Oral Microbiome: Composition, Interactions and Shaping Factors

Jesse Willis

TESI DOCTORAL UPF / 2021

Thesis supervisor:
Dr. Toni Gabaldón

Comparative Genomics group
Until March 2019: Bioinformatics Department
Center for Genomic Regulation
From April 2019: Life Sciences Department
Barcelona Supercomputing Center

Pompeu Fabra University
Department of Experimental and Health Sciences

For Raquel,
As we continue to level those lifts together

“*...And it didn’t stop being magic*
just because you found out how it was done”

- Terry Pratchett

This day before dawn I ascended a hill and look’d at the crowded heaven,
And I said to my spirit
When we become the enfolders of those orbs, and the pleasure and knowledge of every thing in them, shall we be fill’d and satisfied then?
And my spirit said
No, we but level that lift to pass and continue beyond.

- Walt Whitman

Acknowledgements

A glance at the citations of the papers that make up this thesis will show that I could not have accomplished any of it alone. The bioinformatics analyses that I performed were merely the final steps, so I must express my gratitude to the rest of the Saca La Lengua team that allowed it all to come together, especially those that I worked with directly. Annick and Elisabetta, you were the heralds of these projects, spreading the word to schools and institutions all across Spain and promoting interest from and participation by the public. I was very grateful that you both kept me informed and involved, knowing that I would be the eventual bottleneck for publishing the results and strove to make the process as smooth as possible. Ester, Ewa, and Susana, you had the skill to perform the wetlab experiments for which we drylabbers are too clumsy, always helped in crafting the manuscripts, and were central to our group's good times at dinners, post-dinners, retreats, meetings, fire drills, karaokes, etc. Alex, you introduced me to the wonders of R and guided my entry into microbiome studies, so I could hardly have made it this far without you, and you had a smoother jumpshot than most of the scientists I have gone up against. Pedro, thank you for helping to give some context to my analyses, and for always being available for a coffee or ping pong break, even while you were busy starting your own company. Olfat, thank you for bringing your desire, drive, and joy to the work every day - that makes for the best kind of student, one that requires very little teaching yet inspires the teacher to continue to learn.

Our group, Comparative Genomics, in its incarnations at both the Center for Genomic Regulation and the Barcelona Supercomputing

Center, has always been quite large. So unfortunately I limited name drops above to those members that were directly involved in my microbiome-related analyses. Still, I would be remiss not to express my gratitude toward all of the members past and present who, over the last seven years, have made it a joy to come to the lab … or to turn on my laptop camera. Scientific research can be mentally arduous at times, as each of you can surely attest, so it has been uplifting to be surrounded by colleagues and friends with whom I could both commiserate and celebrate.

Toni, thank you for coming to me time and again with different projects that needed analyzing, whether it was microbiome-related or otherwise. I was always interested in trying my hand at new types of data and tools for processing them, and you let me just about run the gamut of our group's various areas of study. Despite the size of our group, you have always been able to find the time to stay personally involved in everyone's projects with vital insights for analyses, as well as for the best ways to thrive in science, without letting any of us forget that we are not scientists for 24 hours a day, certainly not during those post-dinners.

To my mother, Melodi, I owe a great deal. You have always been a tireless worker, and made sure I worked hard myself, though never with undue pressure. You always let me be myself, helped me to become myself. You encouraged me, applauded my choices, even when they took me half a world away. Thank you for regularly sending me your bits of wisdom, and for being my first home; any time I come back it feels like I never left.

Raquel, my wife, my best friend, my muse - scientific and otherwise. I watched you go through this process a few years ago, helped in the little ways that I was able, and was impressed and inspired by your tenacity. I have tried to port that into my own work here. There have been some challenges, not that anyone at this stage

has ever said differently. During the pandemic, we, like so many others in basic research, have lost much of the boon that comes from working in a group — the camaraderie, the emotional and technical support from those sitting around us. But I, unlike many of those others, was lucky because you and I were able to form our own mini group from the safety of our home. I don't know how I would have managed without you there to motivate me, to push me, to brighten each day. So much of the work for this thesis was done from home, and without you I am homeless. Obviously, I cannot hope to match the message from the acknowledgements of your thesis — all I can do is humbly try to express how much you have meant to my own.

Abstract

Microbiome studies have burgeoned in the last few decades, largely thanks to the innovations in high throughput sequencing, affording researchers the opportunity to categorize nearly the entirety of the microbial population in a particular habitat. The oral cavity has been shown to harbor one of the most diverse and unique segments of the human microbiome, and one that has important implications in health and disease, as well as various links to lifestyle habits. In this thesis, we analyzed nearly 3000 oral rinse samples using 16S rRNA gene sequencing from a citizen science project called “Saca La Lengua” (“Stick Out Your Tongue” in English). The study design allowed our team to travel to high schools and other locations all across Spain to collect samples, as well as to disseminate information about the project to the public and to gather their input on interesting ways to analyze the data. Here, we first describe trends in the oral microbiome among youths in relative health, including general stable conformations of its composition and associations with drinking water and lifestyle. Then we focus on the connections between the oral microbiome and a few relevant chronic disorders, including Down Syndrome and cystic fibrosis. Finally, we display trends in abundances of specific taxa and in the overall composition across age, we compare the relative impacts of important health and lifestyle factors, and we highlight the importance of shared environments in shaping the oral microbiome. Taken together, this thesis provides some of the first snapshots of the oral microbiome across the Spanish population, revealing significant connections with oral and systemic health, as well as a multitude of lifestyle factors, ultimately pointing to its inherent ecological tendencies.

Resumen

En las últimas décadas se ha producido un aumento exponencial en el número de estudios sobre el microbioma humano. Este auge, debido en gran medida a los avances en las tecnologías de secuenciación masiva, ha permitido clasificar la mayoría de la población microbiana de cualquier muestra de microbioma. Estudios previos han demostrado que la cavidad oral alberga una de las partes más diversas y características del microbioma humano, han vinculado su composición con hábitos cotidianos y han revelado que tiene importantes implicaciones para la salud. En esta tesis, hemos analizado casi 3.000 muestras de la cavidad oral a través de la secuenciación del gen 16S ARN ribosómico, en un proyecto de ciencia ciudadana titulado “Saca La Lengua.” Primero, describimos las características generales del microbioma oral entre los jóvenes sanos, que presentan conformaciones estables en lo referente a su composición y establecemos asociaciones con la calidad del agua potable y estilo de vida. En segundo lugar, investigamos las conexiones entre el microbioma oral y algunas enfermedades crónicas, tales como el Síndrome de Down y fibrosis quística. Finalmente, examinamos cambios en las tendencias de la composición general y de las abundancias de particulares taxones con la edad, comparamos el impacto relativo de factores importantes para la salud y estilo de vida, y destacamos la importancia de entornos compartidos en la conformación del microbioma oral. En resumen, esta tesis proporciona por primera vez una visión global del microbioma oral de la población Española, revela conexiones significativas con la salud oral y la salud sistémica, y con una multitud de factores del estilo de vida, que en definitiva revelan tendencias ecológicas inherentes del microbioma.

Preface

"No man is an Iland, intire of itself; every man is a peece of the Continent, a part of the maine." That was John Donne in 1624 and his musings on the importance of human community and comradery still resonate today. But this was 40 years before Robert Hooke published *Micrographia*, the book of his observations of cells through a microscope, so Donne's anthropocentrism can perhaps be forgiven. Certainly I cannot speak against it. After all, while the focus of this thesis is the oral microbiome, it is more accurately the *human* oral microbiome. But I digress before I have begun; what I mean to suggest is that humans are more akin to islands than Donne could ever have conceived, physically separated from, but communicable to others of their kind, composed of a series of niches with distinct environmental conditions, each inhabited by bustling populations of microscopic denizens, who themselves might quote Donne with the same sentiment, given the interconnected nature of each segment of the human microbiome, merely swapping "man" for "microbe" and "island" for "organ" and "continent" for "(wo)man." The complexity and breadth of diversity of these microbial islanders make for an attractive challenge to the microbiomist, and the humidity and aeration unique to the oral cavity, as well as its primary or secondary communication to many other microbial habitats of the human island, make this particular site a fascinating destination.

Studies of the microbiome in their modern form, largely based on next generation sequencing techniques, are relatively new and burgeoning, and consensus definitions of precise terms, parameters, and proper practices are still sometimes debated. A recent retrospective review of the field, which surveyed leading microbiome researchers, attempted to provide a comprehensive description of the microbiome and to fill the gaps typical to the framework of its study (Berg et al. 2020). One thing that is clear is that, in addition to the microorganisms themselves, and their

interactions and byproducts, the surrounding environment also forms an integral part of the microbiome, simultaneously contributing to its composition and being manipulated by its activities. In the context of the human oral microbiome, the immediate habitat is the oral cavity, particularly the mucosal surfaces on which the microorganisms reside, like the gums or the tongue. This environment has a set of conditions (like the climate of any terrestrial biome), including the salivary pH or the host's body temperature. However, the mouth is highly vascularized (Naumova et al. 2013) and the gateway for ingested materials and occasionally for respiration, and therefore is subject to, and potentially implicated in, systemic changes across the human body (Willis and Gabaldón 2020).

