1, 2013.
- [41] F. Sanger, S. Nicklen, and A. R. Coulson. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, dec 1977.
- [42] Jason A. Reuter, Damek V. Spacek, and Michael P. Snyder. High-throughput sequencing technologies. *Molecular Cell*, 58(4):586–597, may 2015.
- [43] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–14, jun 2012.
- [44] Jason Lloyd-Price, Anup Mahurkar, Gholamali Rahnavard, Jonathan Crabtree, Joshua Orvis, A. Brantley Hall, Arthur Brady, Heather H. Creasy, Carrie McCracken, Michelle G. Giglio, Daniel McDonald, Eric A. Franzosa, Rob Knight, Owen White, and Curtis Huttenhower. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, 550(7674):61–66, sep 2017.
- [45] Thomas S.B. Schmidt, Jeroen Raes, and Peer Bork. The Human Gut Microbiome: From Association to Modulation. *Cell*, 172(6):1198–1215, 2018.
- [46] Alexandra Zhernakova, Alexander Kurilshikov, Marc Jan Bonder, Etti F. Tigchelaar, Melanie Schirmer, Tommi Vatanen, Zlatan Mujagic, Arnau Vich Vila, Gwen Falony, Sara Vieira-Silva, Jun Wang, Floris Imhann, Eelke Brandsma, Soesma A. Jankipersadsing, Marie Joossens, Maria Carmen Cenit, Patrick Deelen, Morris A. Swertz, Rinse K. Weersma, Edith J. M. Feskens, Mihai G. Netea, Dirk Gevers, Daisy Jonkers, Lude Franke, Yurii S. Aulchenko, Curtis Huttenhower, Jeroen Raes, Marten H. Hofker, Ramnik J. Xavier, Cisca Wijmenga, and Jingyuan Fu. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, 352(6285):565–569, apr 2016.

- [47] J. Gregory Caporaso, Christian L. Lauber, Elizabeth K. Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, Pawel Gajer, Jacques Ravel, Noah Fierer, Jeffrey I. Gordon, and Rob Knight. Moving pictures of the human microbiome. *Genome Biology*, 12:R50, 2011.
- [48] Gwen Falony, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, Marc Jan Bonder, Mireia Valles-Colomer, Doris Vandeputte, Raul Y. Tito, Samuel Chaffron, Leen Rymenans, Chloë Verspecht, Lise De Sutter, Gipsi Lima-Mendez, Kevin D'hoë, Karl Jonckheere, Daniel Homola, Roberto Garcia, Etti F. Tigchelaar, Linda Eeckhautd, Jingyuan Fu, Liesbet Henckaerts, Alexandra Zhernakova, Cisca Wijmenga, and Jeroen Raes. Population-level analysis of gut microbiome variation. *Science*, 352(6285):560–564, apr 2016.
- [49] Duy Tin Truong, Adrian Tett, Edoardo Pasolli, Curtis Huttenhower, and Nicola Segata. Microbial strain-level population structure & genetic diversity from metagenomes. *Genome Research*, 27(4):626–638, 2017.
- [50] Manimozhian Arumugam, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R. Mende, Gabriel R. Fernandes, Julien Tap, Thomas Bruls, Jean-Michel Batto, Marcelo Bertalan, Natalia Borruel, Francesc Casellas, Leyden Fernandez, Laurent Gautier, Torben Hansen, Masahira Hattori, Tetsuya Hayashi, Michiel Kleerebezem, Ken Kurokawa, Marion Leclerc, Florence Levenez, Chaysavanh Manichanh, H. Bjørn Nielsen, Trine Nielsen, Nicolas Pons, Julie Poulain, Junjie Qin, Thomas Sicheritz-Ponten, Sebastian Tims, David Torrents, Edgardo Ugarte, Erwin G. Zoetendal, Jun Wang, Francisco Guarner, Oluf Pedersen, Willem M. de Vos, Søren Brunak, Joel Doré, María Antolín, François Artiguenave, Hervé M. Blottiere, Mathieu Almeida, Christian Brechot, Carlos Cara, Christian Chervaux, Antonella Cultrone, Christine Delorme, Gérard Denariáz, Rozenn Dervyn, Konrad U. Foerstner, Carsten Friss, Maarten van de Guchte, Eric Guedon, Florence Haimet, Wolfgang Huber, Johan van Hylckama-Vlieg, Alexandre Jamet, Catherine Juste, Ghalia Kaci, Jan Knol, Omar Lakhdari, Severine Layec, Karine Le Roux, Emmanuelle Maguin, Alexandre Mérieux, Raquel Melo Minardi, Christine M'rini, Jean Muller, Raish Oozeer, Julian Parkhill, Pierre Renault, Maria Rescigno, Nicolas Sanchez, Shinichi Sunagawa, Antonio Torrejon, Keith Turner, Gaetana Vandemeulebrouck, Encarna Varela, Yohanan Winogradsky, Georg Zeller, Jean Weissenbach, S. Dusko Ehrlich, and Peer Bork. Enterotypes of the human gut microbiome. *Nature*, 473(7346):174–180, apr 2011.
- [51] Allyson L. Byrd, Yasmine Belkaid, and Julia A. Segre. The human skin microbiome. *Nature Reviews Microbiology*, 16(3):143–155, 2018.
- [52] E. K. Costello, C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight. Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694–1697, nov 2009.
- [53] Julia Oh, Allyson L. Byrd, Clay Deming, Sean Conlan, Heidi H. Kong, and Julia A. Segre. Biogeography and individuality shape function in the human skin metagenome. *Nature*, 514(7520):59–64, oct 2014.
- [54] M. Kilian, I. L. C. Chapple, M. Hannig, P. D. Marsh, V. Meuric, A. M. L. Pedersen, M. S. Tonetti, W. G. Wade, and E. Zaura. The oral microbiome – an update for oral healthcare professionals. *Bdj*, 221(10):657–666, 2016.

- [55] Daniel Belstrøm, Palle Holmstrup, Allan Bardow, Alexis Kokaras, Nils-Erik Fiehn, and Bruce J. Paster. Temporal stability of the salivary microbiota in oral health. *PLOS ONE*, 11(1):e0147472, jan 2016.
- [56] Simon J. S. Cameron, Sharon A. Huws, Matthew J. Hegarty, Daniel P. M. Smith, and Luis A. J. Mur. The human salivary microbiome exhibits temporal stability in bacterial diversity. *FEMS Microbiology Ecology*, 91(9):fiv091, jul 2015.
- [57] Pedro Belda-Ferre, Luis David Alcaraz, Ral Cabrera-Rubio, Héctor Romero, Aurea Simón-Soro, Miguel Pignatelli, and Alex Mira. The oral metagenome in health and disease. *ISME Journal*, 6(1):46–56, 2012.
- [58] Ron Sender, Shai Fuchs, and Ron Milo. Revised estimates for the number of human and bacteria cells in the body. *PLOS Biology*, 14(8):e1002533, aug 2016.
- [59] Lawrence A David, Arne C Materna, Jonathan Friedman, Maria I Campos-Baptista, Matthew C Blackburn, Allison Perrotta, Susan E Erdman, and Eric J Alm. Host lifestyle affects human microbiota on daily timescales. *Genome biology*, 15(7):R89, 2014.
- [60] Vinod K. Gupta, Sandip Paul, and Chitra Dutta. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Frontiers in Microbiology*, 8(JUN), 2017.
- [61] Tanya Yatsunenko, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N. Baldassano, Andrey P. Anokhin, Andrew C. Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J. Gregory Caporaso, Catherine A. Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I. Gordon. Human gut microbiome viewed across age and geography. *Nature*, 3(Ivic), 2012.
- [62] Matthew R. Mason, Haikady N. Nagaraja, Terry Camerlengo, Vinayak Joshi, and Purnima S. Kumar. Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. *PLoS ONE*, 8(10):e77287, oct 2013.
- [63] Alexandra J. Obregon-Tito, Raul Y. Tito, Jessica Metcalf, Krithivasan Sankaranarayanan, Jose C. Clemente, Luke K. Ursell, Zhenjiang Zech Xu, Will Van Treuren, Rob Knight, Patrick M. Gaffney, Paul Spicer, Paul Lawson, Luis Marin-Reyes, Omar Trujillo-Villaruel, Morris Foster, Emilio Guija-Poma, Luzmila Troncoso-Corzo, Christina Warinner, Andrew T. Ozga, and Cecil M. Lewis. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Communications*, 6(1), mar 2015.
- [64] Stephanie L. Schnorr, Marco Candela, Simone Rampelli, Manuela Centanni, Clarissa Consolandi, Giulia Basaglia, Silvia Turrone, Elena Biagi, Clelia Peano, Marco Severgnini, Jessica Fiori, Roberto Gotti, Gianluca De Bellis, Donata Luiselli, Patrizia Brigidi, Audax Mabulla, Frank Marlowe, Amanda G. Henry, and Alyssa N. Crittenden. Gut microbiome of the hadza hunter-gatherers. *Nature Communications*, 5, apr 2014.

- [65] Elise R. Morton, Joshua Lynch, Alain Froment, Sophie Lafosse, Evelyne Heyer, Molly Przeworski, Ran Blekhman, and Laure Ségurel. Variation in rural african gut microbiota is strongly correlated with colonization by entamoeba and subsistence. *PLOS Genetics*, 11(11):e1005658, nov 2015.
- [66] Toshitaka Odamaki, Kumiko Kato, Hirosuke Sugahara, Nanami Hashikura, Sachiko Takahashi, Jin Zhong Xiao, Fumiaki Abe, and Ro Osawa. Age-related changes in gut microbiota composition from newborn to centenarian: A cross-sectional study. *BMC Microbiology*, 16(1):1–12, 2016.
- [67] Masaru Tanaka and Jiro Nakayama. Development of the gut microbiota in infancy and its impact on health in later life. *Allergology International*, 66(4):515–522, 2017.
- [68] Angela Moya-Pérez, Pauline Luczynski, Ingrid B. Renes, Shugui Wang, Yuliya Borre, C. Anthony Ryan, Jan Knol, Catherine Stanton, Timothy G. Dinan, and John F. Cryan. Intervention strategies for cesarean section–induced alterations in the microbiota-gut-brain axis. *Nutrition Reviews*, 75(4):225–240, apr 2017.
- [69] A. Bergstrom, T. H. Skov, M. I. Bahl, H. M. Roager, L. B. Christensen, K. T. Ejlerskov, C. Molgaard, K. F. Michaelsen, and T. R. Licht. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of danish infants. *Applied and Environmental Microbiology*, 80(9):2889–2900, feb 2014.
- [70] Quang N. Nguyen, Jonathon E. Himes, David R. Martinez, and Sallie R. Permar. The impact of the gut microbiota on humoral immunity to pathogens and vaccination in early infancy. *PLOS Pathogens*, 12(12):e1005997, dec 2016.
- [71] Andrés Moya and Manuel Ferrer. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends in Microbiology*, 24(5):402–413, may 2016.
- [72] Didier Gonze, Leo Lahti, Jeroen Raes, and Karoline Faust. Multi-stability and the origin of microbial community types. *The ISME Journal*, 11(10):2159–2166, oct 2017.
- [73] Ana Durbán, Juan J. Abellán, Nuria Jiménez-Hernández, Alejandro Artacho, Vicente Garrigues, Vicente Ortiz, Julio Ponce, Amparo Latorre, and Andrés Moya. Instability of the faecal microbiota in diarrhoea-predominant irritable bowel syndrome. *FEMS Microbiology Ecology*, 86(3):581–589, 2013.
- [74] Xochitl C Morgan, Timothy L Tickle, Harry Sokol, Dirk Gevers, Kathryn L Devaney, Doyle V Ward, Joshua A Reyes, Samir A Shah, Neal LeLeiko, Scott B Snapper, Athos Bousvaros, Joshua Korzenik, Bruce E Sands, Ramnik J Xavier, and Curtis Huttenhower. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology*, 13(9):R79, 2012.
- [75] Manuel Ferrer, Alicia Ruiz, Francesca Lanza, Sven-Bastiaan Haange, Andreas Oberbach, Holger Till, Rafael Bargiela, Cristina Campoy, Maria Teresa Segura, Michael Richter, Martin von Bergen, Jana Seifert, and Antonio Suarez. Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environmental Microbiology*, 15(1):211–226, aug 2013.

- [76] Brandilyn A. Peters, Christine Dominianni, Jean A. Shapiro, Timothy R. Church, Jing Wu, George Miller, Elizabeth Yuen, Hal Freiman, Ian Lustbader, James Salik, Charles Friedlander, Richard B. Hayes, and Jiyoung Ahn. The gut microbiota in conventional and serrated precursors of colorectal cancer. *Microbiome*, 4(1), dec 2016.
- [77] Sergio Serrano-Villar, David Rojo, Mónica Martínez-Martínez, Simon Deusch, Jorge F. Vázquez-Castellanos, Talía Sainz, Mar Vera, Santiago Moreno, Vicente Estrada, María José Gosalbes, Amparo Latorre, Abelardo Margolles, Jana Seifert, Coral Barbas, Andrés Moya, and Manuel Ferrer. HIV infection results in metabolic alterations in the gut microbiota different from those induced by other diseases. *Scientific Reports*, 6(1), may 2016.
- [78] J F Vázquez-Castellanos, S Serrano-Villar, A Latorre, A Artacho, M L Ferrús, N Madrid, A Vallejo, T Sainz, J Martínez-Botas, S Ferrando-Martínez, M Vera, F Drona, M Leal, J Del Romero, S Moreno, V Estrada, M J Gosalbes, and A Moya. Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. *Mucosal Immunology*, 8(4):760–772, nov 2014.
- [79] Richard R. Stein, Vanni Bucci, Nora C. Toussaint, Charlie G. Buffie, Gunnar Räscher, Eric G. Pamer, Chris Sander, and João B. Xavier. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS computational biology*, 9(12):e1003388, dec 2013.
- [80] Stephen M. Collins. A role for the gut microbiota in IBS. *Nature Reviews Gastroenterology & Hepatology*, 11(8):497–505, 2014.
- [81] Ian B Jeffery, Paul W O'Toole, Lena Öhman, Marcus J Claesson, Jennifer Deane, Eamonn M M Quigley, and Magnus Simrén. An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut*, 61(7):997–1006, dec 2011.
- [82] Anna Lyra, Teemu Rinttilä, Janne Nikkilä, Lotta Krogius-Kurikka, Kajsa Kajander, Erja Malinen, Jaana Mättö, Laura Mäkelä, and Airi Palva. Diarrhoea-predominant irritable bowel syndromedistinguishable by 16s rRNA gene phylotype quantification. *World Journal of Gastroenterology*, 15(47):5936, 2009.
- [83] Bédís Dridi, Didier Raoult, and Michel Drancourt. Archaea as emerging organisms in complex human microbiomes. *Anaerobe*, 17(2):56–63, apr 2011.
- [84] Victoria Pascal, Marta Pozuelo, Natalia Borruel, Francesc Casellas, David Campos, Alba Santiago, Xavier Martinez, Encarna Varela, Guillaume Sarrabayrouse, Kathleen Machiels, Severine Vermeire, Harry Sokol, Francisco Guarner, and Chaysavanh Manichanh. A microbial signature for crohn's disease. *Gut*, 66(5):813–822, feb 2017.
- [85] Parth J. Parekh, Luis A. Balart, and David A. Johnson. The Influence of the Gut Microbiome on Obesity, Metabolic Syndrome and Gastrointestinal Disease. *Clinical and Translational Gastroenterology*, 6(6):e91–e91, jun 2015.
- [86] Hua V. Lin, Andrea Frassetto, Edward J. Kowalik Jr, Andrea R. Nawrocki, Mofei M. Lu, Jennifer R. Kosinski, James A. Hubert, Daphne Szeto, Xiaorui Yao, Gail Forrest, and Donald J. Marsh. Butyrate and propionate protect against diet-induced obesity

- and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. *PLoS ONE*, 7(4):e35240, apr 2012.
- [87] Fredrik Bäckhed, Jill K. Manchester, Clay F. Semenkovich, and Jeffrey I. Gordon. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proceedings of the National Academy of Sciences*, 104(3):979–984, jan 2007.
- [88] J.E. Tate, A.H. Burton, C. Boschi-Pinto, U.D. Parashar, M. Agocs, F. Serhan, L. De Oliveira, J.M. Mwenda, R. Mihigo, P. Ranjan Wijesinghe, N. Abeyasinghe, K. Fox, and F. Paladin. Global, regional, and national estimates of rotavirus mortality in children <5 years of age, 2000-2013. *Clinical Infectious Diseases*, 62:S96–S105, 2016. cited By 63.
- [89] Meiling Zhang, Menghui Zhang, Chenhong Zhang, Huimin Du, Guifang Wei, Xiaoyan Pang, Haokui Zhou, Binbin Liu, and Liping Zhao. Pattern extraction of structural responses of gut microbiota to rotavirus infection via multivariate statistical analysis of clone library data. *FEMS Microbiology Ecology*, 70(2):177–185, nov 2009.
- [90] Robert F Ramig. Pathogenesis of intestinal and systemic rotavirus infection. *Journal of virology*, 78(19):10213–10220, 2004.
- [91] Alfredo Guarino, Shai Ashkenazi, Dominique Gendrel, Andrea Lo Vecchio, Raanan Shamir, and Hania Szajewska. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition/European Society for Pediatric Infectious Diseases Evidence-Based Guidelines for the Management of Acute Gastroenteritis in Children in Europe. *Journal of Pediatric Gastroenterology and Nutrition*, 59(1):132–152, jul 2014.
- [92] Ener Cagri Dinleyici, Makbule Eren, Metehan Ozen, Zeynel Abidin Yargic, and Yvan Vandenplas. Effectiveness and safety of *Saccharomyces boulardii* for acute infectious diarrhea. *Expert Opinion on Biological Therapy*, 12(4):395–410, apr 2012.
- [93] Ener Cagri Dinleyici and Yvan Vandenplas. *Lactobacillus reuteri* DSM 17938 effectively reduces the duration of acute diarrhoea in hospitalised children. *Acta Paediatrica*, 103(7):n/a–n/a, mar 2014.
- [94] Ener Cagri Dinleyici, Ates Kara, Metehan Ozen, and Yvan Vandenplas. *Saccharomyces boulardii* CNCM I-745 in different clinical conditions. *Expert Opinion on Biological Therapy*, 14(11):1593–1609, nov 2014.
- [95] Andreas J. Bäumlner and Vanessa Sperandio. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*, 535(7610):85–93, 2016.
- [96] Katharine M. Ng, Jessica A. Ferreyra, Steven K. Higginbottom, Jonathan B. Lynch, Purna C. Kashyap, Smita Gopinath, Natasha Naidu, Biswa Choudhury, Bart C. Weimer, Denise M. Monack, and Justin L. Sonnenburg. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature*, 502(7469):96–99, sep 2013.
- [97] Amy Langdon, Nathan Crook, and Gautam Dantas. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Medicine*, 8(1), apr 2016.