In this thesis, I explore our current understanding of the symbiosis between the oral microbiome and its human host, primarily through the lens of a large-scale, citizen science-based project called "Saca La Lengua" (Spanish for "stick out your tongue", http://www.sacalalengua.org). In two separate editions (abbreviated SLL1 and SLL2), spaced two years apart, we went around Spain collecting thousands of oral rinse samples to sequence and study their microbiomes, and this large and diverse sample set allowed us to explore many of the factors affecting the oral microbiome, like age, geographical location, drinking water composition, chronic disorders, lifestyle, and shared environments. The results from the studies we performed with this data together provide what we feel is a strong foundation for moving toward an understanding of eubiosis in the oral microbiome of the Spanish population, as well as what may constitute dysbiosis in that group.

But first, I begin this thesis with an overall introduction to the field of microbiome studies, and to the oral microbiome in particular, in **Chapter 1**. This chapter was published as a review in early 2020 titled "The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems" in the journal *Microorganisms*. Then I present the four publications that came from Saca La Lengua thus far. The first from SLL1 comprises **Chapter 2**, and focuses on teenagers in relative health. The latter three are based on the SLL2

dataset, which included a wide range of ages and a focus on different lifestyle factors, as well as particular chronic disorders. **Chapter 3** presents the results regarding the subset of samples from individuals with Down Syndrome compared to matched controls, and **Chapter 4** does the same for cystic fibrosis. **Chapter 5** is based on the full SLL2 dataset, with analyses of differences in the oral microbiome across age among participants ranging from 13 to 85 years old, comparisons of chronic disorders and different lifestyle factors, as well as the impact of shared environments on the composition of the oral microbiome. These chapters have been published in *Microbiome* (**Chapter 2**), and *Journal of Oral Microbiology* (**Chapters 3** and **4**), or have been submitted for publication (**Chapter 5**). Finally, in **Chapter 6,** I provide an overarching discussion and synthesis of the results in the context of the relevant literature. I start by describing our field's current conception of the nature of the oral microbiome, as well as the progression of the methodologies that reveal it to us, including within my own work. From there, I move into what we understand as dysbiosis, how we can distinguish it from eubiosis, and whether we can rightfully claim to recognize real eubiosis (this is clearly a leading statement that underscores my skepticism, but also belies my optimism for the future of the field, which remains bright). This leads into my **Conclusions** on where my work fits into the field of microbiome studies, how I believe the field will continue to evolve (and the ways that I hope it will evolve), and how the ever increasing knowledge might be applied to other areas of health and science.

Objectives

This thesis aimed to describe the composition of the oral microbiome and the factors which shape it in the Spanish population. To that end, we defined the following objectives:

- Produce a cartography of the oral microbiome composition in the Spanish population, with information on how it varies according to geographical, age, lifestyle and health factors.
- Explore the correlative relationships between microbiome composition and other types of collected data that may reveal the influences of various health and lifestyle factors on the oral microbiome.
- Evaluate the effects of additional externally sourced metadata that may point to effects on the oral microbiome at a societal scale, as opposed to the individual scale which applies to much of the internally collected metadata. This includes information on the ionic composition of public drinking water throughout Spain.
- Explore the presence of clusters of samples based on their overall oral microbiome compositions, and assess their impact and relevance in the context of the available metadata.
- Analyze the differences in the oral microbiome between specifically targeted disease cohorts and control samples in relative health. In particular, Down Syndrome and cystic fibrosis.
- Track the differences in the oral microbiome across a wide range of ages, from children to senior citizens.

Table of Contents

Acknowledgements......vii
Abstract......xi
Resumen......xiii
Preface......xv
Objectives......xix

Chapter 1: The human oral microbiome in health and disease: from sequences to ecosystems......1
1.1 Abstract......3
1.2 Introduction......4
1.3 Technical approaches to study the oral microbiome......10
1.4 The oral cavity and its microbial niches......18
1.5 The healthy oral microbiome and definition of stomatotypes......22
1.6 Non-bacterial oral microbes......29
1.7 Oral microbiome and oral diseases......31
1.8 Oral microbiome and non-oral diseases......34
1.9 Clinical potentials of the oral microbiome / Manipulations and perturbations of the oral microbiome......39
1.10 Conclusions and future outlook......42

Chapter 2: Citizen science charts two major "stomatotypes" in the oral microbiome of adolescents and reveals links with habits and drinking water composition......47
2.1 Abstract......49
2.2 Introduction......50
2.3 Material and methods......52
2.3.1 Sample collection......52
2.3.2 DNA extraction and sequencing......53
2.3.3 Pre-processing of 16S rRNA sequence reads and operational taxonomic unit assignment......55
2.3.4 Microbiome composition profiling......56
2.3.5 Diversity measures......56
2.3.6 Sample clustering......57
2.3.7 Gradients of abundances......58

2.3.8 Co-occurrence networks....58
2.3.9 Questionnaire and other metadata....59
2.3.10 Statistical analyses....60
2.3.11 Distribution maps....60
2.4 Results....61
2.4.1 Data collection and analysis....61
2.4.2 Oral microbiome diversity is structured into two major stomatotypes....62
2.4.3 Lifestyle and social parameters....70
2.4.4 Hygiene habits and saliva pH....72
2.4.5 Tap water influences the oral microbiome....74
2.5 Discussion....77
2.6 Conclusions....81

Chapter 3: Oral microbiome in Down Syndrome and its implications on oral health....85
3.1 Abstract....87
3.2 Introduction....88
3.2 Materials and methods....91
3.2.1 Sample collection....91
3.2.2 DNA extraction and sequencing....92
3.2.3 Fungal composition analysis....94
3.2.4 Pre-processing of 16S rRNA sequence reads and taxonomy assignment....98
3.2.5 Diversity measures....99
3.2.6 Statistical analyses....99
3.3 Results....101
3.3.1 Increased abundance of periodontal pathogens in DS 101
3.3.2 Lower pH and species diversity in DS....103
3.3.3 Opportunistic pathogenic *Candida* species were more prevalent in DS....105
3.3.4 No evidence for premature aging in the oral microbiome of DS....105
3.4 Discussion....109
3.5 Conclusions....118

Chapter 4: Citizen-science based study of the oral microbiome

in cystic fibrosis and matched controls reveals major differences in diversity and abundance of bacterial and fungal species..121
4.1 Abstract..123
4.2 Introduction..124
4.3 Materials and methods...127
4.3.1 Sample collection..127
4.3.2 DNA extraction and sequencing...............................128
4.3.3 Fungal composition analysis....................................129
4.3.4 Pre-processing of 16S rRNA sequence reads and taxonomy assignment...131
4.3.5 Diversity measures..132
4.3.6 Statistical analyses..133
4.3.7 Inferred co-occurrence networks..............................134
4.4 Results..134
4.4.1 Increased abundances of airway pathogens and decreased abundances of some periodontal pathogens in CF 134
4.4.2 Candida albicans *more prevalent in CF*....................139
4.4.3 Lower alpha diversity and salivary pH in CF............141
4.4.4 Networks of co-occurring taxa differ significantly between CF and controls..141
4.5 Discussion...145
4.6 Conclusions...153

Chapter 5: Citizen-science study of the oral microbiome reveals changes through age and the relative impacts of diverse biological, social and lifestyle factors.......................155
5.1 Abstract..157
5.2 Background...158
5.3 Materials and Methods...161
5.3.1 Sample collection..161
5.3.2 Subsampling for analyses..162
5.3.3 Statistical analyses..163
5.3.3.1 Comparisons of compositions...........................163
5.3.3.2 Differential abundance and diversity calculations 164
5.3.3.3 Inferred co-occurrence networks.......................165

5.4 Results..166
5.4.1 Oral microbiome changes through age....................166
5.4.2 Chronic disorders, smoking and the presence of yeasts in the oral cavity, are important drivers of the oral microbiome composition..171
5.4.3 Similarity of the oral microbiome composition among family members and classmates.......................................178
5.5 Discussion..180
5.5.1 Oral microbiome changes through age....................180
5.5.2 Chronic disorders, smoking and the presence of yeasts in the oral cavity, are important drivers of the oral microbiome composition..183
5.5.3 Similarity of the oral microbiome composition among family members and classmates.......................................185
5.6 Conclusions..186

Chapter 6: Summarizing Discussion.....................................195
6.1 Lessons from Saca La Lengua....................................197
6.1.1 SLL1 - Drinking water, stomatotypes, and the “core” oral microbiome..199
6.1.2 SLL2 - Chronic disorders and the oral microbiome 202
6.1.3 SLL2 - Aging in the oral microbiome....................205
6.1.4 SLL2 - The non-bacterial segment of the microbiome 207
6.2 Notes on Methodologies...209
6.3 The Being and the Becoming of Microbiomics.............214
6.3.1 Reductionism in Microbiome studies....................215
6.3.2 Panta rhei..219
6.3.3 The Evolving Microbiome Simulacrum..................222
6.4 The conceit of SLL...228

Conclusions..231

Appendix I: Drinking water impacts the oral microbiome composition..233

Bibliography..237

Chapter 1: The human oral microbiome in health and disease: from sequences to ecosystems

Willis JR & Gabaldón T. (2020). The human oral microbiome in health and disease: from sequences to ecosystems. *Microorganisms*. 8(2). doi: 10.3390/microorganisms8020308.

Chapter 1: The human oral microbiome in health and disease: from sequences to ecosystems

1.1 Abstract

The human oral cavity is home to an abundant and diverse microbial community (i.e., the oral microbiome), whose composition and roles in health and disease have been the focus of intense research in recent years. Thanks to developments in sequencing-based approaches, such as 16S ribosomal RNA metabarcoding, whole metagenome shotgun sequencing, or meta-transcriptomics, we now can efficiently explore the diversity and roles of oral microbes, even if unculturable. Recent sequencing-based studies have charted oral ecosystems and how they change due to lifestyle or disease conditions. As studies progress, there is increasing evidence of an important role of the oral microbiome in diverse health conditions, which are not limited to diseases of the oral cavity. This, in turn, opens new avenues for microbiome-based diagnostics and therapeutics that benefit from the easy accessibility of the oral cavity for microbiome monitoring and manipulation. Yet, many challenges remain ahead. In this review, we survey the main sequencing-based methodologies that are currently used to explore the oral microbiome and highlight major findings enabled by these approaches. Finally, we discuss future prospects in the field.

Keywords: Oral microbiome, next generation sequencing, oral diseases, systemic diseases, stomatotypes, microbiome perturbations

1.2 Introduction

Much like the various terrestrial biomes that make up the Earth, the human microbiome is a series of distinct communities of bacteria, fungi, viruses, archaea, protists, and other microorganisms, whose compositions are dependent upon environmental conditions (Ursell et al. 2012). Different sites of the human body can be seen as unique biomes, with drastically different environments and nutrient availabilities, which in turn promote different communities. Yet even within a particular body site, the microbiome composition can be highly variable between individuals in different states of health, with distinct lifestyles, or due to a number of other factors (Integrative HMP (iHMP) Research Network Consortium 2014). The focus of this review will be the human oral microbiome, techniques to approaching its analysis, and outlining its typical composition as we currently know it, as well as its deviations under atypical conditions.

The oral cavity contains one of the most diverse and unique communities of microbes in the human body (Human Microbiome Project Consortium 2012; K. Li et al. 2012), yet this niche is relatively understudied as compared to the gut—at the time of writing this review, a PubMed search with "oral microbiome" resulted in 746 articles, as compared to 5605 with "gut microbiome". A milliliter of saliva contains approximately 10^8 microbial cells (Philip D. Marsh et al. 2016), and an array of studies have detected up to 700 distinct prokaryotic taxa (Floyd E. Dewhirst et al. 2010), with a typical healthy microbiome comprised of a range of about 100 to 200 distinct bacterial organisms (Paster et al. 2006). The advent of next generation sequencing (NGS) techniques has opened new avenues for large-scale metagenomic studies in diverse populations, allowing for characterization of the microbiome structure and, in some cases, the functional roles and implications for health.

The mouth as a biome is home to multiple unique habitats, each of which has its own community of microorganisms. The microbiomes of the saliva, tongue, buccal mucosa, teeth surfaces, gums, palate,

both subgingival and supragingival plaque, as well as the throat and tonsils, have all been characterized in multiple studies, showing overall similarities but with small-scale differences, such as higher levels of the genus *Corynebacterium* in both types of plaque (Human Microbiome Project Consortium 2012; Segata et al. 2012) or higher levels of the phylum Firmicutes in both saliva and buccal mucosa as compared to plaque (Segata et al. 2012; X. Xu et al. 2015). While some metagenomics studies look at these individual habitats separately, it is also not uncommon to use an oral rinse as a sample collection method, in order to obtain a representative sample of the overall oral microbiome (**Chapter 2**, (Willis et al. 2018; H. Wang et al. 2017; Kato et al. 2016)).

Regardless of the particular biome or habitat being explored, the current trend in microbiome studies is largely in taking advantage of culture-independent NGS technologies, as they continue to decrease in both financial and computational cost, alongside the continuous expansions of databases of microbial genetic sequences. According to the expanded Human Oral Microbiome Database (HOMD) (Escapa et al. 2018), only 57% of the oral bacterial species have been officially named, 13% have been cultivated yet remain unnamed, and 30% are uncultivated. Hence, not only do the NGS techniques make analyses relatively quick and easy, but they have also vastly expanded our awareness of unculturable and/or rare microbiota.

The mouth can be affected by several pathologies that have high prevalence among human populations, including periodontitis, gingivitis, and dental caries, all of which have been clearly related to alterations in the oral microbiome (see references in **Table 1.1**). However, the mouth constitutes an entry point to the respiratory and digestive systems, and it is highly vascularized, resulting in potential implications of the oral microbiome in other systemic diseases. Indeed, a growing number of studies have shown associations between other diseases and changes in the oral microbiome (**Table 1.2**). This suggests that oral microbiota may provide potential biomarkers in the diagnosis of some systemic diseases.

Table 1.1: Examples of metagenomic studies of associations between the oral microbiome and oral diseases. The first column indicates a disease, the second indicates organisms that have been found at higher abundances in individuals presenting with the disease, the third indicates organisms at lower abundances, and the fourth contains the references to the literature, which displays these findings. (*) indicates taxa associated with oral cancer from a study in which samples were from tumor and non-tumor sites in the same patients and disease treatment is not specified.

Disease	Associated organisms	Inhibited organisms	Reference
Periodontitis	**Phyla**: Spirochaetes, Synergistetes and Bacteroidetes **Classes**: Clostridia, Negativicutes and Erysipelotrichia **Genera**: *Prevotella, Fusobacterium* **Species**: *Porphyromonas gingivalis, Treponema denticola, Tannerella forsythia, Filifactor alocis, Parvimonas micra, Aggregatibacter actinomycetemcomitans* **Archaea**: *Methanobrevibacter oralis, Methanobacterium curvum/congolense*, and *Methanosarcina mazeii*	**Phyla**: Proteobacteria **Classes**: Bacilli **Genera**: *Streptococcus, Actinomyces, Granulicatella*	(Matarazzo et al. 2011; Lepp et al. 2004; Griffen et al. 2012; Vartoukian, Palmer, and Wade 2009; Costalonga and Herzberg 2014; B. Liu et al. 2012; Jorth et al. 2014; Haubek 2010)
Dental caries	**Genera**: *Neisseria, Selenomonas, Propionibacterium* **Species**: *Streptococcus mutans, Lactobacillus* spp. **Fungi**: *Candida albicans*	**Species**: non-mutans *Streptococci, Corynebacterium matruchotii, Capnocytophaga gingivalis, Eubacterium IR009, Campylobacter*	(Gross et al. 2010; Koo and Bowen 2014)

		rectus, Lachnospiraceae sp. C1	
Oral cancer	**Species**: *Capnocytophaga gingivalis, Prevotella melaninogenica* and *Streptococcus mitis, Peptostreptococcus stomatis*, Streptococcus salivarius*, Streptococcus gordonii*, Gemella haemolysans*, Gemella morbillorum*, Johnsonella ignava* and Streptococcus parasanguinis I**	**Species**: *Granulicatella adiacens**	(Mager et al. 2005; Pushalkar et al. 2012; L. Wang and Ganly 2014)
Esophageal cancer	**Species**: *Tannerella forsythia, Porphyromonas gingivalis*	**Genera**: *Neisseria* **Species**: *Streptococcus pneumoniae*	(Peters, Wu, Pei, et al. 2017)

Table 1.2: Examples of metagenomic studies of associations between the oral microbiome and systemic diseases. The first column indicates a disease, the second indicates organisms that have been found at higher abundances in individuals presenting with the disease, the third indicates organisms at lower abundances, and the fourth contains the references to literature which displays these findings.