- [98] Aurea Simón-Soro and Alex Mira. Solving the etiology of dental caries. *Trends in Microbiology*, 23(2):76–82, feb 2015.
- [99] William Loeshe. Role of streptococcus mutans in human dental decay. *Microbiological Reviews*, 50(4):353—380, 1986.
- [100] John T. Curnutte, Dana M. Whitten, and Bernard M. Babior. Defective superoxide production by granulocytes from patients with chronic granulomatous disease. *New England Journal of Medicine*, 290(11):593–597, mar 1974.
- [101] Carl Nathan and Amy Cunningham-Bussel. Beyond oxidative stress: an immunologist's guide to reactive oxygen species. *Nature Reviews Immunology*, 13(5):349–361, may 2013.
- [102] F. Ahmadi-Motamayel, MT. Goodarzi, SS. Hendi, S. Kasraei, and A. Moghimbeigi. Total antioxidant capacity of saliva and dental caries. *Medicina Oral Patología Oral y Cirugía Bucal*, pages e553–e556, 2013.
- [103] SVijay Kumar, RHemanth Kumar, Nagaratna Bagewadi, and NitinAnand Krishnan. A study to correlate dental caries experience with total antioxidant levels of saliva among adolescents in mangalore. *Journal of Indian Association of Public Health Dentistry*, 13(2):122, 2015.
- [104] Natália Kamodyová, Gabriel Minárik, Július Hodosy, and Peter Celec. Single consumption of bryndza cheese temporarily affects oral microbiota and salivary markers of oxidative stress. *Current Microbiology*, 69(5):716–724, jul 2014.
- [105] Theodosius Dobzhansky. Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, 35(3):125–129, 1973.
- [106] Charles Darwin. *On the Origin of Species by Means of Natural Selection*. Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.
- [107] Mark Isalan, Caroline Lemerle, Konstantinos Michalodimitrakis, Carsten Horn, Pedro Beltrao, Emanuele Raineri, Mireia Garriga-Canut, and Luis Serrano. Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, 452(7189):840–845, apr 2008.
- [108] Eugene V. Koonin. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Research*, 5:1805, jul 2016.
- [109] Chris S Smillie, Mark B Smith, Jonathan Friedman, Otto X Cordero, Lawrence a David, and Eric J Alm. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, 480(7376):241–4, dec 2011.
- [110] Sabrina Duranti, Gabriele Andrea Lugli, Leonardo Mancabelli, Francesca Turrone, Christian Milani, Marta Mangifesta, Chiara Ferrario, Rosaria Anzalone, Alice Viappiani, Douwe van Sinderen, and Marco Ventura. Prevalence of antibiotic resistance genes among human gut-derived bifidobacteria. *Applied and Environmental Microbiology*, 83(3):e02894–16, nov 2016.

- [111] Stilianos Louca, Martin F. Polz, Florent Mazel, Michaeline B. N. Albright, Julie A. Huber, Mary I. O'Connor, Martin Ackermann, Aria S. Hahn, Diane S. Srivastava, Sean A. Crowe, Michael Doebeli, and Laura Wegener Parfrey. Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, Accepted, apr 2018.
- [112] Mathieu Groussin, Florent Mazel, Jon G. Sanders, Chris S. Smillie, Sébastien Lavergne, Wilfried Thuiller, and Eric J. Alm. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications*, 8:14319, feb 2017.
- [113] A. H. Moeller, A. Caro-Quintero, D. Mjungu, A. V. Georgiev, E. V. Lonsdorf, M. N. Muller, A. E. Pusey, M. Peeters, B. H. Hahn, and H. Ochman. Cospeciation of gut microbiota with hominids. *Science*, 353(6297):380–382, jul 2016.
- [114] Charles K. Fisher and Pankaj Mehta. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS ONE*, 9(7):e102451, 2014.
- [115] Amir Bashan, Travis E. Gibson, Jonathan Friedman, Vincent J. Carey, Scott T. Weiss, Elizabeth L. Hohmann, and Yang-Yu Liu. Universality of human microbial dynamics. *Nature*, 534(7606):259–262, 2016.
- [116] Karoline Faust, Leo Lahti, Didier Gonze, Willem M de Vos, and Jeroen Raes. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current Opinion in Microbiology*, 25(May):56–66, jun 2015.
- [117] Elisa Benincà, Jef Huisman, Reinhard Heerkloss, Klaus D. Jöhnk, Pedro Branco, Egbert H. Van Nes, Marten Scheffer, and Stephen P. Ellner. Chaos in a long-term experiment with a plankton community. *Nature*, 451(7180):822–825, feb 2008.
- [118] Leo Lahti, Jarkko Salojärvi, Anne Salonen, Marten Scheffer, and Willem M. de Vos. Tipping elements in the human intestinal ecosystem. *Nature Communications*, 5, jul 2014.
- [119] Ashley Shade, Jordan S. Read, Nicholas D. Youngblut, Noah Fierer, Rob Knight, Timothy K. Kratz, Noah R. Lottig, Eric E. Roden, Emily H. Stanley, Jesse Stombaugh, Rachel J. Whitaker, Chin H. Wu, and Katherine D. McMahon. Lake microbial communities are resilient after a whole-ecosystem disturbance. *ISME Journal*, 6(12):2153–2167, 2012.
- [120] Les Dethlefsen and David A Relman. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl:4554–61, mar 2011.
- [121] Vanni Bucci, Belinda Tzen, Ning Li, Matt Simmons, Takeshi Tanoue, Elijah Bogart, Luxue Deng, Vladimir Yeliseyev, Mary L. Delaney, Qing Liu, Bernat Olle, Richard R. Stein, Kenya Honda, Lynn Bry, and Georg K. Gerber. MDSINE: Microbial Dynamical Systems INFERENCE Engine for microbiome time-series analyses. *Genome Biology*, 17(1):121, 2016.

- [122] Benjamin Dickens, Charles K. Fisher, and Pankaj Mehta. Analytically tractable model for community ecology with many species. *Physical Review E*, 94(2):022423, aug 2016.
- [123] C. K. Fisher and P. Mehta. The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences*, 111(36):13111–13116, 2014.
- [124] V. Dakos, E. Beninca, E. H. van Nes, C. J. M. Philippart, M. Scheffer, and J. Huisman. Interannual variability in species composition explained as seasonally entrained chaos. *Proceedings of the Royal Society B: Biological Sciences*, 276(1669):2871–2880, may 2009.
- [125] James A. Drake. The mechanics of community assembly and succession. *Journal of Theoretical Biology*, 147(2):213–233, 1990.
- [126] J. A. Capitan, José A. Cuesta, and Jordi Bascompte. Statistical mechanics of ecosystem assembly. *Physical Review Letters*, 103(16):14–17, mar 2009.
- [127] Kathrin Näpflin and Paul Schmid-Hempel. Host effects on microbiota community assembly. *Journal of Animal Ecology*, 87(2):331–340, 2018.
- [128] Thomas Bourguignon, Nathan Lo, Carsten Dietrich, Jan Šobotník, Sarah Sidek, Yves Roisin, Andreas Brune, and Theodore A. Evans. Rampant host switching shaped the termite gut microbiome. *Current Biology*, 28(4):649–654.e2, feb 2018.
- [129] Sören Franzenburg, Sebastian Fraune, Philipp M Altrock, Sven Künzel, John F Baines, Arne Traulsen, and Thomas C G Bosch. Bacterial colonization of Hydra hatchlings follows a robust temporal pattern. *The ISME journal*, 7(4):781–90, 2013.
- [130] Daphna Rothschild, Omer Weissbrod, Elad Barkan, Alexander Kurilshikov, Tal Korem, David Zeevi, Paul Igor Costea, Anastasia Godneva, Iris Nati Kalka, Noam Bar, Smadar Shilo, Dar Lador, Arnau Vich Vila, Niv Zmora, Meirav Pevsner-Fischer, David Israeli, Noa Kosower, Gal Malka, Bat Chen Wolf, Tali Avnit-Sagi, Maya Lotan-Pompan, Adina Weinberger, Zamir Halpern, Shai Carmi, Jingyuan Fu, Cisca Wijmenga, Alexandra Zhernakova, Eran Elinav, and Eran Segal. Environment dominates over host genetics in shaping human gut microbiota. *Nature*, page 150540, feb 2018.
- [131] Kevin R Foster, Jonas Schluter, Katharine Z Coyte, and Seth Rakoff-Nahoum. The evolution of the host microbiome as an ecosystem on a leash. *Nature*, 548(7665):43–51, aug 2017.
- [132] Stéphane Hacquard, Ruben Garrido-Oter, Antonio González, Stijn Spaepen, Gail Ackermann, Sarah Lebeis, Alice C. McHardy, Jeffrey L. Dangl, Rob Knight, Ruth Ley, and Paul Schulze-Lefert. Microbiota and host nutrition across plant and animal kingdoms. *Cell Host & Microbe*, 17(5):603–616, may 2015.
- [133] Petra Louis and Harry J. Flint. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, 19(1):29–41, dec 2016.

- [134] Sarkis K. Mazmanian, Cui Hua Liu, Arthur O. Tzianabos, and Dennis L. Kasper. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, 122(1):107–118, jul 2005.
- [135] Baohong Wang, Mingfei Yao, Longxian Lv, Zongxin Ling, and Lanjuan Li. The Human Microbiota in Health and Disease. *Engineering*, 3(1):71–82, feb 2017.
- [136] Jonas Schluter and Kevin R. Foster. The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biology*, 10(11):e1001424, nov 2012.
- [137] Renee Elizabeth Sockett. Predatory lifestyle of *Bdellovibrio bacteriovorus*. *Annual Review of Microbiology*, 63(1):523–539, oct 2009.
- [138] Petra Louis, Georgina L. Hold, and Harry J. Flint. The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews Microbiology*, 12(10):661–672, sep 2014.
- [139] Kayla C. King, Michael A. Brockhurst, Olga Vasieva, Steve Paterson, Alex Betts, Suzanne A. Ford, Crystal L. Frost, Malcolm J. Horsburgh, Sam Haldenby, and Gregory D.D. Hurst. Rapid evolution of microbe-mediated protection against pathogens in a worm host. *ISME Journal*, 10(8):1915–1924, 2016.
- [140] Aaron G. Wexler, Yiqiao Bao, John C. Whitney, Louis-Marie Bobay, Joao B. Xavier, Whitman B. Schofield, Natasha A. Barry, Alistair B. Russell, Bao Q. Tran, Young Ah Goo, David R. Goodlett, Howard Ochman, Joseph D. Mougous, and Andrew L. Goodman. Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proceedings of the National Academy of Sciences*, 113(13):3639–3644, 2016.
- [141] Daniel M. Cornforth and Kevin R. Foster. Antibiotics and the art of bacterial war: Fig. 1. *Proceedings of the National Academy of Sciences*, 112(35):10827–10828, aug 2015.
- [142] Jessica Ann Thompson, Rita Almeida Oliveira, Ana Djukovic, Carles Ubeda, and Karina Bivar Xavier. Manipulation of the quorum sensing signal AI-2 affects the antibiotic-treated gut microbiota. *Cell Reports*, 10(11):1861–1871, mar 2015.
- [143] Jonas Schluter, Carey D Nadell, Bonnie L Bassler, and Kevin R Foster. Adhesion as a weapon in microbial competition. *The ISME Journal*, 9(1):139–149, oct 2014.
- [144] Jonathan Friedman and Eric J. Alm. Inferring correlation networks from genomic survey data. *PLoS Computational Biology*, 8(9):e1002687, sep 2012.
- [145] Li C Xia, Joshua A Steele, Jacob A Cram, Zoe G Cardon, Sheri L Simmons, Joseph J Vallino, Jed A Fuhrman, and Fengzhu Sun. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Systems Biology*, 5(Suppl 2):S15, 2011.
- [146] Alfred J. Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14(3):271–274, jan 1909.
- [147] Alfred J. Lotka. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7):410–415, jun 1920.

- [148] Kenta Suzuki, Katsuhiko Yoshida, Yumiko Nakanishi, and Shinji Fukuda. An equation-free method reveals the ecological interaction networks within complex microbial ecosystems. *Methods in Ecology and Evolution*, 2017(March):1–12, jun 2017.
- [149] Michael J. McGeachie, Joanne E. Sordillo, Travis Gibson, George M. Weinstock, Yang-Yu Liu, Diane R. Gold, Scott T. Weiss, and Augusto Litonjua. Longitudinal prediction of the infant gut microbiome with dynamic bayesian networks. *Scientific Reports*, 6(1), feb 2016.
- [150] Mahdi Shafiei, Katherine A Dunn, Eva Boon, Shelley M MacDonald, David A Walsh, Hong Gu, and Joseph P Bielawski. BioMiCo: a supervised bayesian model for inference of microbial community structure. *Microbiome*, 3(1):8, 2015.
- [151] Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Second Edition (Studies in Nonlinearity) (Volume 1)*. Westview Press, 2014.
- [152] ROBERT M. MAY. Will a large complex system be stable? *Nature*, 238(5364):413–414, aug 1972.
- [153] K. Z. Coyte, J. Schluter, and K. R. Foster. The ecology of the microbiome: Networks, competition, and stability. *Science*, 350(6261):663–666, nov 2015.
- [154] Raffaella Di Cagno, Maria De Angelis, Ilaria De Pasquale, Maurice Ndagijimana, Pamela Vernocchi, Patrizia Ricciuti, Francesca Gagliardi, Luca Laghi, Carmine Crecchio, Maria Guerzoni, Marco Gobetti, and Ruggiero Francavilla. Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization. *BMC Microbiology*, 11(1):219, 2011.
- [155] J. C. Clemente, E. C. Pehrsson, M. J. Blaser, K. Sandhu, Z. Gao, B. Wang, M. Magris, G. Hidalgo, M. Contreras, O. Noya-Alarcon, O. Lander, J. McDonald, M. Cox, J. Walter, P. L. Oh, J. F. Ruiz, S. Rodriguez, N. Shen, S. J. Song, J. Metcalf, R. Knight, G. Dantas, and M. G. Dominguez-Bello. The microbiome of uncontacted amerindians. *Science Advances*, 1(3):e1500183–e1500183, apr 2015.
- [156] Pieter Van den Abbeele, Tom Van de Wiele, Willy Verstraete, and Sam Possemiers. The host selects mucosal and luminal associations of coevolved gut microorganisms: a novel concept. *FEMS Microbiology Reviews*, 35(4):681–704, jul 2011.
- [157] H. J. Kim, J. Q. Boedicker, J. W. Choi, and R. F. Ismagilov. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proceedings of the National Academy of Sciences*, 105(47):18188–18193, nov 2008.
- [158] Seth Rakoff-Nahoum, Kevin R. Foster, and Laurie E. Comstock. The evolution of cooperation within the gut microbiota. *Nature*, 533(7602):255–259, apr 2016.
- [159] Melanie Ghouil and Sara Mitri. The ecology and evolution of microbial competition. *Trends in Microbiology*, 24(10):833–845, oct 2016.
- [160] Christoph Ratzke, Jonas Denk, and Jeff Gore. Ecological suicide in microbes. *Nature Ecology & Evolution*, page 161398, apr 2018.

- [161] Sergi Valverde, Jordi Piñero, Bernat Corominas-Murtra, Jose Montoya, Lucas Joppa, and Ricard Solé. The architecture of mutualistic networks as an evolutionary spandrel. *Nature Ecology & Evolution*, 2(1):94–99, jan 2018.
- [162] Nicholas Blumm, Gourab Ghoshal, Zalán Forró, Maximilian Schich, Ginestra Bianconi, Jean-Philippe Bouchaud, and Albert-László Barabási. Dynamics of ranking processes in complex systems. *Physical Review Letters*, 109(12), sep 2012.
- [163] D. P. Bakker, B. R. Postmus, H. J. Busscher, and H. C. van der Mei. Bacterial strains isolated from different niches can exhibit different patterns of adhesion to substrata. *Applied and Environmental Microbiology*, 70(6):3758–3760, jun 2004.
- [164] Aaron Lerner, Torsten Matthias, and Rustam Aminov. Potential effects of horizontal gene exchange in the human gut. *Frontiers in Immunology*, 8, nov 2017.
- [165] Aleksandar D. Kostic, Dirk Gevers, Heli Siljander, Tommi Vatanen, Tuulia Hyötyläinen, Anu-Maaria Hämäläinen, Aleksandr Peet, Vallo Tillmann, Päivi Pöhö, Ismo Mattila, Harri Lähdesmäki, Eric A. Franzosa, Outi Vaarala, Marcus de Goffau, Hermie Harmsen, Jorma Ilonen, Suvi M. Virtanen, Clary B. Clish, Matej Orešič, Curtis Huttenhower, Mikael Knip, and Ramnik J. Xavier. The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes. *Cell Host & Microbe*, 17(2):260–273, 2015.
- [166] Catherine a. Lozupone, Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight. Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415):220–230, sep 2012.
- [167] S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmiento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, E. Boss, C. Bowler, M. Follows, L. Karp-Boss, U. Krzic, E. G. Reynaud, C. Sardet, M. Sieracki, and D. Velayoudon and. Structure and function of the global ocean microbiome. *Science*, 348(6237):1261359–1261359, may 2015.
- [168] Jeremiah J Faith, Janaki L Guruge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L Goodman, Jose C Clemente, Rob Knight, Andrew C Heath, Rudolph L Leibel, Michael Rosenbaum, and Jeffrey I Gordon. The Long-Term Stability of the Human Gut Microbiota. *Science*, 341(6141):1237439–1237439, jul 2013.
- [169] Michelle I Smith, Tanya Yatsunenko, Mark J Manary, Indi Trehan, Rajhab Mkakosya, Jiye Cheng, Andrew L Kau, Stephen S Rich, Patrick Concannon, Josyf C Mychaleckyj, Jie Liu, Eric Houtp, Jia V Li, Elaine Holmes, Jeremy Nicholson, Dan Knights, Luke K Ursell, Rob Knight, and Jeffrey I Gordon. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science (New York, N.Y.)*, 339(6119):548–54, 2013.

- [170] Lawrence A David, Corinne F Maurice, Rachel N Carmody, David B Gootenberg, Julie E Button, Benjamin E Wolfe, Alisha V Ling, A Sloan Devlin, Yug Varma, Michael A Fischbach, Sudha B Biddinger, Rachel J Dutton, and Peter J Turnbaugh. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484):559–63, 2014.
- [171] Rachel Poretsky, Luis M. Rodriguez-R, Chengwei Luo, Despina Tsementzi, and Konstantinos T. Konstantinidis. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE*, 9(4):e93827, apr 2014.
- [172] Benjamin J. Callahan, Paul J. McMurdie, and Susan P. Holmes. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, 11(12):2639–2643, 2017.
- [173] Benjamin J. Callahan, Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–583, 2016.
- [174] Michael Tessler, Johannes S. Neumann, Ebrahim Afshinnekoo, Michael Pineda, Rebecca Hersch, Luiz Felipe M. Velho, Bianca T. Segovia, Fabio A. Lansac-Toha, Michael Lemke, Rob Desalle, Christopher E. Mason, and Mercer R. Brugler. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Scientific Reports*, 7(1):1–14, 2017.
- [175] Ravi Ranjan, Asha Rani, Ahmed Metwally, Halvor S. McGee, and David L. Perkins. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications*, 469(4):967–977, 2015.
- [176] Bo Yang, Yong Wang, and Pei Yuan Qian. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, 17(1):1–8, 2016.
- [177] Kamila Polgarova, Michal Behuliak, and Peter Celec. Effect of saliva processing on bacterial dna extraction. *New Microbiologica*, 33(1):373–379, may 2010.
- [178] Heng Zhu, Feng Qu, and Li huang Zhu. Isolation of genomic DNAs from plants, fungi and bacteria using benzyl chloride. *Nucleic Acids Research*, 21(22):5279–5280, 1993.
- [179] Peter J. A. Cock, Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. The sanger FASTQ file format for sequences with quality scores, and the solexa/illumina FASTQ variants. *Nucleic Acids Research*, 38(6):1767–1771, dec 2009.
- [180] A Gordon and GJ Hannon. Fastx-toolkit. fastq/a shortreads pre-processing tools v0.0.13, 2010.
- [181] Robert Schmieder and Robert Edwards. Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*, 27(6):863–864, March 2011. PMID: 21278185.

- [182] Robert R Sokal and Peter Sneath. *Principles of Numerical Taxonomy*. W. H. Freeman, 1963.
- [183] M. Blaxter, J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462):1935–1943, oct 2005.
- [184] David P Clark, Daniel P Buckley, David A Stahl, Michael T Madiga, and John M Martinko. *Brock Biology of Microorganisms with MasteringMicrobiology, Global Edition, 14/E*. Pearson Education Limited, 2014.
- [185] Ramon Rosselló-Mora and Rudolf Amann. The species concept for prokaryotes. *FEMS Microbiology Reviews*, 25(1):39–67, jan 2001.
- [186] J Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, Antonio Gonzalez Peña, Julia K Goodrich, Jeffrey I Gordon, Gavin a Huttley, Scott T Kelley, Dan Knights, Jeremy E Koenig, Ruth E Ley, Catherine a Lozupone, Daniel McDonald, Brian D Muegge, Meg Pirrung, Jens Reeder, Joel R Sevinsky, Peter J Turnbaugh, William a Walters, Jeremy Widmann, Tanya Yatsunenko, Jesse Zaneveld, and Rob Knight. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5):335–336, may 2010.
- [187] Robert C. Edgar. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461, aug 2010.
- [188] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7):5069–5072, jul 2006.
- [189] Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1):D590–D596, nov 2012.
- [190] Christopher Quince, Alan W. Walker, Jared T. Simpson, Nicholas J. Loman, and Nicola Segata. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9):833–844, 2017.
- [191] Despoina D. Roumpeka, R. John Wallace, Frank Escalettes, Ian Fotheringham, and Mick Watson. A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Frontiers in Genetics*, 8(MAR):1–10, 2017.
- [192] Sasha K. Ames, David A. Hysom, Shea N. Gardner, G. Scott Lloyd, Maya B. Gokhale, and Jonathan E. Allen. Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics*, 29(18):2253–60, sep 2013.
- [193] Sasha K. Ames, Shea N. Gardner, Jose Manuel Marti, Tom R. Slezak, Maya B. Gokhale, and Jonathan E. Allen. Using populations of human and microbial genomes for organism detection in metagenomes. *Genome Research*, 25(7):1056–1067, jul 2015.

- [194] Ľubomíra Tóthová, Natália Kamodyová, Tomáš Červenka, and Peter Celec. Salivary markers of oxidative stress in oral diseases. *Frontiers in Cellular and Infection Microbiology*, 5(October), oct 2015.
- [195] Isabella Dalle-Donne, Ranieri Rossi, Daniela Giustarini, Aldo Milzani, and Roberto Colombo. Protein carbonyl groups as biomarkers of oxidative stress. *Clinica Chimica Acta*, 329(1-2):23–38, mar 2003.
- [196] Ozcan Erel. A novel automated direct measurement method for total antioxidant capacity using a new generation, more stable ABTS radical cation. *Clinical Biochemistry*, 37(4):277–285, apr 2004.
- [197] Lenka Baňasová, Natália Kamodyová, Katarína Janšáková, Ľubomíra Tóthová, Peter Stanko, Ján Turňa, and Peter Celec. Salivary DNA and markers of oxidative stress in patients with chronic periodontitis. *Clinical Oral Investigations*, 19(2):201–207, mar 2014.
- [198] Michal Behuliak, Roland Pálffy, Roman Gardlík, Július Hodosy, Lukáč Halčák, and Peter Celec. Variability of thiobarbituric acid reacting substances in saliva. *Disease Markers*, 26(2):49–53, 2009.
- [199] Jian-Ting Chen. Advanced oxidation protein products as a novel marker of oxidative stress in postmenopausal osteoporosis. *Medical Science Monitor*, 21:2428–2432, 2015.
- [200] G Münch, R Keis, a Wessels, P Riederer, U Bahner, a Heidland, T Niwa, H D Lemke, and R Schinzel. Determination of advanced glycation end products in serum by fluorescence spectroscopy and competitive ELISA. *European journal of clinical chemistry and clinical biochemistry : journal of the Forum of European Clinical Chemistry Societies*, 35(9):669–677, 1997.
- [201] Iris F.F. Benzie and J.J. Strain. The ferric reducing ability of plasma (FRAP) as a measure of “antioxidant power”: The FRAP assay. *Analytical Biochemistry*, 239(1):70–76, jul 1996.
- [202] R. H. Whittaker. Vegetation of the siskiyou mountains, oregon and california. *Ecological Monographs*, 30(3):279–338, feb 1960.
- [203] Lou Jost. Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10):2427–2439, 2007.
- [204] Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlenn, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, and Helene Wagner. *vegan: Community Ecology Package*, 2017. R package version 2.4-3.
- [205] Torsten Hothorn, Kurt Hornik, Mark A. van de Wiel, and Achim Zeileis. Implementing a class of permutation tests: The coin package. *Journal of Statistical Software*, 28(8):1–23, 2008.

- [206] Marti J. Anderson, Thomas O. Crist, Jonathan M. Chase, Mark Vellend, Brian D. Inouye, Amy L. Freestone, Nathan J. Sanders, Howard V. Cornell, Liza S. Comita, Kendi F. Davies, Susan P. Harrison, Nathan J. B. Kraft, James C. Stegen, and Nathan G. Swenson. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*, 14(1):19–28, nov 2010.
- [207] Hanna Tuomisto. A diversity of beta diversities: straightening up a concept gone awry. part 1. defining beta diversity as a function of alpha and gamma diversity. *Ecography*, 33(1):2–22, feb 2010.
- [208] Hanna Tuomisto. A diversity of beta diversities: straightening up a concept gone awry. part 2. quantifying beta diversity and related phenomena. *Ecography*, 33(1):23–45, feb 2010.
- [209] C. Lozupone and R. Knight. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12):8228–8235, dec 2005.
- [210] J. Roger Bray and J. T. Curtis. An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs*, 27(4):325–349, feb 1957.
- [211] Nicola Segata, Jacques Izard, Levi Waldron, Dirk Gevers, Larisa Miropolsky, Wendy S Garrett, and Curtis Huttenhower. Metagenomic biomarker discovery and explanation. *Genome biology*, 12(6):R60, jan 2011.
- [212] Anna Heintz-Buschart, Patrick May, Cédric C. Laczny, Laura A. Lebrun, Camille Bellora, Abhimanyu Krishna, Linda Wampach, Jochen G. Schneider, Angela Hogan, Carine de Beaufort, Paul Wilmes, M. Arumugam, C. Huttenhower, T. Yatsunenko, S. Schloissnig, J. J. Faith, P. J. Turnbaugh, J. Qin, N. Qin, E. A. Franzosa, M.-C. Arrieta, N. C. Verberkmoes, P. J. Turnbaugh, M. J. Gosalbes, A. R. Erickson, M. Ferrer, W. Xiong, R. J. Giannone, M. J. Morowitz, J. F. Banfield, R. L. Hettich, A. E. Pérez-Cobas, M. K. Waldor, F. Beulig, J. Hultman, J. K. Goodrich, P. D. Schloss, K. D. Iverson, J. F. Petrosino, S. J. Schloss, S. J. Song, C. C. Patterson, K. M. Gillespie, M. A. Atkinson, A. Chervonsky, S. M. Cabrera, A. M. Henschel, M. J. Hessner, T. Rodriguez-Calvo, O. Ekwall, N. Amirian, J. Zapardiel-Gonzalo, M. G. von Herrath, A. Giongo, C. T. Brown, D. Endesfelder, A. D. Kostic, M. Murri, M. C. de Goffau, A. G. Davis-Richardson, E. E. L. Muller, H. Roume, S. Sunagawa, E. A. Franzosa, B. K. Tan, R. Adya, H. S. Randeve, D. Legrand, A. T. French, Y. Akiyama, L. Bertuccini, M. Uhlen, M. A. Atkinson, J. Walter, R. Ley, L. Engelen, C. D. A. Stehouwer, C. G. Schalkwijk, S. M. Nabokina, R. Joice, K. Yasuda, A. Shafquat, X. C. Morgan, C. Huttenhower, A. A. Welch, R. Luben, K. T. Khaw, S. A. Bingham, W. Ammerlaan, H. Roume, A. Heintz-Buschart, E. E. L. Muller, P. Wilmes, J. R. Kultima, R. Li, C. Quast, B. Langmead, S. L. Salzberg, D. E. Wood, S. L. Salzberg, J. Li, P. J. McMurdie, S. Holmes, A. P. Reynolds, G. Richards, B. de la Iglesia, V. J. Rayward-Smith, L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, Y. Peng, H. C. M. Leung, S. M. Yiu, F. Y. L. Chin, D. R. Zerbino, E. Birney, M. Scholz, C.-C. Lo, P. S. G. Chain, D. Hyatt, C. L. Dupont, M. Albertsen, M. Wu, A. J. Scott, R. D. Finn, S. R. Eddy, R. D. Finn, D. H. Haft, R. Caspi, N. Christian, P. May, S. Kempa, T. Handorf, O. Ebenhöf, H. Li, C. C. Laczny, N. Pinel, N. Vlassis, P. Wilmes, M. Ester, H.-P. Kriegel, J. Sander, X. Xu, T. Benaglia, D. Chauveau, D. Hunter, R. Drmanac, J. Schubert, K. Wang,

- M. Li, H. Hakonarson, A. Rimmer, P. Danecek, B. H. McArdle, M. J. Anderson, C.-H. Chiu, Y.-T. Wang, B. A. Walther, A. Chao, A. Baselga, C. D. L. Orme, E. Paradis, J. Claude, K. Strimmer, C. Meng, B. Kuster, A. C. Culhane, A. M. Gholami, M. I. Love, W. Huber, S. Anders, G. Csardi, T. Nepusz, D. Beisser, G. W. Klau, T. Dandekar, T. Muller, M. T. Dittrich, M. T. Dittrich, G. W. Klau, A. Rosenwald, T. Dandekar, T. Müller, W. Luo, C. Brouwer, F. Meyer, J. A. Vizcaino, and R. K. Aziz. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nature Microbiology*, 2(1):16180, 2016.
- [213] Ohad Manor and Elhanan Borenstein. Systematic Characterization and Analysis of the Taxonomic Drivers of Functional Shifts in the Human Resource Systematic Characterization and Analysis of the Taxonomic Drivers of Functional Shifts in the Human Microbiome. *Cell Host and Microbe*, 21(2):1–14, 2017.
- [214] Xochitl C Morgan and Curtis Huttenhower. Meta’omic Analytic Techniques for Studying the Intestinal Microbiome. *Gastroenterology*, 146(6):1437–1448.e1, may 2014.
- [215] Stefanía Magnúsdóttir, Almut Heinken, Laura Kutt, Dmitry A. Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, Christian Jäger, Joanna Baginska, Paul Wilmes, Ronan M.T. Fleming, and Ines Thiele. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1):81–89, 2017.
- [216] Morgan G I Langille, Jesse Zaneveld, J Gregory Caporaso, Daniel McDonald, Dan Knights, Joshua A Reyes, Jose C Clemente, Deron E Burkepille, Rebecca L Vega Thurber, Rob Knight, Robert G Beiko, and Curtis Huttenhower. Predictive functional profiling of microbial communities using 16s rRNA marker gene sequences. *Nature Biotechnology*, 31(9):814–821, aug 2013.
- [217] M. Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, jan 2000.
- [218] Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, nov 2016.
- [219] Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1):D457–D462, oct 2015.
- [220] Taiyun Wei and Viliam Simko. *R package "corrplot": Visualization of a Correlation Matrix*, 2017. (Version 0.84).
- [221] A. Eklund. *beeswarm: The Bee Swarm Plot, an Alternative to Stripchart*, 2016. R package version 0.2.3.
- [222] Wolfram Research Inc. *Mathematica 11.0*, 2017.
- [223] C. F. J. Wu. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, 14(4):1261–1295, dec 1986.

- [224] Xiao Xiao, Ethan P. White, Mevin B. Hooten, and Susan L. Durham. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*, 92(10):1887–1894, oct 2011.
- [225] Don S. Lemons and Anthony Gythiel. Paul langevin’s 1908 paper “on the theory of brownian motion” [“sur la théorie du mouvement brownien,” c. r. acad. sci. (paris) 146, 530–533 (1908)]. *American Journal of Physics*, 65(11):1079–1081, nov 1997.
- [226] A. Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 322(8):549–560, 1905.
- [227] J. Weber. Fluctuation dissipation theorem. *Physical Review*, 101(6):1620–1626, mar 1956.
- [228] Ricard V. Solé and Sergi Valverde. Information Theory of Complex Networks: On Evolution and Architectural Constraints. In *Complex networks*, volume 207, pages 189–207. Springer, aug 2004.
- [229] Konstantin Klemm and Víctor M. Eguíluz. Growing scale-free networks with small-world behavior. *Physical Review E*, 65(5), may 2002.
- [230] Karoline Faust and Leo Lahti. *The seqtime R package 2017*, 2017. R package version 1.1.1.
- [231] Jose Manuel Martí, Daniel Martínez-Martínez, Teresa Rubio, César Gracia, Manuel Peña, Amparo Latorre, Andrés Moya, and Carlos P. Garay. Health and Disease Imprinted in the Time Variability of the Human Microbiome. *mSystems*, 2(2):e00144–16, apr 2017.
- [232] MATLAB. *version 9.3.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2017.
- [233] Andrew D. Horchler. *Matlab Toolbox for Numerical Solution of Stochastic Differential Equations (SDETools)*, 2011.
- [234] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, March 2014.
- [235] Zoltán Eisler, Imre Bartos, and János Kertész. Fluctuation scaling in complex systems: Taylor’s law and beyond1. *Advances in Physics*, 57(1):89–142, jan 2008.
- [236] L. R. Taylor. Aggregation, variance and the mean. *Nature*, 189(4766):732–735, mar 1961.
- [237] César A. Hidalgo, Nicholas Blumm, Albert-László Barabási, and Nicholas A. Christakis. A dynamic network approach for the study of human phenotypes. *PLoS Computational Biology*, 5(4):e1000353, apr 2009.
- [238] Bent Jørgensen, José Raúl Martínez, and Min Tsao. Asymptotic behaviour of the variance function. *Scandinavian Journal of Statistics*, 21(3):223–243, 1994.

- [239] Agata Fronczak and Piotr Fronczak. Origins of Taylor's power law for fluctuation scaling in complex systems. *Physical Review E*, 81(6), jun 2010.
- [240] Wayne S. Kendal and Bent Jørgensen. Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Physical Review E*, 83(6), jun 2011.
- [241] Wayne S. Kendal and Bent Jørgensen. Tweedie convergence: A mathematical basis for Taylor's power law, $1/f$ noise, and multifractality. *Physical Review E*, 84(6), dec 2011.
- [242] Ana Elena Pérez-Cobas, Elisa Maiques, Alexandra Angelova, Purificación Carrasco, Andrés Moya, and Amparo Latorre. Diet shapes the gut microbiota of the omnivorous cockroach *Blattella germanica*. *FEMS Microbiology Ecology*, 91(4), feb 2015.
- [243] Tania Rosas, Carlos García-Ferris, Rebeca Domínguez-Santos, Pablo Llop, Amparo Latorre, and Andrés Moya. Rifampicin treatment of *Blattella germanica* evidences a fecal transmission route of their gut microbiota. *FEMS Microbiology Ecology*, 94(2), jan 2018.
- [244] John C. Wooley, Adam Godzik, and Iddo Friedberg. A primer on metagenomics. *PLoS Computational Biology*, 6(2):e1000667, feb 2010.
- [245] A. Camelo-Castillo, A. Benitez-Paez, P. Belda-Ferre, R. Cabrera-Rubio, and A. Mira. *Streptococcus dentisani* sp. nov., a novel member of the mitis group. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 64(Pt 1):60–65, jan 2014.
- [246] J. R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje. The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 37(Database):D141–D145, jan 2009.
- [247] S. Dray and A.B. Dufour. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4):1–20, 2007.
- [248] Wei Chen, Clarence K. Zhang, Yongmei Cheng, Shaowu Zhang, and Hongyu Zhao. A comparison of methods for clustering 16S rRNA sequences into OTUs. *PLoS ONE*, 8(8):e70837, aug 2013.
- [249] Jack A. Gilbert, Robert A. Quinn, Justine Debelius, Zhenjiang Z. Xu, James Morton, Neha Garg, Janet K. Jansson, Pieter C. Dorrestein, and Rob Knight. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature*, 535(7610):94–103, jul 2016.
- [250] A. M. Kilpatrick and A. R. Ives. Species interactions can explain Taylor's power law for ecological time series. *Nature*, 422(6927):65–68, mar 2003.
- [251] Ford Ballantyne IV and Andrew J. Kerkhoff. The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. *Oikos*, 116(1):174–180, jan 2007.

- [252] J. E. Koenig, A. Spor, N. Scalfone, A. D. Fricker, J. Stombaugh, R. Knight, L. T. Angenent, and R. E. Ley. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, 108(Supplement_1):4578–4585, jul 2010.
- [253] Peter E. Larsen and Yang Dai. Metabolome of human gut microbiome is predictive of host dysbiosis. *GigaScience*, 4(1), sep 2015.
- [254] B. J. Gardiner, A. Y. Tai, D. Kotsanas, M. J. Francis, S. A. Roberts, S. A. Ballard, R. K. Junckerstorff, and T. M. Korman. Clinical and microbiological characteristics of *eggerthella lenta* bacteremia. *Journal of Clinical Microbiology*, 53(2):626–635, dec 2014.
- [255] Julia K. Goodrich, Jillian L. Waters, Angela C. Poole, Jessica L. Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, William Van Treuren, Rob Knight, Jordana T. Bell, Timothy D. Spector, Andrew G. Clark, and Ruth E. Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, 2014.
- [256] E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. M. Bohannan, and D. A. Relman. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science*, 336(6086):1255–1262, 2012.
- [257] M. J. Claesson, S. Cusack, O. O’Sullivan, R. Greene-Diniz, H. de Weerd, E. Flannery, J. R. Marchesi, D. Falush, T. Dinan, G. Fitzgerald, C. Stanton, D. van Sinderen, M. O’Connor, N. Harnedy, K. O’Connor, C. Henry, D. O’Mahony, A. P. Fitzgerald, F. Shanahan, C. Twomey, C. Hill, R. P. Ross, and P. W. O’Toole. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences*, 108(Supplement_1):4586–4591, 2011.
- [258] Hao Chung The, Paola Florez De Sessions, Song Jie, Duy Pham Thanh, Corinne N Thompson, Chau Nguyen, Ngoc Minh, Collins Wenhan Chu, Anh Tran, Nicholas R Thomson, Guy E Thwaites, Maia A Rabaa, Stephen Baker, Corinne N Thompson, Chau Nguyen, Ngoc Minh, Collins Wenhan Chu, Tuan-anh Tran, R Thomson, Guy E Thwaites, Maia A Rabaa, Martin Hibberd, Stephen Baker Assessing, Paola Florez De Sessions, Song Jie, Tuan-anh Tran, Nicholas R Thomson, E Thwaites, Maia A Rabaa, Martin Hibberd, and Stephen Baker. Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children. *Gut Microbes*, 0(0):1–17, 2017.
- [259] Mihai Pop, Alan W Walker, Joseph Paulson, Brianna Lindsay, Martin Antonio, M Hos-sain, Joseph Oundo, Boubou Tamboura, Volker Mai, Irina Astrovskaya, Hector Bravo, Richard Rance, Mark Stares, Myron M Levine, Sandra Panchalingam, Karen Kotloff, Usman N Ikumapayi, Chinelo Ebruke, Mitchell Adeyemi, Dilruba Ahmed, Firoz Ahmed, Meer Alam, Ruhul Amin, Sabbir Siddiqui, John B Ochieng, Emmanuel Ouma, Jane Juma, Eunce Mailu, Richard Omere, J Morris, Robert F Breiman, Deb-sish Saha, Julian Parkhill, James P Nataro, and O Stine. Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biology*, 15(6):R76, 2014.
- [260] Sathish Subramanian, Sayeeda Huq, Tanya Yatsunenko, Rashidul Haque, Mustafa Mahfuz, Mohammed A Alam, Amber Benezra, Joseph DeStefano, Martin F Meier,