Disease	Associated organisms	Inhibited organisms	Reference
Colorectal cancer	**Genera**: *Lactobacillus, Rothia* **Species**: *Fusobacterium nucleatum*		(Kato et al. 2016; Broecker et al. 2017; Oh et al. 2016; Wantland et al. 1958)
Pancreatic cancer	**Genera**: *Leptotrichia* (later in progression of	**Genera**: *Leptotrichia* (at	(Fan, Alekseyenko, et

	disease) **Species**: *Porphyromonas gingivalis and Aggregatibacter actinomycetemcomitans* (at onset of disease)	onset of disease) **Species**: *Porphyromonas gingivalis and Aggregatibacter actinomycetemc omitans* (later in progression of disease)	al. 2018; P. J. Torres et al. 2015)
Cystic fibrosis	**Species**: *Streptococcus oralis* (depends on environmental conditions), S. *mitis*, *S. gordonii* and *S. sanguinis*	**Species**: *Streptococcus oralis* (depends on environmental conditions)	(Whiley et al. 2015)
Cardiovascular disease	**Species**: *Campylobacter rectus*, *Porphyromonas gingivalis*, *Porphyromonas endodontalis*, *Prevotella intermedia*, *Prevotella nigrescens*, (oral commensals that were found on athersclerotic plaques - not necessarily at high abundance in oral cavity)		(Teles and Wang 2011; Chhibber-Goel et al. 2016)
Rheumatoid arthritis	**Genera**: *Veillonella*, *Atopobium*, *Prevotella*, *Leptotrichia* **Species**: *Rothia mucilaginosa*, *Rothia dentocariosa*, *Lactobacillus salivarius*, *Cryptobacterium curtum*	**Genera**: *Haemophilus*, *Neisseria* **Species**: *Porphyromonas gingivalis*, *Rothia aeria*	(Roszyk and Puszczewicz 2017; Scher et al. 2012; Xuan Zhang et al. 2015; S. B. Brusca, Abramson, and Scher 2014)
Alzheimer's disease	**Phyla**: Spirochaetes **Species**: *Porphyromonas*		(Dominy et al. 2019; Miklossy 2016; Aguayo et

	gingivalis		al. 2018)
Diabetes	**Genera**: *Aggregatibacter, Neisseria, Gemella, Eikenella, Selenomonas, Actinomyces, Capnocytophaga, Fusobacterium, Veillonella, Streptococcus*	**Genera**: *Porphyromonas, Filifactor, Eubacterium, Synergistetes, Tannerella, Treponema*	(Casarin et al. 2013)

As we search for these deviations within populations, we must also consider the caveat that our knowledge of the human microbiome may be far from complete. Recent studies have collected data from previously unstudied populations and found not only differences in composition but even added many undiscovered species to public databases (Pasolli et al. 2019; Clemente et al. 2015), highlighting strong disparities between different regions of the world. Most studies have been focused on European, North American, Chinese, or other so-called "WEIRD" populations. This refers to Western, educated, industrialized, rich, and democratic nations, an acronym that was proposed originally to denote a bias in psychology studies toward these societies, which at the time comprised about 13% of the world's population, yet accounted for between 60% and 90% of subjects in psychology studies (Henrich, Heine, and Norenzayan 2010). This early evidence of a similar bias in microbiome studies suggests that there remains the possibility of a much broader landscape of "healthy" microbiomes across different cultures. Furthermore, even within healthy sample sets of the same populations, distinct subgroups can be elucidated (**Chapter 2**, (Willis et al. 2018; Arumugam et al. 2011; Zaura et al. 2017; De Filippis et al. 2014; Ding and Schloss 2014; Takeshita et al. 2016)), Therefore, in addition to a focus on understanding the causes and effects of dysbiosis in the microbiome, there must also be a continued emphasis on fully characterizing the healthy microbiome to more reliably detect true deviations from the normal state.

Studies aiming to characterize the composition of the oral microbiome in diverse human populations are progressing rapidly, as are those looking for variation according to a growing number of parameters, particularly those related to health and disease. In this review, we provide a general overview of the state-of-the-art methodologies used to study the oral microbiome, and of the main results obtained during the last decade of intense research. Finally, we will discuss the current challenges and perspectives of this fast-moving field. Throughout the review, we will put a focus on the emerging roles of the oral microbiome in health and disease, and the new opportunities for therapeutics and diagnostics.

1.3 Technical approaches to study the oral microbiome

The costs of sequencing DNA have plummeted thanks to the introduction of NGS technologies, now allowing scientists to sequence several human genomes in a single day at a price of under $1000 per genome, a nearly one million-fold decrease from 20 years ago (Malla et al. 2018). Similarly, there are a number of cost-efficient NGS techniques that can be used today when approaching microbiome studies, depending on what the researcher hopes to learn (**Figure 1.1**). The two most widely used approaches include whole metagenome shotgun sequencing (WMS) and 16S ribosomal RNA amplicon sequencing, both of which involve reading the DNA sequences of the microbes present in a sample and comparing them to a database of sequences to establish the relative quantities of the different organisms present in that sample. In WMS sequencing (**Figure 1.1C**), the DNA is randomly fragmented multiple times, allowing millions of short sequences to be read in parallel, and then they are reassembled into full (or partial) genomic sequences by connecting the overlapping ends (S. Anderson 1981). However, 16S rRNA sequencing (**Figure 1.1A**), also known as 16S barcoding, has been used more frequently in metagenomic studies, since it is less

expensive, both experimentally and computationally, permitting larger scale study designs. The 16S rRNA gene is common to all bacteria and archaea, and has highly conserved regions, which make it a useful marker gene for the use of universal primer sequences to isolate it for sequencing. Scattered amongst the conserved regions of the gene are nine hypervariable regions (named V1 through to V9), and it is these segments that allow the taxonomic identification of organisms when mapping reads to a database of known 16S rRNA sequences (Weisburg et al. 1991).

Of course, even with the use of ASVs for identification, 16S sequencing still lacks significant taxonomic resolution as compared to WMS sequencing, often only permitting distinction up to the genus level. Alternatives to 16S sequencing have also been proposed in order to improve resolution or to avoid bias due to the varying number of copies of the 16S gene in different species (Kembel et al. 2012) (though there are methods to correct for it (Louca, Doebeli, and Parfrey 2018)). The *rpoB* gene, for instance, has the advantage of generally being single copy and having greater variation, which allows for deeper taxonomic resolution. However, the corresponding lack of conservation makes it less applicable as a universal marker (Vos et al. 2012). A database of *rpoB* gene sequences is available from the FROGS (Find, Rapidly, Otus with Galaxy Solution) website (Escudié et al. 2018). Some have proposed that one or more housekeeping genes, like *rpoB*, should be sequenced along with the 16S gene, since they are ubiquitous and rapidly evolving, allowing for better taxonomic resolution than the 16S gene alone (Ogier et al. 2019). To distinguish closely related organisms, others have suggested a multilocus sequence analysis (MLSA) approach, wherein multiple housekeeping genes from distinct chromosomal loci are sequenced in parallel (Martens et al. 2008). Nevertheless, the 16S rRNA gene remains the current standard for marker gene analyses of the microbiome.

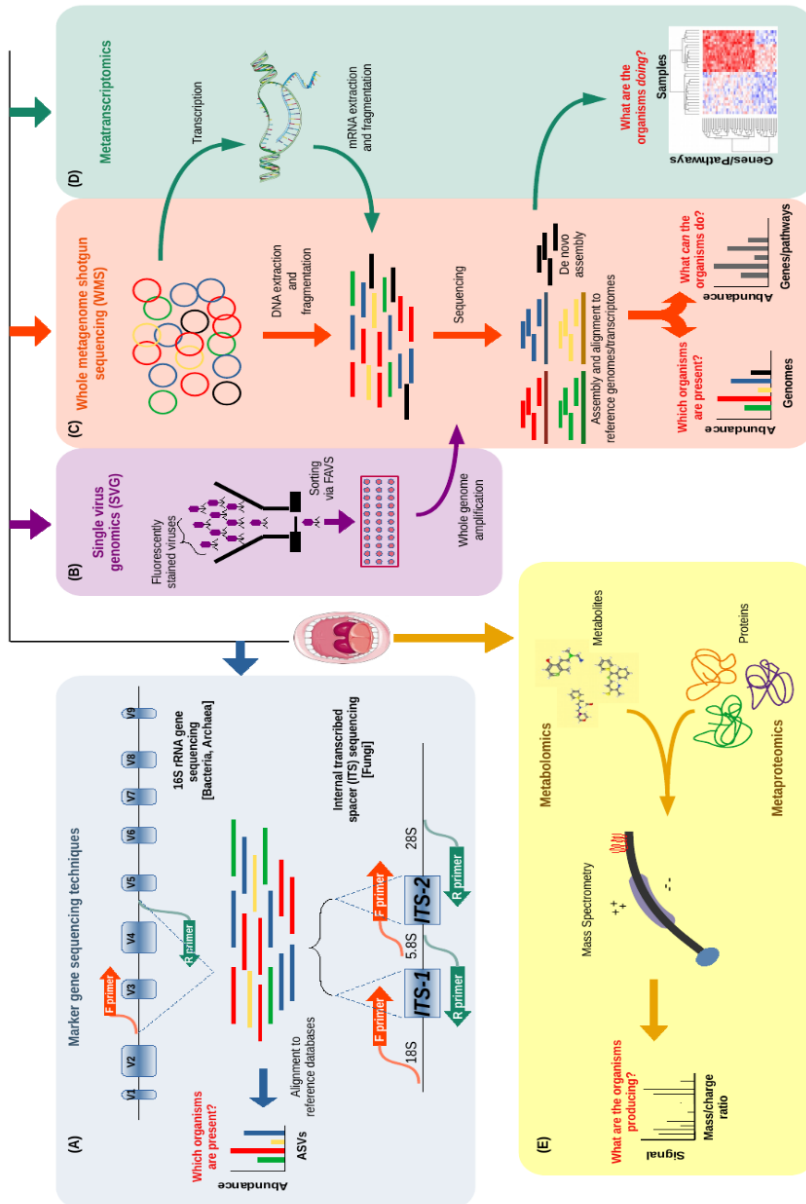

Figure 1.1: Schematics of standard techniques used in microbiome studies. (**A**) Marker gene sequencing techniques can use primers to target certain conserved regions of a genome to capture intermittent variable regions, which can then be

used to identify organisms in a sample rapidly and inexpensively. The 16S rRNA gene is the most commonly used marker gene in bacteria and archaea, and in the figure, primers are used to capture the V3 and V4 variable regions together, a common approach for 16S sequencing. The internal transcribed spacer (ITS) region of the nuclear rRNA cistron in fungi is made of two segments, which can be captured with primers targeting the 18S, 5.8S, and 28S rRNA sections that surround them. (**B**–**D**) Instead of targeting one small segment of the genome, these techniques capture the entirety of the genetic material from an organism. (**B**) Single virus genomics (SVG) uses a fluorescent stain to isolate individual virus particles in a sample by fluorescence-activated virus sorting (FAVS), wherein they are embedded in an agarose bead before undergoing whole genome amplification and sequencing. (**C**) Whole metagenome shotgun sequencing (WMS) involves the fragmentation of all DNA in a sample, sequencing of the fragments, and assembly of the sequences, which can then be mapped to reference genomes, or de novo assembly can be performed. (**D**) Metatranscriptomics also involves a shotgun sequencing approach, but it is performed after mRNA extraction. The outputs then allow for differential gene expression analysis. (**E**) Metabolomics and metaproteomics allow for quantification of the metabolites and proteins produced by the microbiome in a sample, respectively. Mass spectrometry is a common approach to quantification. Mock metabolite shapes in **Figure 1.1** were generated using the JSME Molecular Editor by Peter Ertl and Bruno Bienfait licensed under CC-BY-NC-SA 3.0. Images of body sites and organs in **Figure 1.1** and **Figure 1.2** were obtained from Servier Medical Art by Servier licensed under CC-BY 3.0.Traditionally, 16S sequences were clustered into groups with at least 97% identity, called operational taxonomic units (OTUs), which have been used as proxies for species-level or, more commonly, genus-level taxonomic identification. A number of software tools are available, which convert reads to sample-by-OTU feature tables, such as QIIME (Caporaso et al. 2010) and mothur (Schloss et al. 2009). However, newer approaches are better able to control for amplicon sequencing errors, and thereby obviate the use of arbitrary identity thresholds, allowing for single-nucleotide resolution with amplicon sequence variants (ASVs) (Benjamin J. Callahan, McMurdie, and Holmes 2017). Software options for ASV methods include DADA2 (Benjamin J. Callahan et al. 2016) and Deblur (Amir et al. 2017).

All of the marker gene techniques mentioned are useful when asking the question, “What microorganisms are present in a sample?”, giving an overview of the microbial makeup across many samples. However, WMS sequencing can allow for the detection of species or even strains, in addition to functional annotations of microbiome samples (Ranjan et al. 2016), which can only be

predicted based on known full genome sequences when performing 16S sequencing. So, WMS additionally gives insight into the functional potential of the microbiome, allowing researchers to ask the question "What can the microorganisms present actually do?"

Metagenome studies can be further bolstered by the use of metatranscriptomics (Moran 2009), metaproteomics (Heyer et al. 2017), and metabolomics (Fiehn 2002), though only the first of these utilizes NGS technologies. Metatranscriptomics (**Figure 1.1D**) answers the question, "what are the microorganisms doing?" Here, the idea is to profile the total microbial gene expression in a sample by capturing the total messenger RNA (mRNA) content, so this is particularly useful when exploring the functional activity of the microbiome in different conditions, like disease vs. health, different diets, or different times of the day. Metaproteomics (**Figure 1.1E**) is another approach to assessing the functional activity of a microbiome sample, but instead of sequencing genetic material, the idea is to catalog the abundances of the microbial proteins present in a sample. This is typically done by protein extraction and tandem mass spectrometry analysis (MS/MS) (Heyer et al. 2017). Metabolomics (**Figure 1.1E**), on the other hand, answers the question, "what are the microorganisms producing in a given sample?" The metabolome is the total set of small molecules produced by the microbiome (and the host) in a sample, and can be a strong indicator of the health or dysbiosis of a sample (Bernini et al. 2009). Metabolites are typically quantified by use of chromatography and detection techniques like mass spectrometry (MS) and nuclear magnetic resonance (NMR). Each of these techniques also has its own drawbacks, which prevent it from being as widely used as metagenomics. Metatranscriptomics can be hindered by the instability of mRNA and the excess of rRNA (though methods have been developed to counteract this (Peano et al. 2013)), and by the limited reference databases of transcriptomes (Aguiar-Pulido et al. 2016). Metaproteomics suffers from computational limitations when querying protein databases (which, nevertheless, remain incomplete), as well as a redundancy in annotations due to identical peptides in homologous proteins from different organisms, which may use them in different processes,

thereby leaving the resulting taxonomic and functional quantifications ambiguous (Heyer et al. 2017). Recent tools, however, claim to combat both of these issues (Easterly et al. 2019). With metabolomics, the challenges lie in determining whether the metabolites are produced by the host or the microbiome, and associating them with the relevant genes and pathways, highlighting the need for the integration of this technique with other omics data (Aguiar-Pulido et al. 2016).

All of these tools and techniques are frequently aimed at investigating the bacteriome, which makes up the most significant portion of the microbiome but not its entirety. To classify the composition of the mycobiome, the fungal component of the microbiome, researchers often use a marker region, much like the 16S rRNA gene, called the internal transcribed spacer region of the nuclear ribosomal RNA cistron, referred to as the ITS region (**Figure 1.1A**), which provides a similar taxonomic resolution to that of 16S sequencing for bacteria (Schoch et al. 2012). The virome, the viral component of the microbiome, can be difficult to approach since there are no conserved marker regions like in the 16S rRNA gene in bacteria or the ITS region in fungi. Thus, the full virome must be sampled and compared to known viral sequences. Two problems arise from this, as current viral databases lack the characterization of many viruses, and consequently, any new viral sequences that do not match closely to those in current databases would be difficult to classify (Wylie, Weinstock, and Storch 2012). Another challenge for virome studies is the relatively low proportion of viral nucleic acid content alongside that of other microbes. However, there have been enrichment procedures proposed to increase the content of viral nucleic acids (Thurber et al. 2009). More recently, a new approach called single virus genomics (SVG) has been proposed (**Figure 1.1B**), in which individual viruses are isolated via fluorescence-activated viral sorting (FAVS), and genomic material is amplified and sequenced (Allen et al. 2011).

Whatever the technique being employed, it is important that researchers come to a consensus on the exact procedure for

collection and sequencing to ensure reproducibility. While some of these referenced studies have shown that the microbiome profile of a sample is not heavily influenced by the collection technique (Y. Lim et al. 2017; Fan, Peters, et al. 2018), these are focused on large-scale differences. However, as sequencing technologies become more efficient and microbiome-associated databases become more complete, researchers will continue to compare samples at finer scales, so minor technical variability stemming from different techniques of swabbing collection sites or using a different solution for oral rinse collections could potentially impact results. Some reviews of the current best practices have been published (Mallick et al. 2017; Knight et al. 2018) and these should continue to be improved and built upon.