- Brian D Muegge, Michael J Barratt, Laura G VanArendonk, Qunyuan Zhang, Michael A Province, William A Petri, Tahmeed Ahmed, and Jeffrey I Gordon. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*, 510(7505):417–21, 2014.
- [261] Lawrence A. David, Ana Weil, Edward T. Ryan, Stephen B. Calderwood, Jason B. Harris, Fahima Chowdhury, Yasmin Begum, Firdausi Qadri, Regina C. LaRocque, and Peter J. Turnbaugh. Gut Microbial Succession Follows Acute Secretory Diarrhea in Humans. *mBio*, 6(3):e00381–15, jul 2015.
- [262] Karen L. Kotloff, James P. Nataro, William C. Blackwelder, Dilruba Nasrin, Tamer H. Farag, Sandra Panchalingam, Yukun Wu, Samba O. Sow, Dipika Sur, Robert F. Breiman, Abu S.G. Faruque, Anita K.M. Zaidi, Debasish Saha, Pedro L. Alonso, Boubou Tamboura, Doh Sanogo, Uma Onwuchekwa, Byomkesh Manna, Thandavarayan Ramamurthy, Suman Kanungo, John B. Ochieng, Richard Omoro, Joseph O. Oundo, Anowar Hossain, Sumon K. Das, Shahnawaz Ahmed, Shahida Qureshi, Farheen Quadri, Richard A. Adegbola, Martin Antonio, M. Jahangir Hossain, Adebayo Akinsola, Inacio Mandomando, Tacilta Nhampossa, Sozinho Acácio, Kousick Biswas, Ciara E. O'Reilly, Eric D. Mintz, Lynette Y. Berkeley, Khitam Muhsen, Halvor Sommerfelt, Roy M. Robins-Browne, and Myron M. Levine. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet*, 382(9888):209–222, jul 2013.
- [263] G Binda, E Domenichini, A Gottardi, B Orlandi, E Ortelli, B Pacini, and G Fowst. Rifampicin, a general review. *Arzneimittel-forschung*, 21(12):1907, 1971.
- [264] Dipanjan Dutta, Parikshit Bagchi, Arunachal Chatterjee, Mukti Kant Nayak, Anupam Mukherjee, Shiladitya Chattopadhyay, Shigeo Nagashima, Nobumichi Kobayashi, Satoshi Komoto, Koki Taniguchi, and Mamta Chawla-Sarkar. The molecular chaperone heat shock protein-90 positively regulates rotavirus infection. *Virology*, 391(2):325–333, 2009.
- [265] Erin L. Gross, Clifford J. Beall, Stacey R. Kutsch, Noah D. Firestone, Eugene J. Leys, and Ann L. Griffen. Beyond streptococcus mutans: Dental caries onset linked to multiple species by 16s rRNA community analysis. *PLoS ONE*, 7(10):e47722, oct 2012.
- [266] Scott N. Peterson, Erik Snestrud, Nicholas J. Schork, and Walter A. Bretz. Dental caries pathogenicity: a genomic and metagenomic perspective. *International Dental Journal*, 61:11–22, jul 2011.
- [267] Fang Yang, Xiaowei Zeng, Kang Ning, Kuan-Liang Liu, Chien-Chi Lo, Wei Wang, Jie Chen, Dongmei Wang, Ranran Huang, Xingzhi Chang, Patrick S Chain, Gary Xie, Junqi Ling, and Jian Xu. Saliva microbiomes distinguish caries-active from healthy human populations. *The ISME Journal*, 6(1):1–10, jun 2011.
- [268] Justine Debelius, Se Jin Song, Yoshiki Vazquez-Baeza, Zhenjiang Zech Xu, Antonio Gonzalez, and Rob Knight. Tiny microbes, enormous impacts: what matters in gut microbiome studies? *Genome Biology*, 17(1), oct 2016.

- [269] Viera Celecová, Natália Kamodyová, Ľubomíra Tóthová, Matúš Kúdela, and Peter Celec. Salivary markers of oxidative stress are related to age and oral health in adult non-smokers. *Journal of Oral Pathology & Medicine*, 42(3):263–266, sep 2012.
- [270] A. Mardinoglu, S. Shoaie, M. Bergentall, P. Ghaffari, C. Zhang, E. Larsson, F. Backhed, and J. Nielsen. The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Molecular Systems Biology*, 11(10):834–834, oct 2015.
- [271] Jose Manuel Almerich-Silla, Jose María Montiel-Company, Sara Pastor, Felipe Serano, Miriam Puig-Silla, and Francisco Dasí. Oxidative stress parameters in saliva and its association with periodontal disease and types of bacteria. *Disease Markers*, 2015:1–7, 2015.
- [272] Ľubomíra Tóthová, Daniela Ostatníková, Katarína Šebeková, Peter Celec, and Július Hodosy. Sex differences of oxidative stress markers in young healthy subjects are marker-specific in plasma but not in saliva. *Annals of Human Biology*, 40(2):175–180, jan 2013.
- [273] Kei Fujishima, Miki Kawada-Matsuo, Yuichi Oogai, Masayuki Tokuda, Mitsuo Torii, and Hitoshi Komatsuzawa. dprandsodin streptococcus mutans are involved in coexistence with s. sanguinis, and PerR is associated with resistance to h₂O₂. *Applied and Environmental Microbiology*, 79(5):1436–1443, dec 2012.
- [274] P. Naaber. Inhibition of clostridium difficile strains by intestinal lactobacillus species. *Journal of Medical Microbiology*, 53(6):551–554, jun 2004.
- [275] Sebastian E. Winter, Parameth Thiennimitr, Maria G. Winter, Brian P. Butler, Douglas L. Huseby, Robert W. Crawford, Joseph M. Russell, Charles L. Bevins, L. Garry Adams, Renée M. Tsois, John R. Roth, and Andreas J. Bäumlner. Gut inflammation provides a respiratory electron acceptor for salmonella. *Nature*, 467(7314):426–429, sep 2010.
- [276] Francesc Peris-Bondia, Amparo Latorre, Alejandro Artacho, Andrés Moya, and Giuseppe D'Auria. The active human gut microbiota differs from the total microbiota. *PLoS ONE*, 6(7):e22448, jul 2011.
- [277] Corinne Ferrier Maurice, Henry Joseph Haiser, and Peter James Turnbaugh. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell*, 152(1-2):39–50, jan 2013.
- [278] Giuseppe D'Auria, Francesc Peris-Bondia, Mária Džunková, Alex Mira, Maria Carmen Collado, Amparo Latorre, and Andrés Moya. Active and secreted IgA-coated bacterial fractions from the human gut reveal an under-represented microbiota core. *Scientific Reports*, 3(1), dec 2013.
- [279] David Rojo, Celia Méndez-García, Beata Anna Raczkowska, Rafael Bargiela, Andrés Moya, Manuel Ferrer, and Coral Barbas. Exploring the human microbiome from multiple perspectives: factors altering its composition and function. *FEMS Microbiology Reviews*, 41(4):453–478, feb 2017.

- [280] Benoit Delahaye, Damien Eveillard, and Nicholas Bouskill. On the Power of Uncertainties in Microbial System Modeling: No Need To Hide Them Anymore. *mSystems*, 2(6):e00169–17, dec 2017.
- [281] Keith C. Heyde and Warren C. Ruder. Exploring host-microbiome interactions using an in silico model of biomimetic robots and engineered living cells. *Scientific Reports*, 5(1), jul 2015.
- [282] Raymond E Goldstein. Are theoretical results ‘results’? *eLife*, 7, jul 2018.
- [283] Ravi U. Sheth, Vitor Cabral, Sway P. Chen, and Harris H. Wang. Manipulating bacterial communities by in situ microbiome engineering. *Trends in Genetics*, 32(4):189–200, apr 2016.
- [284] David Bikard, Chad W Euler, Wenyan Jiang, Philip M Nussenzweig, Gregory W Goldberg, Xavier Duportet, Vincent A Fischetti, and Luciano A Marraffini. Exploiting CRISPR-cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology*, 32(11):1146–1150, oct 2014.
- [285] S. M. Nayeemul Bari, Forrest C. Walker, Katie Cater, Barbaros Aslan, and Asma Hatoum-Aslan. Strategies for editing virulent staphylococcal phages using CRISPR-cas10. *ACS Synthetic Biology*, 6(12):2316–2325, sep 2017.

Appendix

A.1 Chapter I - Supplementary information

A.1.1 Supplementary Tables

Table A.1 Taylor's parameters of individuals with either animal-based (A) or plant-based (P) diets [170]. Previous to the diet, the population sampled was described by $\bar{V} = 0.09 \pm 0.05$, $\bar{\beta} = 0.77 \pm 0.04$.

Metadata	V	β	\bar{R}^2	V_{st}	β_{st}
A	0.26 ± 0.05	0.826 ± 0.025	0.918	3.1 ± 0.9	1.2 ± 0.6
A	0.32 ± 0.06	0.857 ± 0.025	0.924	4.4 ± 1.1	2.0 ± 0.6
A	0.194 ± 0.033	0.813 ± 0.024	0.918	1.9 ± 0.6	0.9 ± 0.6
A	0.24 ± 0.04	0.824 ± 0.020	0.924	2.7 ± 0.7	1.2 ± 0.5
A	0.34 ± 0.06	0.855 ± 0.024	0.931	4.7 ± 1.1	1.9 ± 0.6
A	0.30 ± 0.05	0.847 ± 0.022	0.921	3.9 ± 1.0	1.7 ± 0.5
A	0.133 ± 0.021	0.784 ± 0.023	0.916	0.7 ± 0.4	0.2 ± 0.6
A	0.25 ± 0.04	0.831 ± 0.024	0.929	3.0 ± 0.8	1.4 ± 0.6
P	0.23 ± 0.05	0.804 ± 0.035	0.885	2.6 ± 0.9	0.7 ± 0.8
P	0.097 ± 0.018	0.705 ± 0.031	0.891	0.03 ± 0.34	-1.6 ± 0.7
P	0.037 ± 0.006	0.642 ± 0.025	0.881	-1.12 ± 0.11	-3.1 ± 0.6
P	0.118 ± 0.019	0.723 ± 0.025	0.895	0.4 ± 0.4	-1.2 ± 0.6
P	0.17 ± 0.04	0.78 ± 0.04	0.842	1.5 ± 0.7	0.1 ± 0.9
P	0.123 ± 0.020	0.757 ± 0.026	0.914	0.5 ± 0.4	-0.4 ± 0.6
P	0.19 ± 0.05	0.77 ± 0.04	0.871	1.8 ± 0.9	-0.0 ± 0.9
P	0.121 ± 0.020	0.736 ± 0.027	0.921	0.5 ± 0.4	-0.9 ± 0.6
P	0.187 ± 0.034	0.771 ± 0.030	0.908	1.8 ± 0.7	-0.1 ± 0.7
P	0.097 ± 0.015	0.735 ± 0.025	0.922	0.05 ± 0.28	-0.9 ± 0.6

Table A.2 Taylor's parameters for individuals taking antibiotics (Ab) in the antibiotics study [120], persons diagnosed with irritable bowel syndrome (IBS) in the IBS study [73] and for special intervals concerning gut microbiota (HLS) in the host lifestyle study [59]. Prior to the antibiotics intake, the population sampled in the antibiotics study [120] was described by $\bar{V} = 0.12 \pm 0.05$, $\bar{\beta} = 0.75 \pm 0.04$. Healthy individuals sampled in the IBS study [73] were characterized by $\bar{V} = 0.135 \pm 0.010$, $\bar{\beta} = 0.692 \pm 0.024$. The healthy and quotidian periods in the host lifestyle study [59] are characterized by $\bar{V} = 0.25 \pm 0.09$, $\bar{\beta} = 0.777 \pm 0.025$.

Metadata	V	β	\bar{R}^2	V_{st}	β_{st}
Ab	0.35 ± 0.07	0.81 ± 0.04	0.925	4.3 ± 1.4	1.3 ± 0.9
Ab	0.41 ± 0.09	0.82 ± 0.04	0.908	5.6 ± 1.8	1.6 ± 0.9
Ab	0.23 ± 0.04	0.770 ± 0.031	0.920	2.1 ± 0.8	0.5 ± 0.7
Ab	0.165 ± 0.029	0.738 ± 0.031	0.928	0.9 ± 0.6	-0.3 ± 0.7
Ab	0.34 ± 0.06	0.812 ± 0.032	0.936	4.1 ± 1.2	1.5 ± 0.7
Ab	0.26 ± 0.05	0.798 ± 0.033	0.931	2.8 ± 0.9	1.1 ± 0.8
IBS (minor)	0.205 ± 0.034	0.740 ± 0.029	0.917	6.9 ± 3.3	2.0 ± 1.2
IBS (severe)	0.35 ± 0.06	0.793 ± 0.025	0.934	21 ± 6	4.2 ± 1.0
HLS (abroad)	0.51 ± 0.06	0.820 ± 0.012	0.928	2.8 ± 0.6	1.7 ± 0.5
HLS (infection)	0.49 ± 0.08	0.828 ± 0.018	0.923	2.6 ± 0.9	2.0 ± 0.7
HLS (after infection)	0.36 ± 0.05	0.776 ± 0.015	0.922	1.1 ± 0.6	-0.0 ± 0.6

Table A.3 Taylor's parameters for the healthy subject (DH) and kwashiorkor part (DK) of the discordant twins [169]. The population of healthy twins is characterized by $\bar{V} = 0.25 \pm 0.10$, $\bar{\beta} = 0.863 \pm 0.028$.

Metadata	V	β	\bar{R}^2	V_{st}	β_{st}
DH	0.27 ± 0.04	0.835 ± 0.016	0.925	0.2 ± 0.4	-1.0 ± 0.6
DH	0.36 ± 0.06	0.858 ± 0.015	0.929	1.1 ± 0.6	-0.2 ± 0.5
DH	0.35 ± 0.06	0.859 ± 0.014	0.926	1.0 ± 0.5	-0.1 ± 0.5
DH	0.25 ± 0.04	0.829 ± 0.014	0.911	0.0 ± 0.4	-1.2 ± 0.5
DH	0.30 ± 0.05	0.844 ± 0.014	0.920	0.5 ± 0.4	-0.7 ± 0.5
DH	0.29 ± 0.05	0.850 ± 0.016	0.915	0.4 ± 0.5	-0.5 ± 0.5
DH	0.28 ± 0.05	0.848 ± 0.016	0.921	0.3 ± 0.5	-0.5 ± 0.6
DH	0.35 ± 0.07	0.861 ± 0.017	0.918	0.9 ± 0.6	-0.0 ± 0.6
DH	0.31 ± 0.04	0.833 ± 0.012	0.916	0.6 ± 0.4	-1.1 ± 0.4
DH	0.33 ± 0.05	0.843 ± 0.013	0.925	0.8 ± 0.5	-0.7 ± 0.5
DH	0.31 ± 0.05	0.852 ± 0.014	0.925	0.6 ± 0.5	-0.4 ± 0.5
DH	0.31 ± 0.05	0.853 ± 0.015	0.930	0.6 ± 0.5	-0.4 ± 0.5
DH	0.203 ± 0.033	0.815 ± 0.015	0.907	-0.44 ± 0.32	-1.7 ± 0.5
DK	0.40 ± 0.07	0.859 ± 0.017	0.926	1.5 ± 0.7	-0.1 ± 0.6
DK	0.44 ± 0.08	0.868 ± 0.016	0.919	1.8 ± 0.8	0.2 ± 0.6
DK	0.196 ± 0.031	0.819 ± 0.014	0.916	-0.50 ± 0.30	-1.5 ± 0.5
DK	0.160 ± 0.026	0.798 ± 0.015	0.904	-0.85 ± 0.25	-2.3 ± 0.5
DK	0.30 ± 0.05	0.845 ± 0.014	0.924	0.5 ± 0.4	-0.6 ± 0.5
DK	0.23 ± 0.04	0.834 ± 0.014	0.908	-0.1 ± 0.4	-1.0 ± 0.5
DK	0.27 ± 0.05	0.848 ± 0.015	0.930	0.2 ± 0.4	-0.5 ± 0.5
DK	0.35 ± 0.07	0.860 ± 0.019	0.916	1.0 ± 0.7	-0.1 ± 0.7
DK	0.34 ± 0.05	0.835 ± 0.012	0.917	0.9 ± 0.5	-1.0 ± 0.4
DK	0.25 ± 0.04	0.831 ± 0.012	0.912	0.0 ± 0.4	-1.1 ± 0.4
DK	0.36 ± 0.06	0.858 ± 0.013	0.918	1.1 ± 0.5	-0.2 ± 0.5
DK	0.31 ± 0.06	0.851 ± 0.016	0.924	0.6 ± 0.6	-0.4 ± 0.6
DK	0.149 ± 0.022	0.799 ± 0.013	0.905	-0.96 ± 0.22	-2.2 ± 0.5

Table A.4 Taylor's parameters for individuals with different degrees of overweight and obesity [168]. Healthy people in this study, who were not obese, are characterized by $\bar{V} = 0.19 \pm 0.06$, $\bar{\beta} = 0.806 \pm 0.034$.

Metadata	V	β	\bar{R}^2	V_{st}	β_{st}
OW	0.59 ± 0.12	0.894 ± 0.034	0.920	6.6 ± 2.0	2.6 ± 1.0
OW	0.22 ± 0.04	0.830 ± 0.030	0.904	0.5 ± 0.6	0.7 ± 0.9
OBI	0.28 ± 0.04	0.855 ± 0.022	0.958	1.5 ± 0.6	1.4 ± 0.6
OBI	0.33 ± 0.07	0.870 ± 0.031	0.916	2.4 ± 1.1	1.9 ± 0.9
OBII	0.223 ± 0.032	0.823 ± 0.023	0.938	0.6 ± 0.5	0.5 ± 0.7
OBII	0.208 ± 0.029	0.844 ± 0.022	0.935	0.4 ± 0.5	1.1 ± 0.7
OBIII	0.34 ± 0.05	0.855 ± 0.025	0.943	2.5 ± 0.9	1.4 ± 0.7
OBIII	0.26 ± 0.04	0.845 ± 0.026	0.954	1.1 ± 0.7	1.2 ± 0.8
OBIII	0.33 ± 0.06	0.870 ± 0.027	0.908	2.4 ± 1.0	1.9 ± 0.8
OBIII	0.200 ± 0.026	0.843 ± 0.020	0.949	0.2 ± 0.4	1.1 ± 0.6
OBIII	0.30 ± 0.05	0.846 ± 0.026	0.929	1.9 ± 0.8	1.2 ± 0.7
OBIII	0.176 ± 0.029	0.826 ± 0.026	0.894	-0.2 ± 0.5	0.6 ± 0.8
OBIII	0.30 ± 0.06	0.841 ± 0.031	0.896	1.8 ± 0.9	1.0 ± 0.9
OBIII	0.28 ± 0.04	0.857 ± 0.025	0.941	1.5 ± 0.7	1.5 ± 0.7
OBIII	0.122 ± 0.018	0.822 ± 0.024	0.930	-1.05 ± 0.30	0.5 ± 0.7
OBIIIId	0.47 ± 0.08	0.872 ± 0.023	0.945	4.7 ± 1.3	1.9 ± 0.7
OBIIIId	0.38 ± 0.06	0.846 ± 0.023	0.951	3.2 ± 1.0	1.2 ± 0.7
OBIIIId	0.36 ± 0.06	0.842 ± 0.022	0.954	2.9 ± 0.9	1.1 ± 0.6

Table A.5 Rank and Rank Stability Index (RSI, as discussed in Material and Methods) over different periods for the taxa listed as *rank stability islands* regarding the gut microbiome of the subject A in the host lifestyle study [59].