Depending on the investigator's study goals, there is a wide variety of potential approaches to data analysis, and there are equally plentiful software packages available. The phyloseq (McMurdie and Holmes 2013) and microbiome (Leo and Shetty 2017) packages for R offer a means to organize the data from sequencing experiments alongside any metadata, and provide a collection of tools and tutorials for calculations and plotting in typical microbiome analyses. This includes functions for calculating the alpha diversity (the relative diversity of taxa present in a sample) and beta diversity (the relative distance between any two samples based on the overall composition, as well as plots for various ordination methods. From there, a bioinformatician can go in any number of directions, so here we will just mention some of the most prominent analyses. The vegan R package (Jari et al. 2016) offers options for multivariate tests like the anosim function (analysis of similarities) to determine differences in microbiome compositions between groups of samples, and the adonis function for permanova (permutational multivariate analysis of variance), which can apply linear models to determine sources of variation amongst samples. Linear mixed effects models can also be applied to determine the effects on various data points, like particular taxa, diversity levels, or other metadata variables. Standard or generalized linear models can be fitted using the lm or glm functions, respectively, from the core R package called stats (R Core Team 2020). Mixed effects models can

also be used with the lmer and glmer functions from the lme4 package (Bates et al. 2015). There are also software options available to predict the functional capability of a microbiome sample sequenced with a marker gene target like 16S rRNA. The tools PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states) (Langille et al. 2013) and Tax4Fun (Aßhauer et al. 2015) use reference genome databases to attempt to reconstruct full metagenomes from each sample. Machine learning techniques have also been implemented to attempt to predict disease based on the microbiome composition (Y.-H. Zhou and Gallins 2019), and some investigators have made their code publicly available (Pasolli et al. 2016; Duvallet et al. 2017). The oral microbiome has even been used in a classifier for colorectal cancer with some success (Flemer et al. 2018).

A recent point of contention in the field is that of the compositional nature of microbiome data and the implication for its analyses, something which some have begun to address from both an experimental (Vandeputte et al. 2017) and a statistical (Gloor et al. 2017) perspective. Since the reads produced in an NGS experiment are essentially random samples of the relative abundances of the organisms present, this cannot account for the implications of differences in the total abundances of organisms, which may be physiologically relevant. Vandeputte et al. described an experimental technique called quantitative microbiome profiling (QMP) (Vandeputte et al. 2017) in which the total microbial load is determined by sampling an equal number of sequences per sample, and then correcting for the 16S copy number bias and the total number of cells from the sample. They were able to use this approach to reveal erroneous results from a standard experiment, suggesting a solution for some technical biases inherent in NGS studies. However, since most datasets to date have not been sequenced with any techniques like QMP, other solutions have been proposed to handle the typical asymmetrical datasets in microbiome studies. Rarefaction of read counts, in which a random sample of the same number of reads is extracted from each sample, was a common approach in earlier microbiome studies, but this practice has been discouraged as it omits biologically relevant information

(McMurdie and Holmes 2014). Gloor et al. instead suggested normalizing the data using a centered log-ratio transformation, which is minimally affected by the depth of reads for a sample (Gloor et al. 2017). They also provided a tutorial for the workflow that accompanies their publication on this topic.

1.4 The oral cavity and its microbial niches

Of all the habitats within the human body for which the microbiome is typically studied, the oral cavity warrants perhaps the most unique approach to study, in that it contains a number of highly distinct niches formed at the various surfaces within the mouth. Changes in the availability of oxygen, nutrients, and the pH-mediating effect of saliva (Wilson 2009) can promote the growth of different organisms, and conversely, these organisms can be involved in their own small niche construction (Laubichler and Renn 2015) via biofilm formation and nutrient metabolism, which can produce effects both within the oral cavity (**Table 1.1**, **Figure 1.2**) and systemically (**Table 1.2**, **Figure 1.2**). Some researchers have chosen to study all these niches in parallel to compare them against each other (Human Microbiome Project Consortium 2012; Segata et al. 2012; X. Xu et al. 2015), some have selected individual sites with particular focuses on localized dysbiosis in disease states (van der Meulen et al. 2018; Ganesan et al. 2017; Abusleme et al. 2013; Moutsopoulos and Konkel 2018; Mark Welch et al. 2016; Wei et al. 2019; Asakawa et al. 2018; Lu et al. 2019; Fukui et al. 2018), and others have used an oral rinse approach to capture an overall view of the oral cavity (**Chapter 2**, (Willis et al. 2018; H. Wang et al. 2017; Kato et al. 2016)).

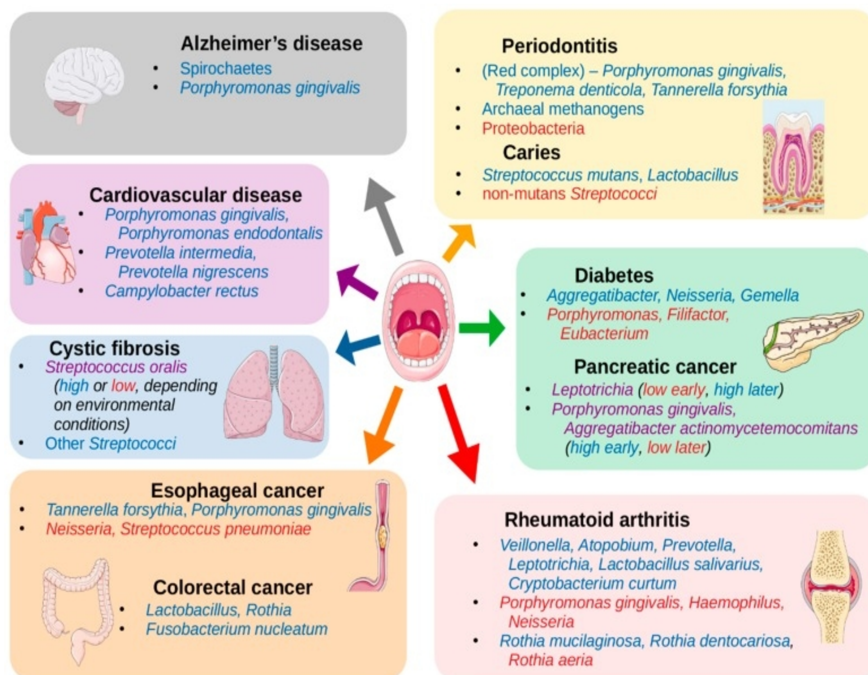

Figure 1.2: Oral and systemic diseases associated with the oral microbiome. A representation of the associations found between diseases with increases or decreases of the abundances of organisms in the oral cavity (listed in **Table 1.1** and **Table 1.2**). Organisms listed in blue have been shown to be increased in abundance in the oral cavity in individuals presenting with the noted disease, and organisms listed in red have been shown to be decreased. Those in purple may be either increased or decreased depending on the conditions or progression of the disease. Images of body sites and organs in **Figure 1.1** and **Figure 1.2** were obtained from Servier Medical Art by Servier licensed under CC-BY 3.0.

Studies focusing on specific oral niches typically aim to explore a disease relevant to that site. For instance, primary Sjögren's syndrome (pSS) in the buccal mucosa was believed to be a potential reservoir for pathogens implicit in the disease, wherein the disease samples were shown to have higher Firmicutes/Proteobacteria ratios as compared to healthy controls, and higher abundances of 19 genera (van der Meulen et al. 2018). Various efforts have

characterized the changes in subgingival and supragingival plaque related to periodontitis (Ganesan et al. 2017; Abusleme et al. 2013; Moutsopoulos and Konkel 2018; Mark Welch et al. 2016), as well as a study of subgingival plaque and buccal mucosa showing that both sites differed between periodontitis samples and healthy controls, with many of the same organisms affected in both sites, though they also displayed unique species colonization (Wei et al. 2019). The tongue microbiome was explored in the elderly in Japan because of a potential connection between ingested microbes and pneumonia, which found that samples with worse dental health were enriched in pneumonia-associated bacteria (Asakawa et al. 2018). The tongue was also targeted as a potential segment in diagnostic tools that would perhaps incorporate the microbiomes of the full gastrointestinal tract to detect pancreatic cancer (Lu et al. 2019). The palatine tonsils were explored in HIV-infected patients to better understand the oral and systemic complications of the disease, and it was shown that the bacteriome was indeed significantly altered in infected individuals, but the mycobiome was not (Fukui et al. 2018).