Period	Genera												
	<i>Actinomyces</i>		<i>Leuconostoc</i>		<i>Lachnobacterium</i>		<i>Eggerthella</i>		<i>Clostridium</i>		<i>Collinsella</i>		
name	days	rank	RSI	rank	RSI	rank	RSI	rank	RSI	rank	RSI	rank	RSI
<i>before</i>	0 to 70	46	72.5	44	76.3	45	70.2	35	73.3	28	77.2	25	84.2
<i>abroad</i>	72 to 122	56	67.1	46	66.2	77	53.3	48	53.4	36	49.9	41	63.5
<i>returned</i>	123 to 256	44	79.3	41	69.5	31	74.2	33	77.5	34	71.6	27	81.0
<i>after</i>	257 to 364	43	79.0	39	72.2	33	68.4	30	78.5	34	76.7	26	80.4
Overall		47	76.4	43	71.0	36	69.2	35	74.1	34	70.7	28	79.5

A.1.2 Supplementary Figures

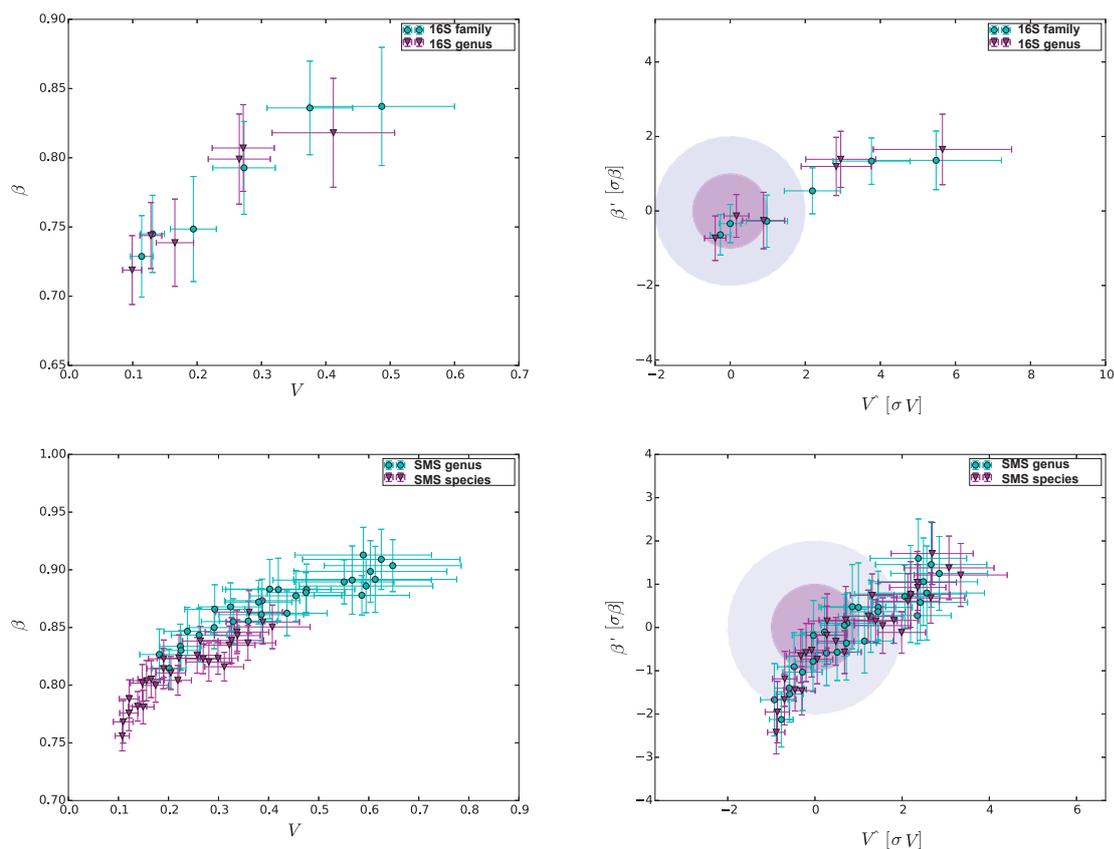


Fig. A.1 Overview of the comparison of different approaches based on adjacent taxonomic levels using plots in the Taylor-parameters space. The former row of subfigures is for 16S, where levels are family (blue circles) vs. genus (purple triangles), whereas the latter row of subfigures is for WGS, where levels are genus (blue circles) vs. species (purple triangles). The left column shows the raw results and the right column plots the standardized results.

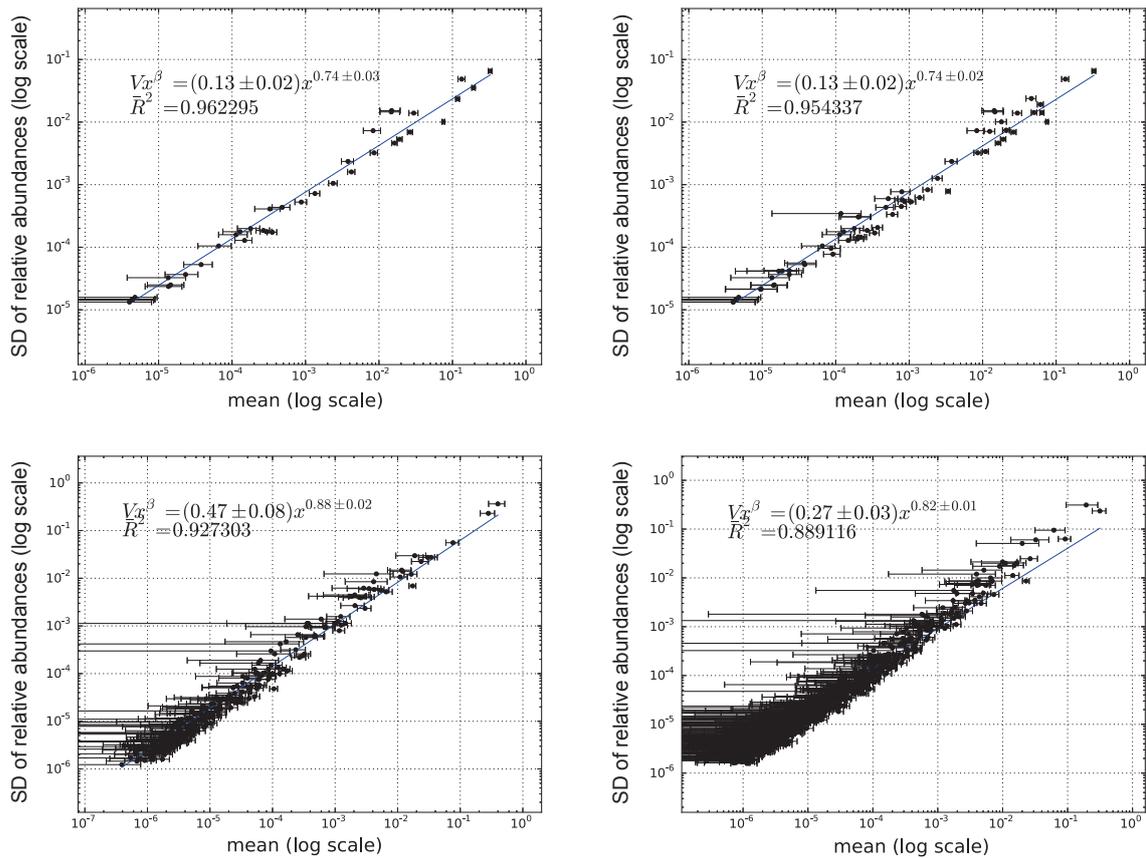


Fig. A.2 Detail of comparison of different approaches based on adjacent taxonomic levels using plots of X-weighted power-law fits (see Material and Methods). The former row of subfigures shows examples for 16S, whereas the latter row of subfigures plots examples for WGS. The left column shows results for the superior taxonomic level (family for 16S, genus for WGS), while the right column shows results for the inferior level (genus for 16S, specie for WGS).

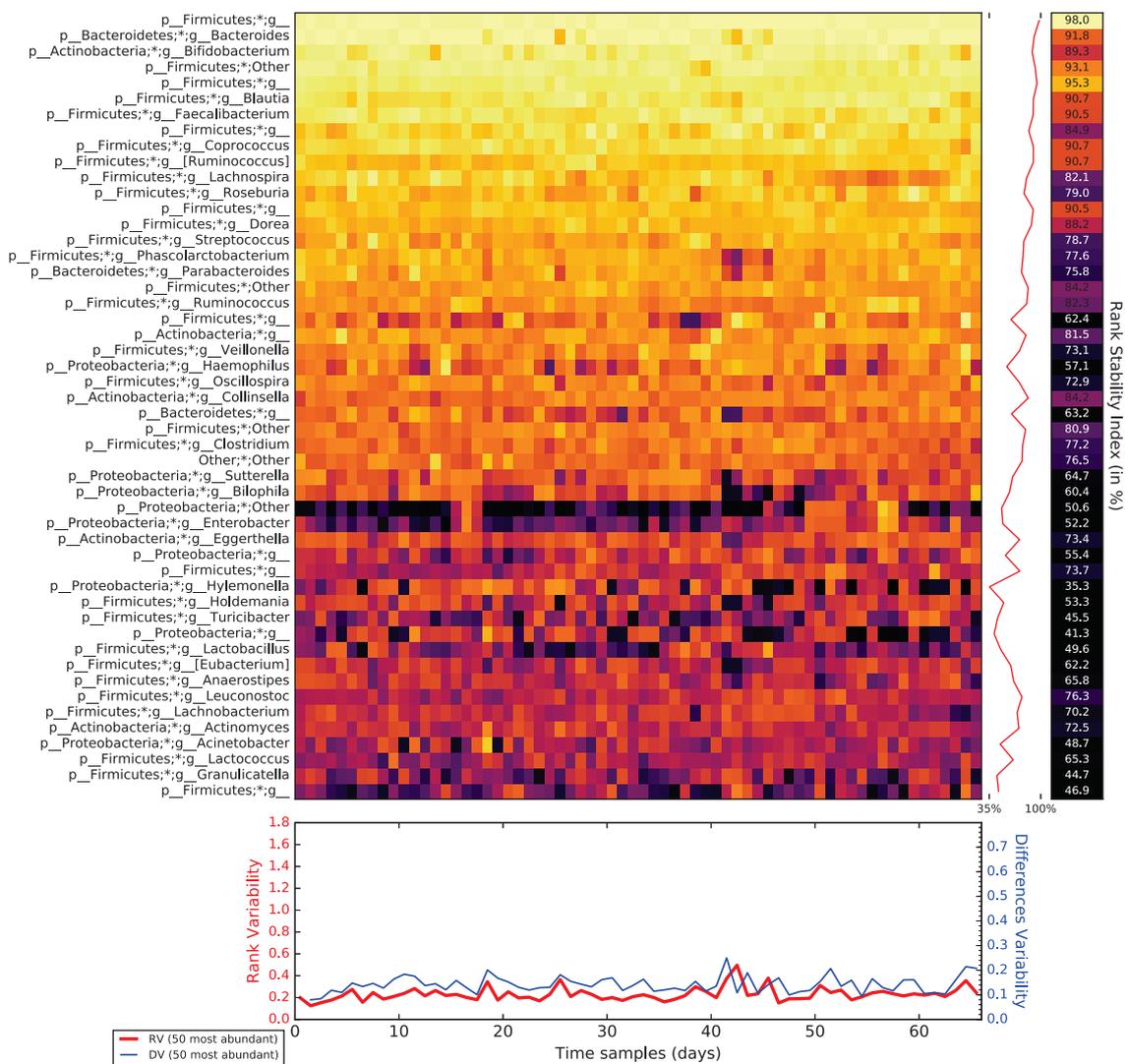


Fig. A.3 Rank variation over time for the 50 most dominant elements (taxa) and their calculated RSI (Rank Stability Index), Rank Variability (RV) and Differences Variability (DV), as detailed in Rank stability and variability in Material and Methods, for an ordinary period (days 0 to 70, before the trip) belonging to subject A in the host lifestyle study [59].

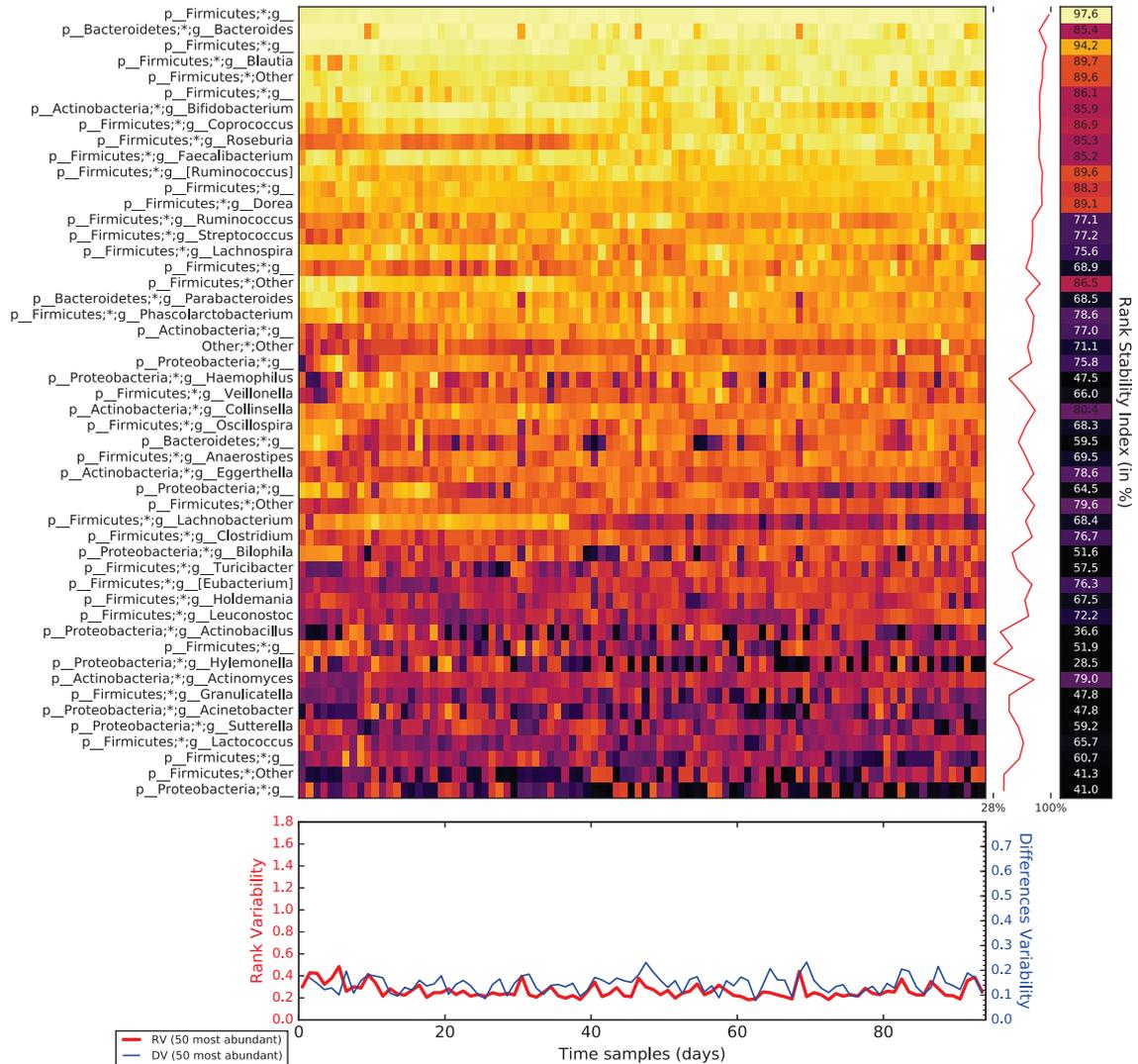


Fig. A.4 Rank variation over time for the 50 most dominant elements (taxa) and their calculated RSI (Rank Stability Index), Rank Variability (RV) and Differences Variability (DV), as detailed in Rank stability and variability in Material and Methods, for an ordinary period (days 257 to 364, further after the trip) belonging to subject A in the host lifestyle study [59].

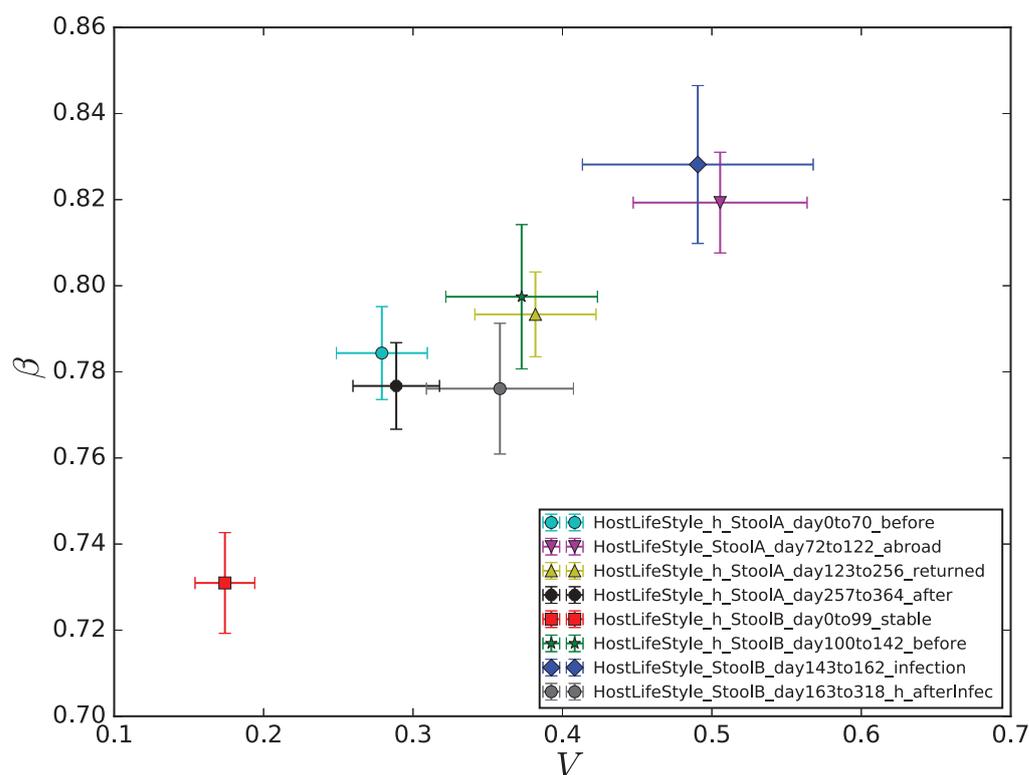


Fig. A.5 Taylor's law parameter space for intervals concerning gut microbiota in the host lifestyle study [59]. We observe that subject *B*, who suffered a *Salmonella* infection during the experiment, had a relevant shift in the parameters from *_before* to *_infection* and a final recovery from the perturbed state to *_afterinfect*, which lies in the parameter area compatible with the healthy and stable intervals (see Supplementary Table A.2). Subject *A* also had a shift in variability from *_before* to *_abroad* and back to *_returned*, also in the proximity zone of healthy and stable periods.

A.2 Chapter II - Supplementary information

A.2.1 Supplementary Figures

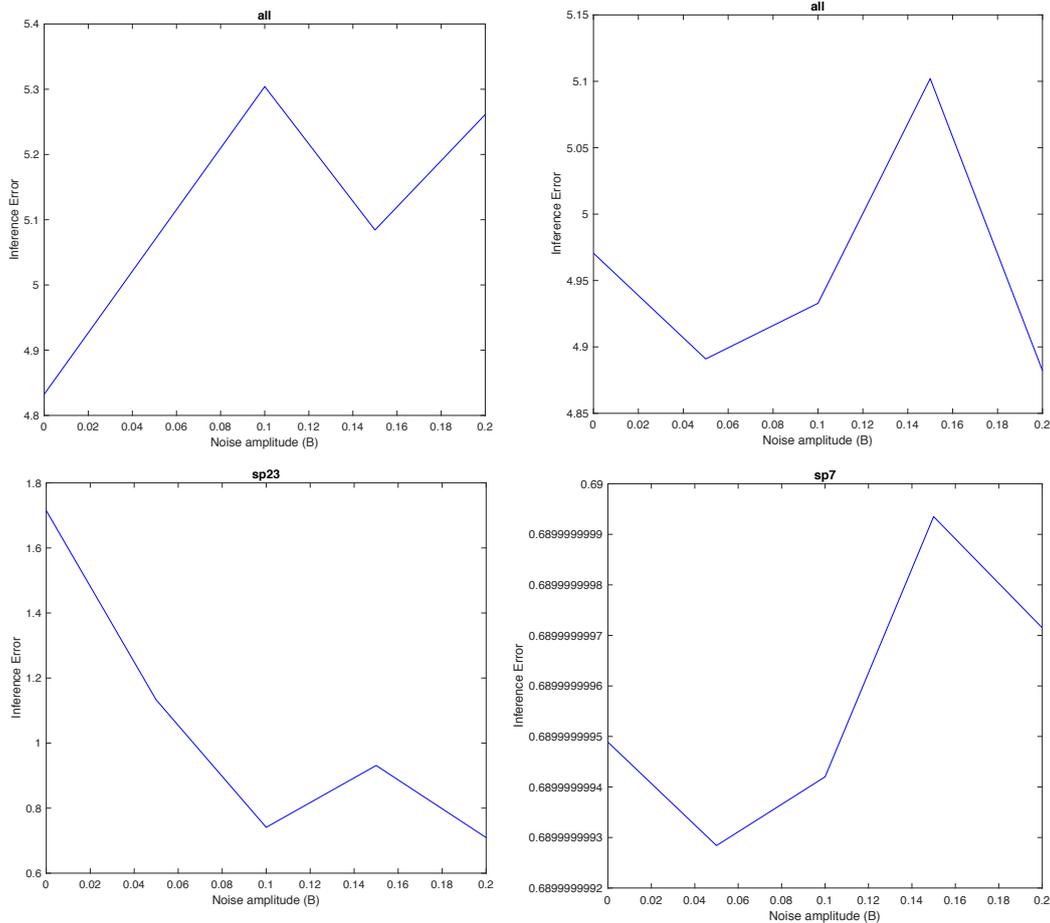


Fig. A.6 Error between original and inferred interaction matrix. In the upper row, it is represented the error between the entire matrices. On the left, the case when sp23 is modified, and on the right, the case when sp7 is modified. In the lower row, it is represented the error between the specific rows of sp23 (left), and sp7 (right) compared to the original interactions.

A.3 Chapter III - Supplementary information

A.3.1 Supplementary Tables

Table A.6 Metadata with children's hospitalisation information. Here it is shown the Patient, the age (in months), the weight (in kg), the fever at the hospitalisation start, the dehydration (1 = positive, 2 = negative), and the diarrhoea duration (in h).

Patient	Age	Weight	Fever	Dehydration	Diarrhoea duration
Case 1	39	14	37.9	1	72
Case 2	38	15.3	36	2	72
Case 3	37	14	36.5	1	48
Case 4	36	14	37	1	48
Case 5	36	16	36	1	72
Case 6	36	20	38	1	48
Case 7	36	18	36	1	72
Case 8	36	17	38	1	24
Case 9	36	14	36	2	72
Case 10	36	15	36.7	2	72

A.3.2 Supplementary Figures

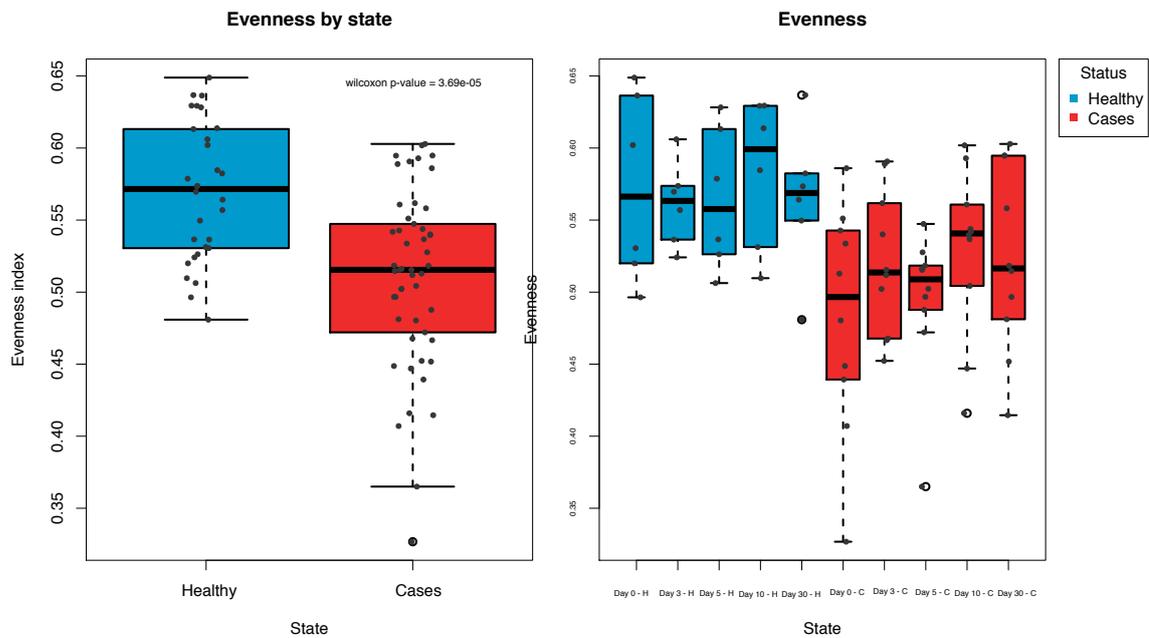


Fig. A.7 Boxplots showing the Evenness index between both health status with all times together; and separated by time of sampling. Blue color represent healthy children, and red color represent children with acute diarrhoea.

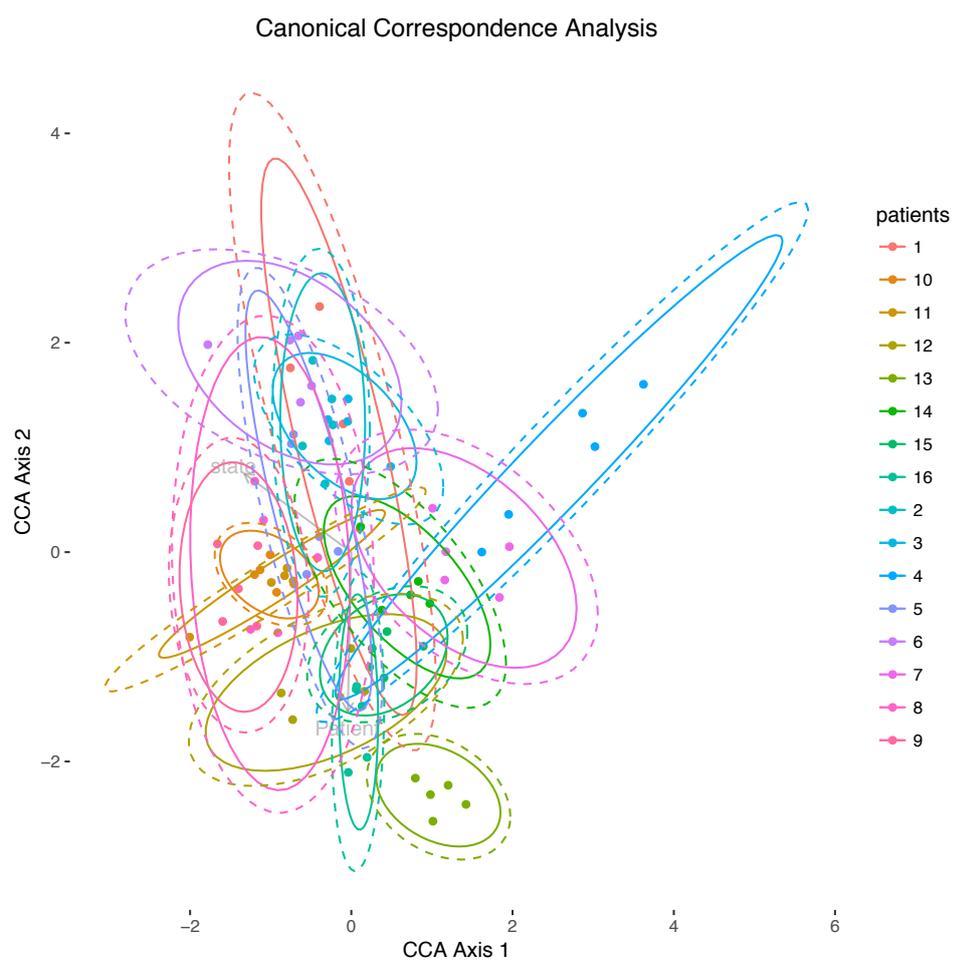


Fig. A.8 Canonical correspondence analysis of all individuals and times, separated by different child.

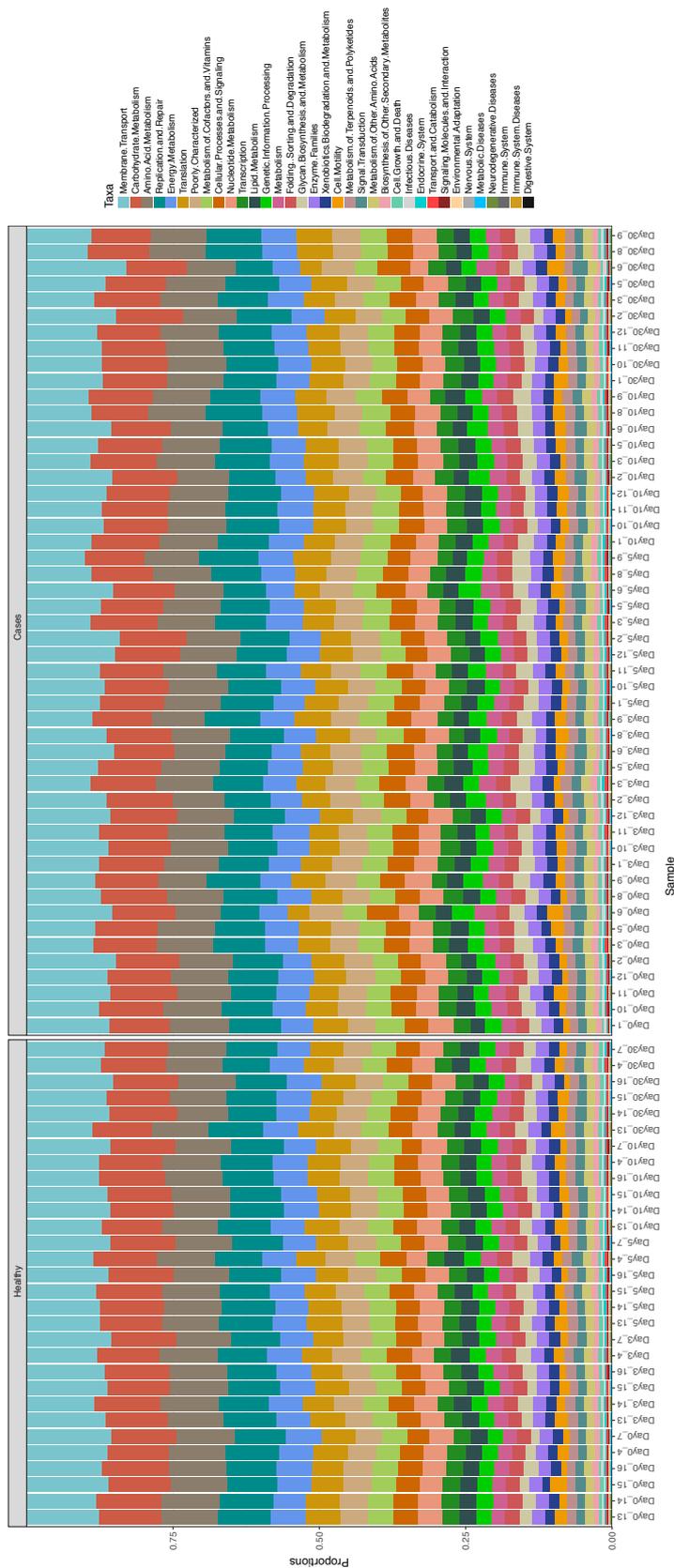


Fig. A.9 Relative abundances of the KEGG pathways inferred by PICRUST at level 2 of the hierarchy.

A.4 Chapter IV - Supplementary information

A.4.1 Supplementary Figures

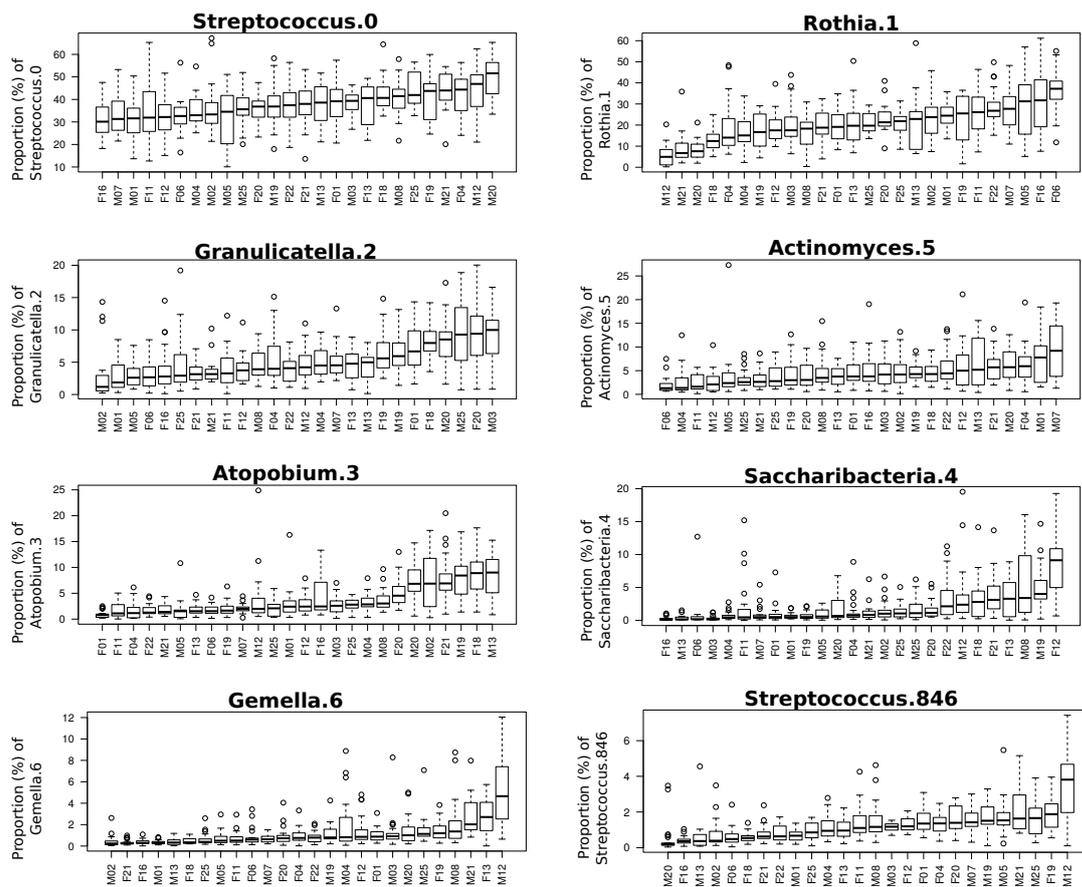


Fig. A.10 Composition of the salivary microbiome of the 26 volunteers. Boxplots show the proportion of the eight most prevalent bacterial OTUs, volunteers are ordered by the average proportion.

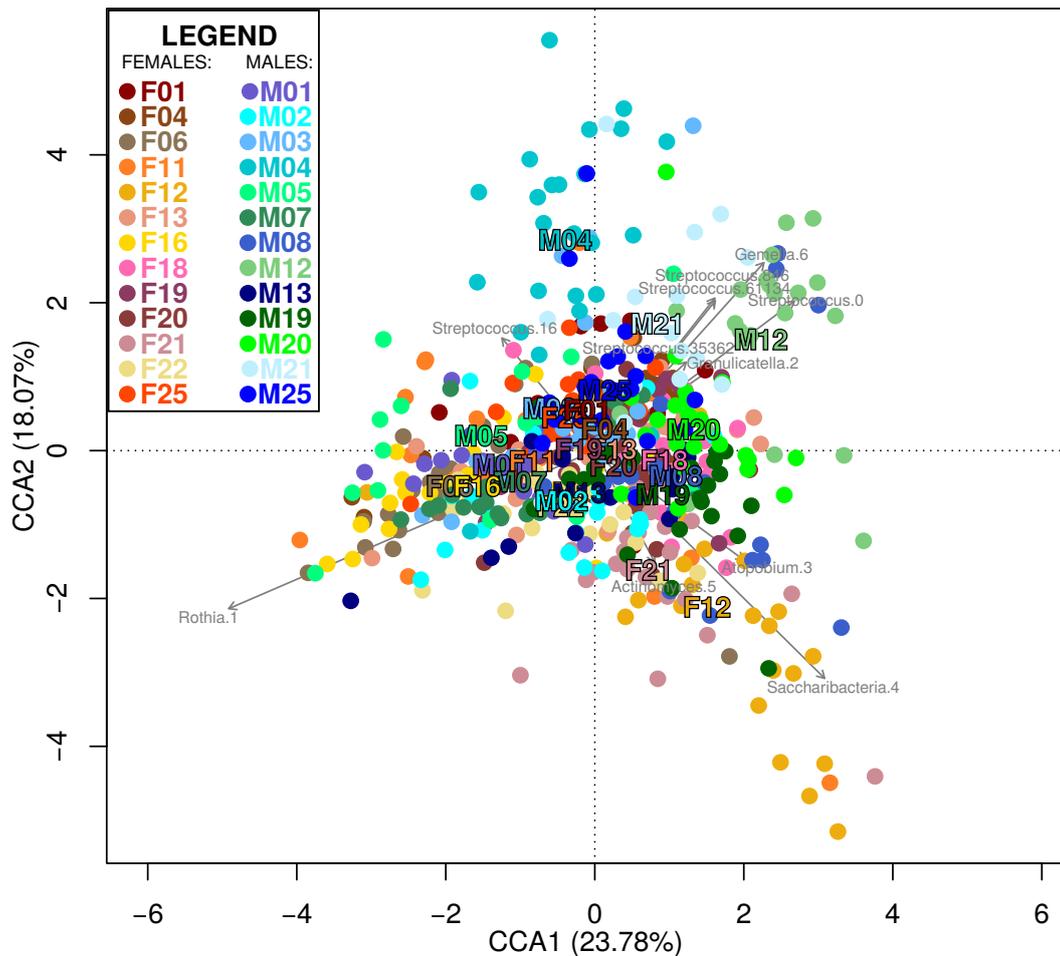


Fig. A.11 In the canonical correspondence analysis plot, the bacterial composition of the 573 samples was tested to assess the fit with the ‘volunteer individuality’ variable. All OTUs with average proportion $>0.001\%$ were included in the analysis. Samples from each volunteer are coloured as indicated the figure legend. The graphics shows only those of the 50 most prevalent OTUs which had significant influence ($p < 0.001$) on the ordination of samples.