At a broad scale, the microbial composition throughout the regions of the oral cavity is fairly consistent, making it easily distinguishable from the microbiomes of other human body habitats (Human Microbiome Project Consortium 2012; Koren et al. 2013; Debelius et al. 2016; Costello et al. 2009; Vázquez-Baeza et al. 2017). However, while the niches in the oral cavity are largely composed of the same organisms, some may be present in different proportions. One study combining samples from 10 niches along the digestive tract in over 200 individuals from the United States placed these niches into four groups based on the similarity of overall composition (Segata et al. 2012). One of the sites was the intestine, represented by stool samples, which were grouped alone, while the other nine were in the mouth and throat. One of the three other groups consisted of buccal mucosa, keratinized gingiva, and hard palate, another of tongue, saliva, palatine tonsils, and throat, while the last group contained subgingival and supragingival plaques. Though all of the non-stool niches were generally dominated by the phyla Firmicutes and Bacteroidetes, they based the groups more on

small-scale differences. The first group was more unique than the other two non-stool groups, and was shown to have a considerably higher abundances of the genus *Streptococcus* and lower overall alpha diversity, which is a measure of the relative diversity of organisms present in a given sample. Group 3, containing the two types of gingival plaque, typically had higher alpha diversity. Comparisons between the compositions of these niches and their diversities have been corroborated in other studies (Human Microbiome Project Consortium 2012; X. Xu et al. 2015; Eren et al. 2014). The authors posit that the level of saliva flow in the mouth is a key factor determining the composition of the microbiome at each niche in the oral habitat because of its capacity to regulate pH and nutrient availability, but other major factors may include the type of surface and oxygen availability. The two plaques, for instance, form on the non-shedding surfaces of teeth where they produce biofilms, within which oxygen is limited, resulting in greater abundances of obligate anaerobic organisms in the subgingival plaque and of facultative anaerobic organisms in the supragingival plaque (Segata et al. 2012). They suggest that one niche from each group could be used to represent all of the niches in that group, such as using the buccal mucosa microbiome as a proxy for both keratinized gingiva and hard palate microbiomes (group 1 niches). However, as sequencing techniques continue to improve and costs continue to decrease, we may find more and more subtle idiosyncrasies within each niche, due to combinations of the microenvironmental factors just mentioned, as well as any others. So, that decision will be up to the discretion of the researchers and the relevance to their studies.

Since these niches tend to have very similar overall microbiome compositions at all but the lowest taxonomic levels, many researchers choose to treat the oral cavity as an individual habitat and analyze global compositions and processes therein (**Chapter 2**, (Willis et al. 2018; Kato et al. 2016; Y. Lim et al. 2018)). A few studies have shown that this is a viable and efficient method for sample collection to investigate the oral microbiome (Y. Lim et al. 2017; Fan, Peters, et al. 2018), and standardized procedures have been proposed (Woo and Lu 2019; “XIT Genomic DNA from Buccal Cells: For Extraction of Genomic DNA from Buccal/Cheek

Cells" 2012). Essentially, an individual would refrain from eating, drinking, brushing, or smoking (anything that might temporarily shift the typical microbial composition) for at least 30 min prior to collection. Then, they would swish with a buffer solution for about 30−60 s, and then spit the contents into a tube, which would later be centrifuged and sequenced. The practical benefits of using an oral rinse are the ease of collection, as it is a quick and non-invasive method to obtain oral microbial DNA from a study participant, as well as the ease of storage and transport, since these samples can be frozen and sequenced later without detriment to the quality of the samples (Fan, Peters, et al. 2018; Pramanik et al. 2012). This is advantageous for large-scale microbiome projects, as it allows for the collection of many samples, which may take weeks or months, that can later be sequenced together to minimize the potential technical bias inherent in sequencing projects, as mentioned in the previous section.

1.5 The healthy oral microbiome and definition of stomatotypes

The field of microbiome research is arguably still in its infancy, as evidenced by the continuing efforts to expand the databases on known microbial genomes, and to combat the bias toward "WEIRD" populations, as mentioned in the introduction (Pasolli et al. 2019; Clemente et al. 2015; Lassalle et al. 2018). These factors alone make it difficult to effectively define what constitutes a "healthy" oral microbiome. However, on top of that, the accumulation of studies over the last decade or so have shown that, even within particular populations, there can potentially be multiple distinct trends of microbiome composition amongst individuals in relative general health (**Chapter 2**, (Willis et al. 2018; Zaura et al. 2017; De Filippis et al. 2014; Ding and Schloss 2014; Takeshita et al. 2016)). As such, while we may be eager to investigate the microbiome's relationship to disease, it is vital that we also continue to further define microbiome compositions in health among

different populations, and explore the causes of shifts within or between these populations.

Projects based on oral rinse samples are ideal for observing how the microbiome of the oral habitat as a whole is affected by external factors. Of course, the phrase “external factors” could cover a wide spectrum of variables, but some with clear connections to the mouth are the water we drink and the food we eat. In a cohort of 1319 samples from healthy adolescents in Spain analyzed by 16S rRNA sequencing, it was shown that differences in the ionic composition of public drinking water was associated with shifts in the overall composition of the oral microbiome (**Chapter 2**, (Willis et al. 2018)). Samples from regions with greater alkalinity and greater levels of ions, such as sulfate (SO4) and sodium (Na), had higher abundances of genera, such as *Porphyromonas* and *Flavobacterium*, while regions with lower levels showed higher abundances of other genera, including *Veillonella*, *Pseudomonas*, and *Ralstonia*. Different diets have also been shown to contribute to variations in the microbiome composition, such as in the WMS-based study comparing the oral microbiomes from populations of hunter-gatherers (HGs) from the Philippines, traditional farmers (TFs) from the Philippines, and Western controls (WCs) from the Human Microbiome Project (samples from the United States) (Lassalle et al. 2018). They showed that the HG samples had higher alpha-diversity while it was lower in WC samples, and TF samples fell in the middle. Likewise, there was a strong gradient in the abundances of the core oral genera *Neisseria* and *Haemophilus*, with high levels of *Neisseria* and low levels of *Haemophilus* in HG samples, the reverse in WC samples, and TF samples again falling in between. The HG samples, despite good oral health, also displayed higher abundances of a number of species typically considered to be oral pathogens associated with gingivitis and periodontitis by Western standards. Functional analyses revealed an increase in vitamin B5 biosynthesis pathways in HG samples and, to a lesser extent, in TF samples. Americans had been shown to consume greater quantities of foods with vitamin B5, so the authors posit that this lack in hunter-gatherer diets would select for organisms that synthesize it

on their own. Conversely, they showed that WC samples, and to a lesser extent, TF samples were enriched in urease activity, particularly from *Haemophilus* spp. This urease counteracts the drops in pH that occur when bacteria degrade sugars into acidic compounds, so selecting for these organisms in WC samples, with their sugar- and starch-heavy diets makes sense. The authors thus suggest that organisms considered to be oral pathogens in Western populations could indeed be part of the healthy microbiomes of different populations like hunter-gatherer societies, and that pathogenic strains of these organisms would be selected based on the nutrient availability tied to diet.

Food and water are obvious influencing factors, but any number of other factors could also impact the microbiome. A common approach in the early stages of the analysis of the microbiome of any body habitat is to first look at the broad effects of such factors by clustering the samples based on the overall microbial composition. The notion of separating samples into clusters was discussed in an early NGS-based study on the gut microbiome, wherein the authors labeled the clusters “enterotypes”, implying the presence of different putative categories of gut microbial composition (Arumugam et al. 2011). They found three distinct enterotypes, the separations of which were driven largely by differences in the abundances of particular organisms, namely the genera *Bacteroides*, *Prevotella*, and *Ruminococcus*, and corroborated these by finding very similar enterotypes from two separate sample sets. From this, they suggested that there may exist some limited number of equilibria of symbiotic states between a human host and its microbiome, which would arise due to different diets and lifestyles.

Many studies have since adopted this technique, and in studies of the oral cavity, similarly composed clusters of samples have also emerged, dubbed “stomatotypes” in one such study, as homage to the original term enterotype, but in reference to the mouth (**Chapter 2**, (Willis et al. 2018)). A summary of some of the genera of bacteria that have been found to co-occur in different stomatotypes across studies is found in **Table 1.3**. There have thus been shown at

least two strongly corroborated stomatotypes, one including higher abundances of the Proteobacteria genera *Neisseria* and *Haemophilus*, and the other with higher abundances of the Bacteroidetes genus *Prevotella* and the Firmicutes genus *Veillonella*. Some studies have shown more than just these two stomatotypes, though the consensus of compositions is more varied. Some of the genera co-occur in different manners, depending on the study.

Table 1.3: Consensus stomatotype-driving genera. Genera that have been shown in the literature to strongly drive the distinction between samples of oral microbiome datasets by differences in their abundances. The stomatotype numbers are arbitrarily assigned. The genera are listed in the second column, along with notes on associations between the organisms where relevant. The third column contains the references to the literature, which shows these stomatotype associations.