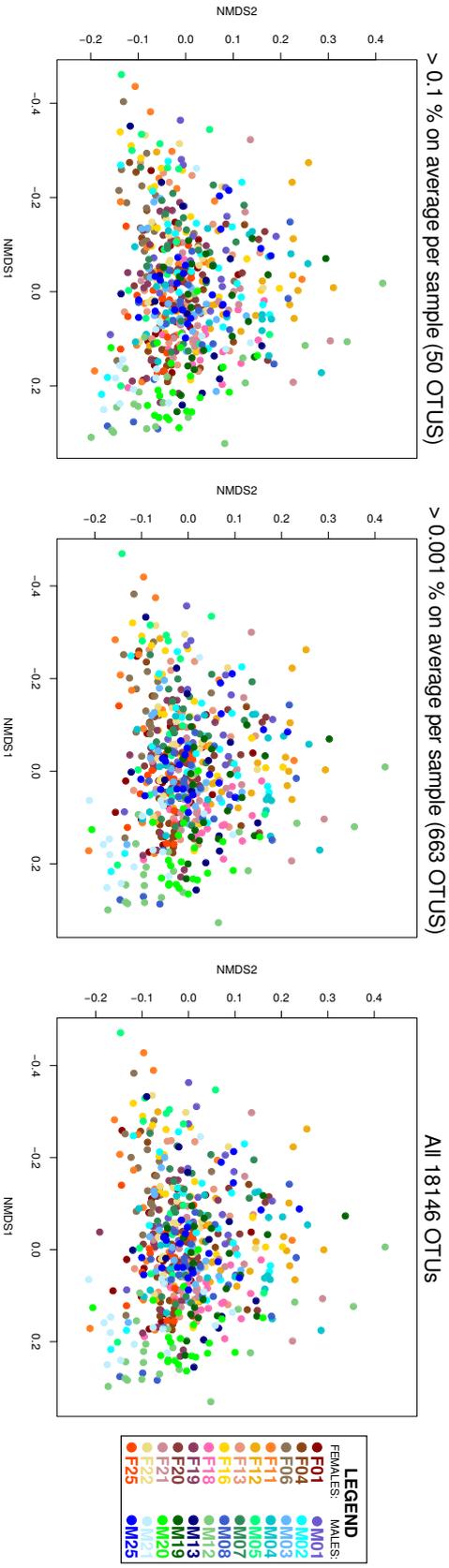


Fig. A.12 Comparison of NMDS ordinations of the 573 saliva samples including all 18,146 OTUs, OTUs with average proportion >0.001 % (687 OTUs) and with average proportion >0.1 % (50 OTUs). These three ordinations are not significantly different ($p < 0.001$).

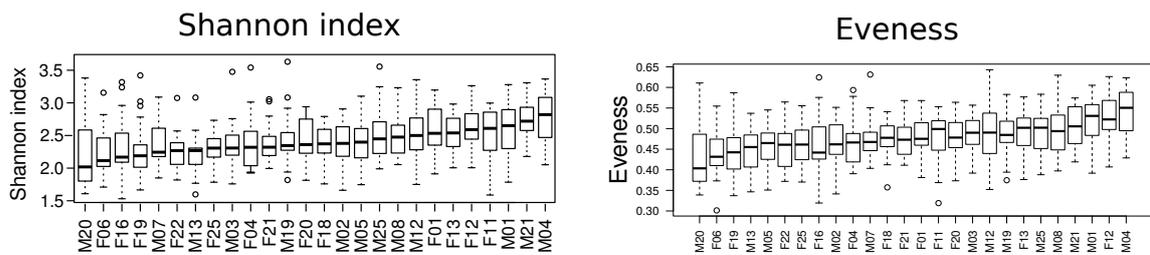


Fig. A.13 Boxplots of Shannon diversity index and evenness index for each of the volunteers, ordered from the lowest to the highest according to the average value.

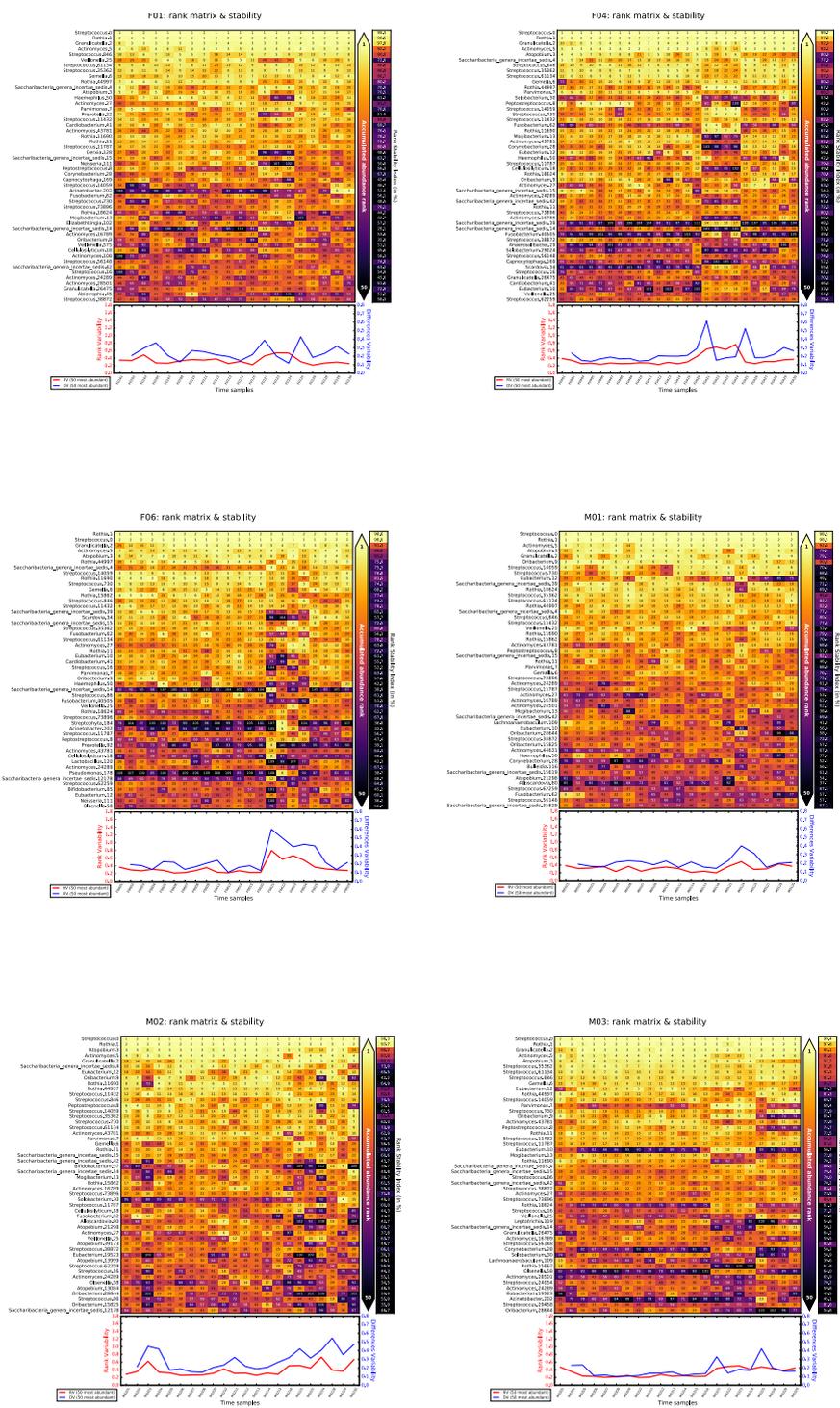


Fig. A.14 Temporal variability of 50 most prevalent OTUs in 6 volunteers. The volunteers F18 and F11 (the most and the less stable) are shown in the Figure 7.5

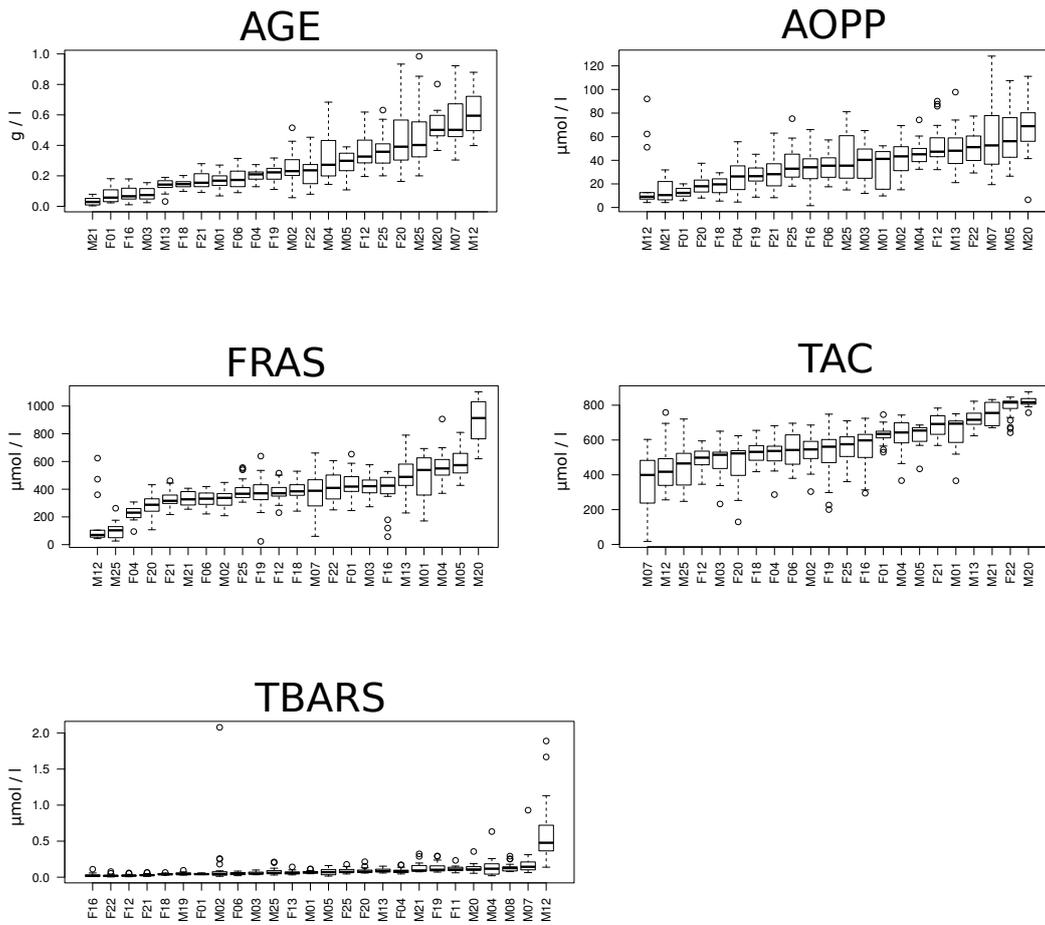


Fig. A.15 Boxplots showing values of AGE, AOPP, FRAS, TAC and TBARS for each volunteer, ordered from the lowest to the highest according to the average values.

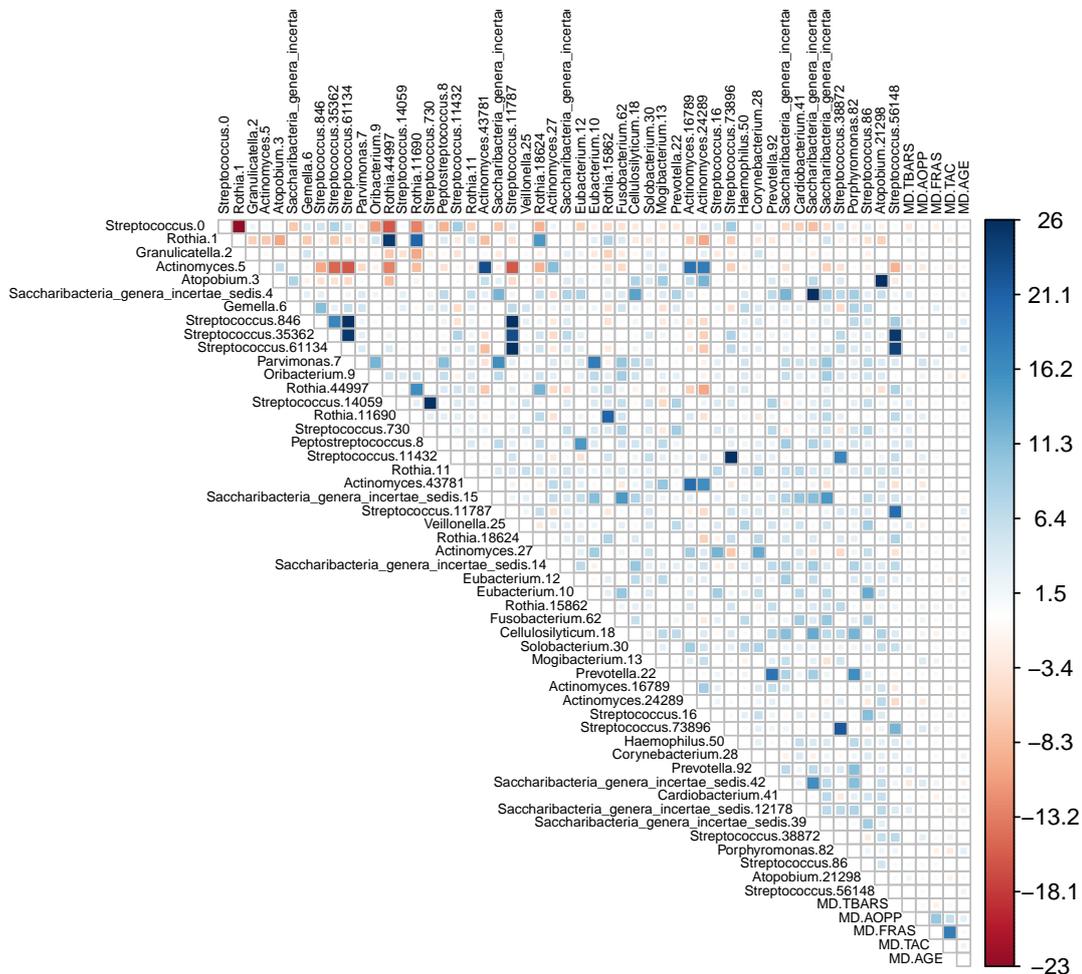
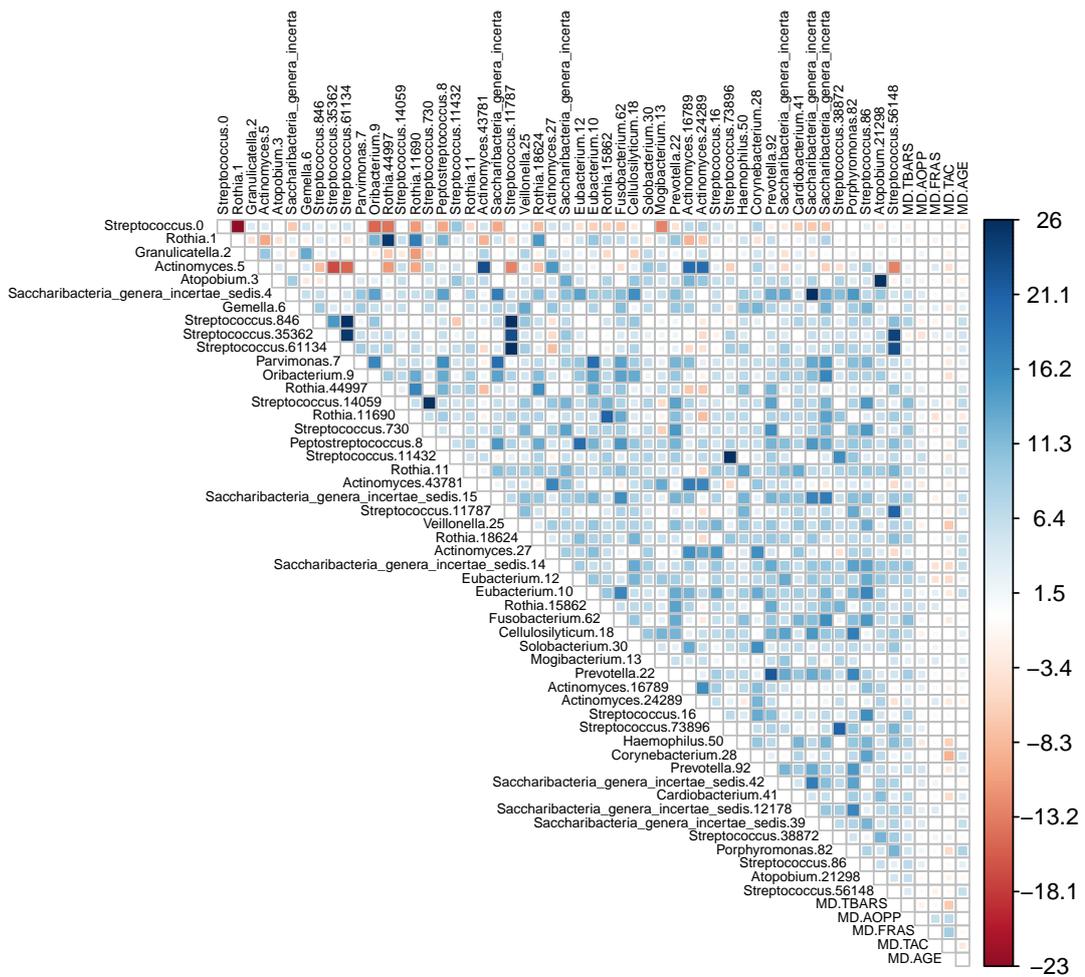


Fig. A.16 Correlation plots summarising Pearson’s and Spearman’s correlations on intra-individual level. The color bar on the right part of the panel explains the number of individuals with a negative (red) or positive (blue) correlation detected.



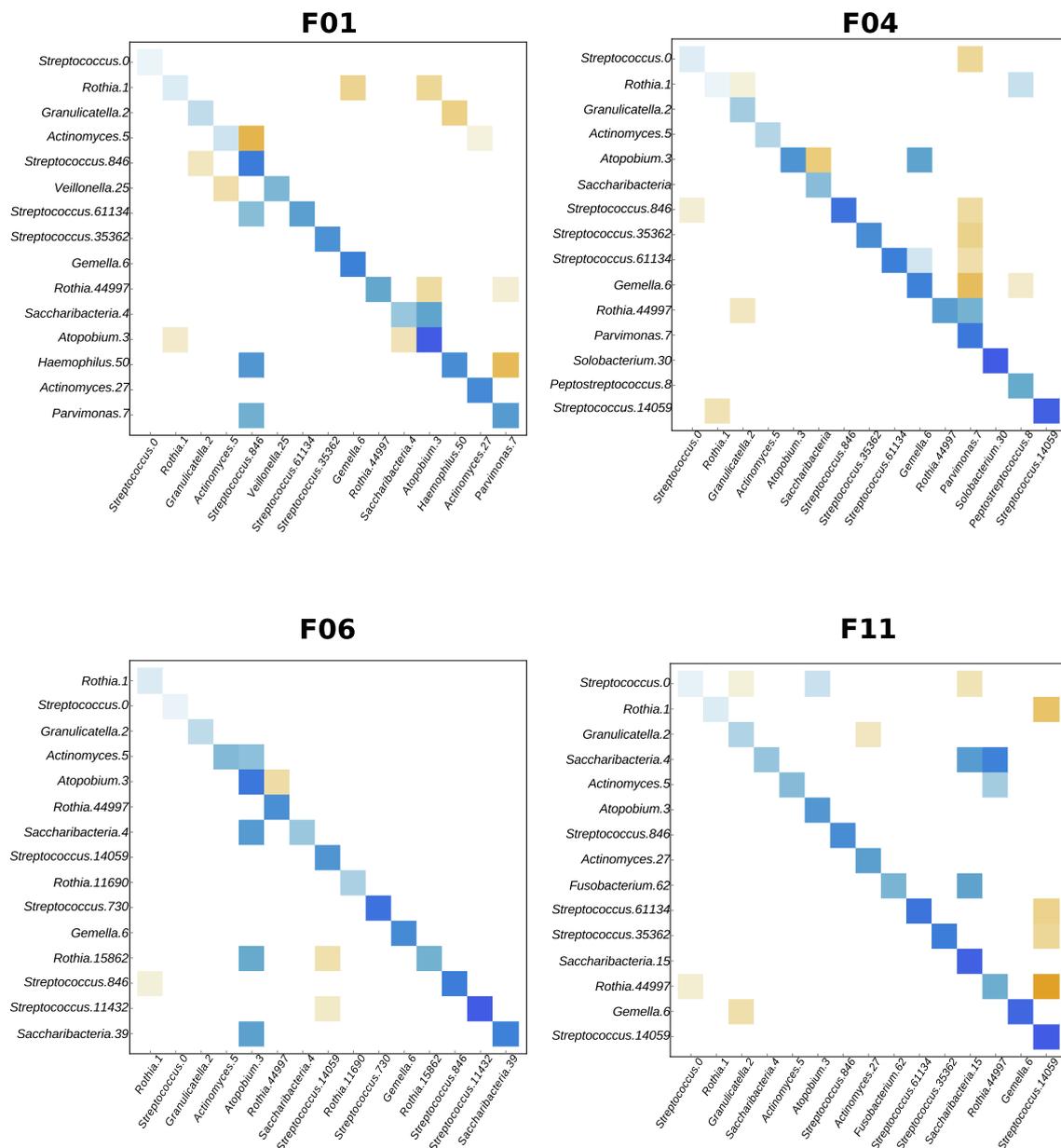
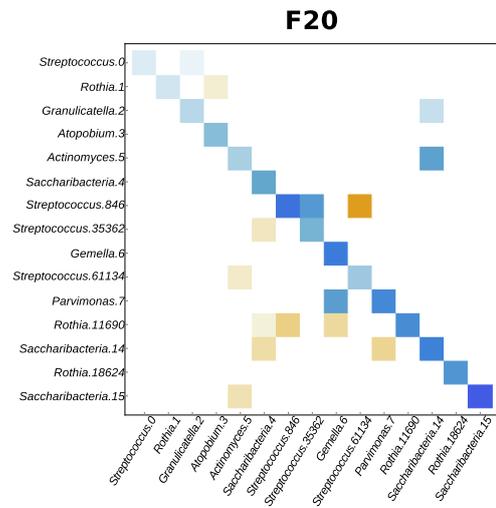
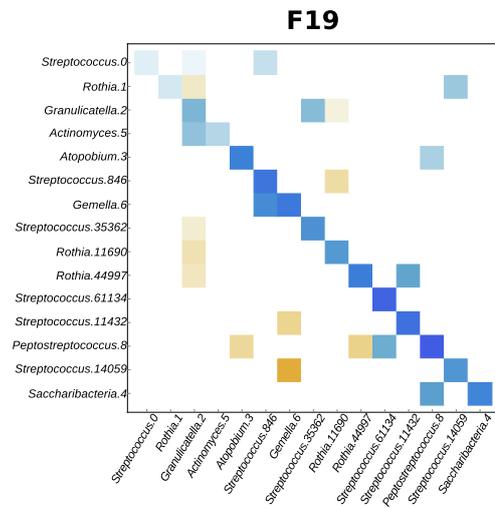
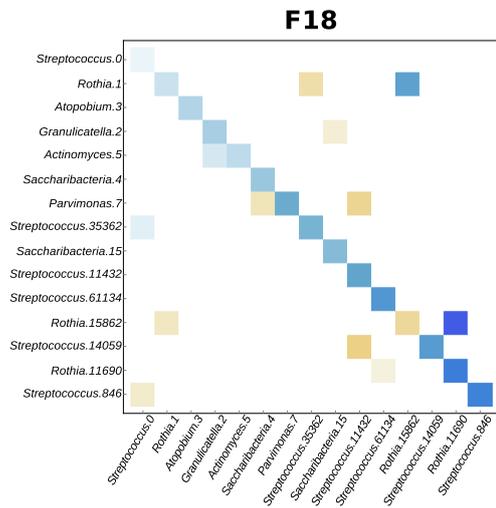
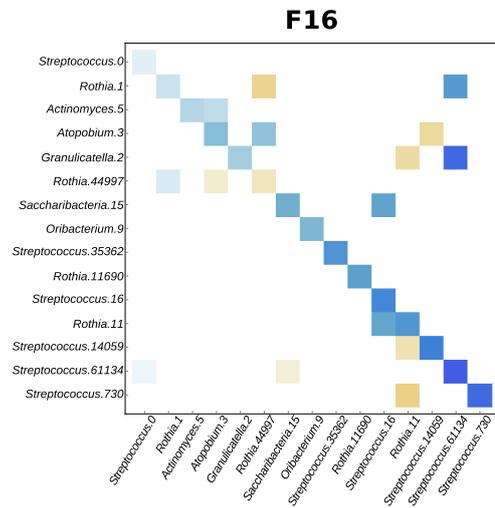
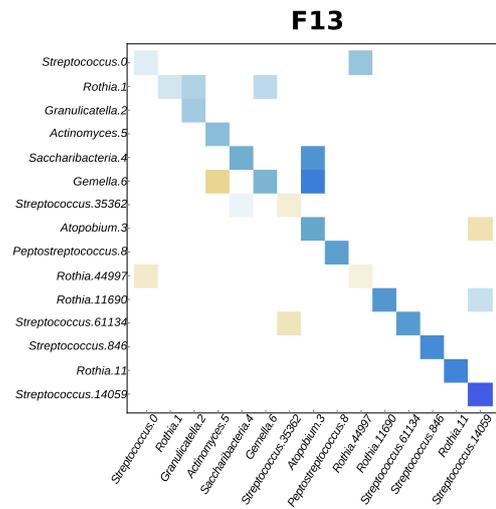
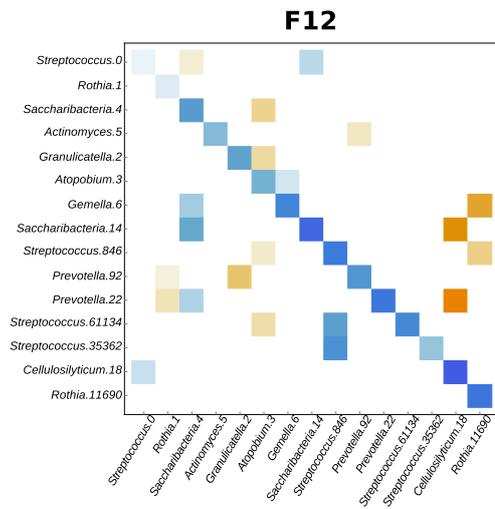
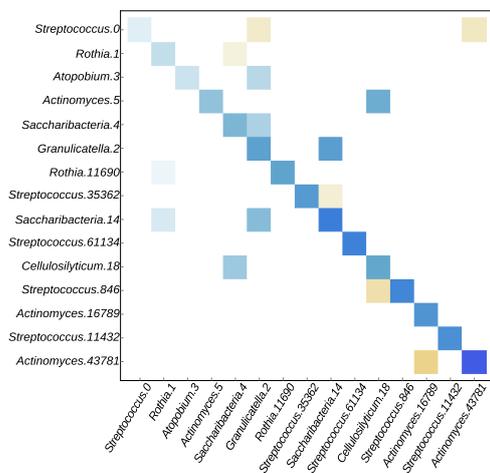


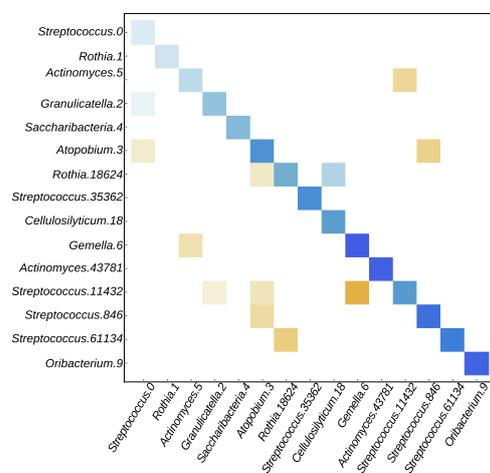
Fig. A.17 Microbial interactions between the 15 most abundant OTUs per sample. Blue color corresponds to negative interaction and orange corresponds to positive interaction. All interactions were analysed with LIMITS algorithm



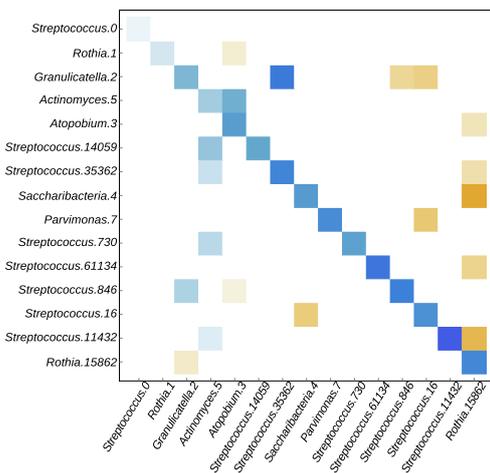
F21



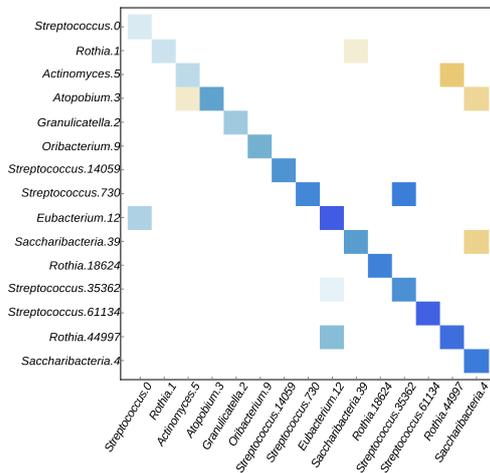
F22



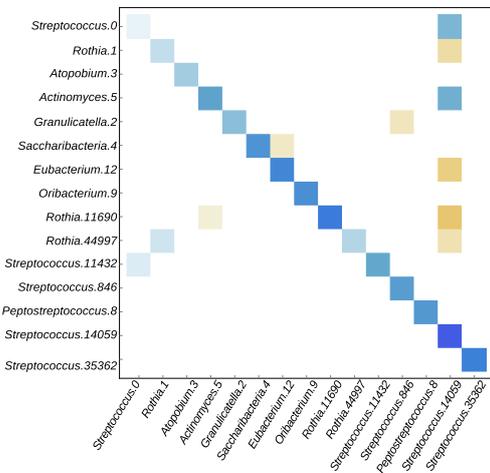
F25



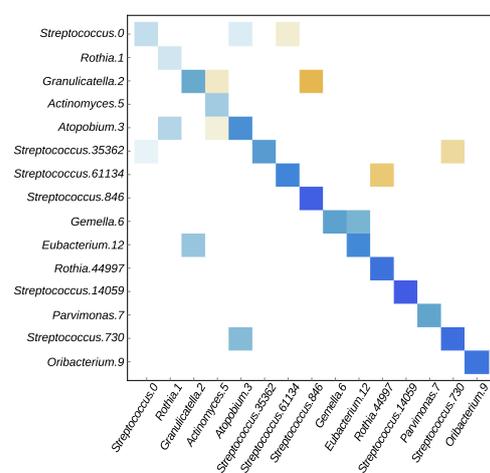
M01

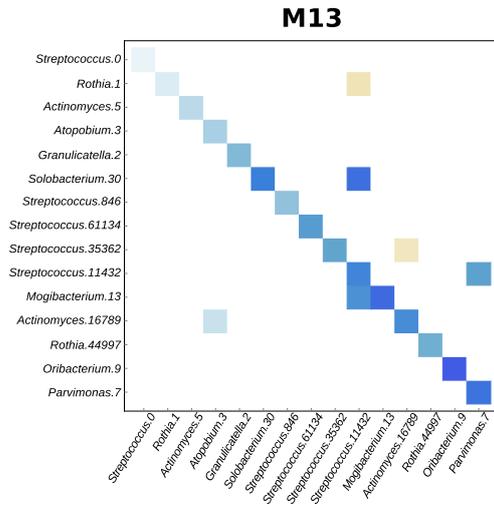
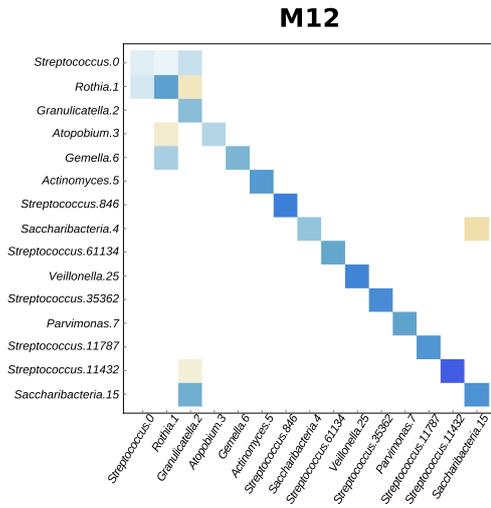
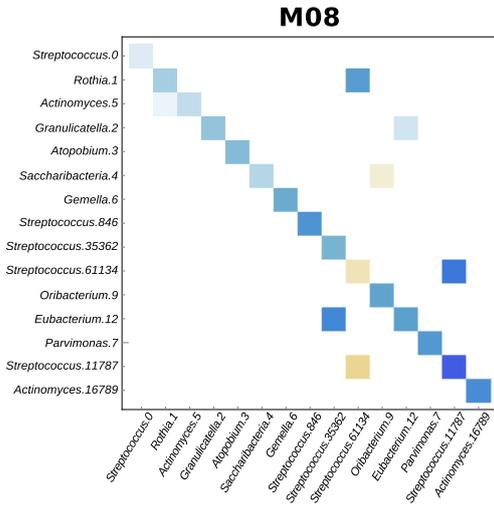
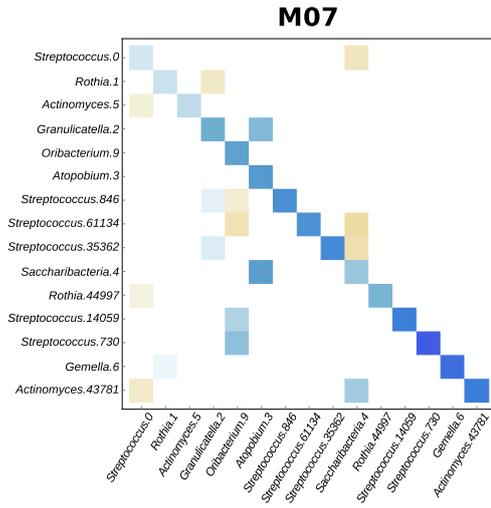
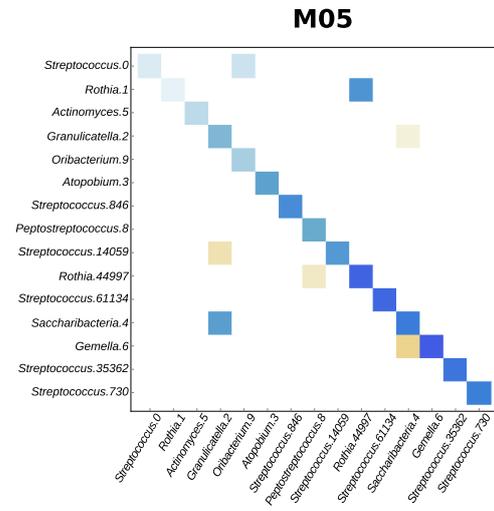
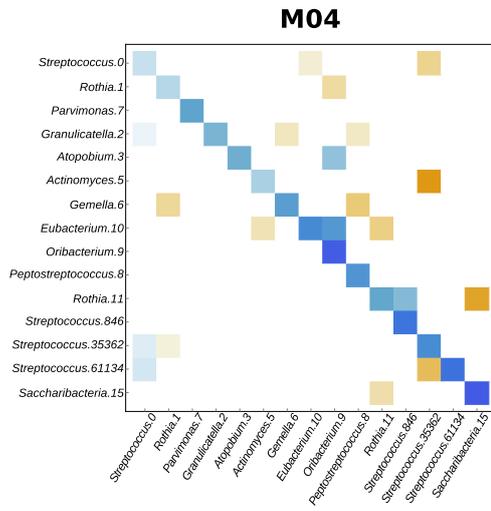


M02

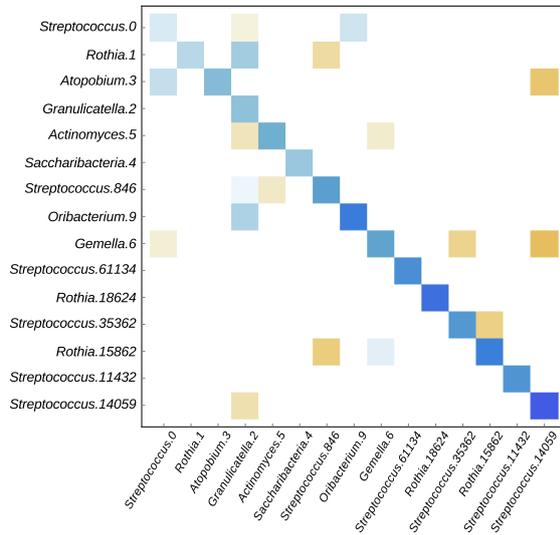


M03

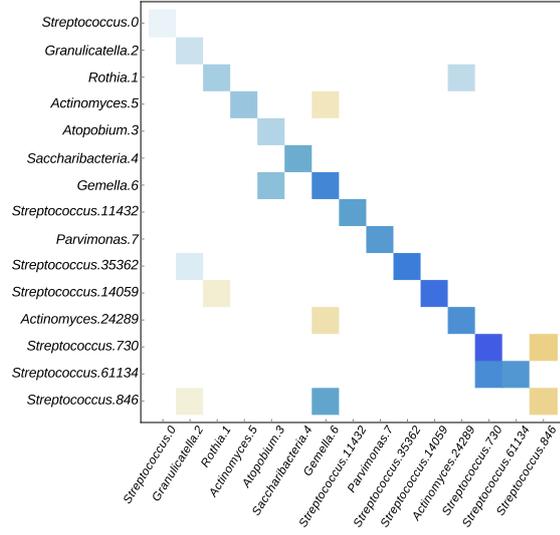




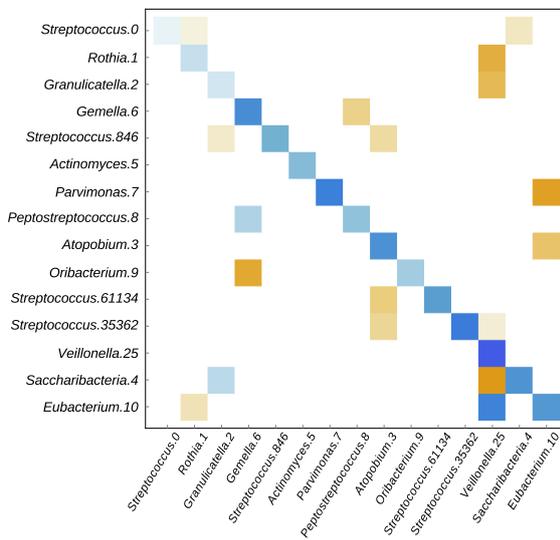
M19



M20



M21



M25