	Genus	**References**
Stomatotype 1	***Neisseria***	(Willis et al. 2018; Zaura et al. 2017; De Filippis et al. 2014; Takeshita et al. 2016)
	Haemophilus	(Willis et al. 2018; Zaura et al. 2017; Takeshita et al. 2016)
Stomatotype 2	***Prevotella***	(Willis et al. 2018; Zaura et al. 2017; De Filippis et al. 2014; Takeshita et al. 2016)
	Veillonella	(Willis et al. 2018; Zaura et al. 2017; Takeshita et al. 2016)
Variable Stomatotypes	***Streptococcus*** – varies depending on study and species	(Willis et al. 2018; Zaura et al. 2017; De Filippis et al. 2014; Takeshita et al. 2016)

	Gemella – co-occurs with *Streptococcus* and *Porphyromonas*	(De Filippis et al. 2014; Takeshita et al. 2016)
	Porphyromonas – may co-occur with *Streptococcus*, *Gemella*, or *Neisseria*	(Zaura et al. 2017; Takeshita et al. 2016)
	Rothia – co-occurs with varying species of *Streptococcus*, depending on study	(De Filippis et al. 2014; Takeshita et al. 2016)

The explicit nature of stomatotypes may be very appealing as a means of differentiating samples, but these discrepancies highlight two major concerns. The first stems from the fragmentary nature of our current understanding of the microbiome due to the cultural biases and technical limitations already discussed. We do not yet have a complete picture of what may constitute the various potential equilibria of microbial abundances that lead to a particular stomatotype, because we have not explored the healthy microbiomes of many unique populations across the world, and many of the studies we already have lack the resolution to explain fine-scale distinctions between samples. As mentioned above, the hunter-gatherer populations from the Philippines that were sampled were enriched in *Neisseria* spp. while the Western controls were enriched in *Haemophilus* spp. (Lassalle et al. 2018). This might suggest that samples from both populations would be clustered into stomatotype 1 from **Table 1.3**, since these two genera drive the equilibria of stomatotype 1, typically together but not in all studies (De Filippis et al. 2014). Meanwhile, Bacteroidetes (the phylum containing *Prevotella*) and *Veillonella*, the drivers of stomatotype 2, fell more in the middle of the gradient between the HG and WC equilibria, further potential evidence that HG and WC may group together in stomatotype 1. How could one reconcile the strong

differences displayed between these two populations and the evidence so far presented for the compositions within stomatotypes? This could likely be partially explained by the “WEIRD” bias in the studies presenting stomatotypes—perhaps these equilibria emerge in populations with westernized diets, medical treatments, and lifestyles. However, the question also ties into the second major concern, which is the statistical relevance of separating samples into discrete clusters. The notion of analyzing the gradients of microbial abundances was proposed as a response to the enterotype concept (Jeffery et al. 2012). This has led to further discussions of the merits of enterotypes/stomatotypes, cautioning their use as predictive or diagnostic tools (Knights et al. 2014), and also suggesting improvements to their calculations while further emphasizing a focus on the gradients of abundance (Koren et al. 2013; Costea et al. 2018). See **Figure 1.3** for an example of the different gradients of the abundances of organisms in **Table 1.3**, which have been shown as stomatotype drivers, and how they associate with the stomatotypes found in a random subset of 500 samples from an oral microbiome dataset (**Chapter 2**, (Willis et al. 2018)). These, and other studies, suggest that stomatotypes are useful as a first step in exploring underlying variation among samples, which can then be further investigated through deeper analysis of shifts in particular organisms. For instance, in the study of adolescents in Spain, they drew connections to tap water composition by first observing maps of the distributions of samples in the two stomatotypes that they found, which were reminiscent of maps of water hardness values across Spain, and then later began to look at the effects on particular organisms within their samples (**Chapter 2**, (Willis et al. 2018)).

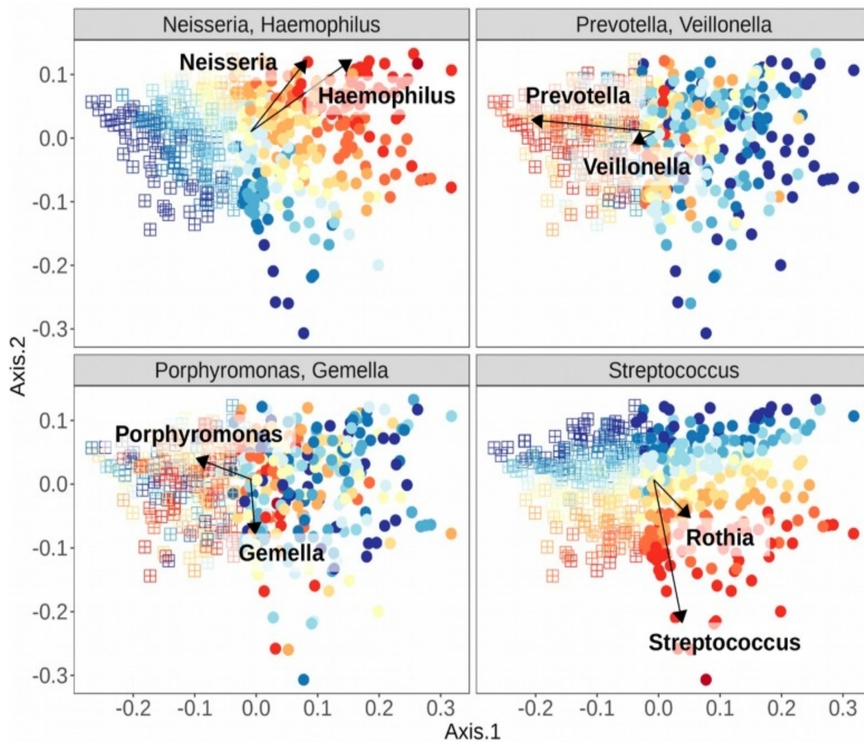

Figure 1.3: Gradients of abundances of consensus stomatotype-driving genera. Using a random subset of 500 samples from an oral microbiome dataset (**Chapter 2**, (Willis et al. 2018)), samples were clustered into two stomatotypes using the weighted Unifrac distance measure. Type 1 samples are represented by circles and type 2 samples by squares. In each box, samples are colored by the total relative abundance of the indicated organisms. Overlaid are arrows indicating the tendency of the abundances of each organism noted in **Table 1.3**. In this subset of samples, *Neisseria* and *Haemophilus* strongly associate with stomatotype 1 samples, *Prevotella* strongly associates with stomatotype 2 samples while *Veillonella* does so weakly. The "variable stomatotype" drivers are indeed variable in their associations in this instance. *Streptococcus* shows a clear gradient but does not conform to either stomatotype. *Gemella* and *Rothia*, which have been shown to co-occur with *Streptococcus* in stomatotypes in the literature, do the same here, with *Rothia* more associated with stomatotype 1. However, *Porphyromonas*, which has been shown to co-occur with *Streptococcus*, *Gemella*, or *Neisseria* previously, associates with none of these here, and instead is strongly associated with stomatotype 2.

1.6 Non-bacterial oral microbes

Bacteria dominate both the research about the human oral microbiome, and the biomass within the oral habitat, with fungi estimated to comprise <0.1% (Jonathon L. Baker et al. 2017). Nevertheless, there is an appreciable diversity of fungal species present in the oral cavity, including species from the genera *Candida*, *Aspergillus*, *Penicillium*, *Schizophyllum*, *Rhodotorula*, and *Gibberella* (Peters, Wu, Hayes, et al. 2017). Yet, two primary complications have limited the exploration of the mycobiome: (1) Difficulty in identifying many fungal species and (2) confusion in fungal nomenclature. Both of these issues have begun to be addressed in large part by the use of NGS technologies. Until recently, the diversity within the oral mycobiome was believed to be quite limited, dominated primarily by a few species of *Candida* (Bandara, Panduwawala, and Samaranayake 2019). This was largely because many fungi are difficult to cultivate in a laboratory, but advances in NGS technologies have revealed a wider array of fungal organisms than previously expected. One study found that the genus *Malassezia* was highly prevalent in the mouth (Dupuy et al. 2014) but had previously gone undetected in this body site because it has particular lipid requirements and needs specialized culture media to grow in a lab, and it was previously believed to be a pathogen on the skin (Saunders, Scheynius, and Heitman 2012). However, even within metagenomic studies, there may be complications in categorizing the true fungal diversity. For instance, a study re-analyzing the samples in which *Malassezia* was detected was not able to find this genus in any of the samples (G. Wu et al. 2015), but it is possibly because this second study did not use the same DNA extraction protocol as the first, which included a step that used beads to help break cell walls and capsules. This highlights an inconsistency in the protocols for fungal metagenomic studies, and the need for standardization.

Independent of the technical concerns surrounding the collection and categorization of genetic material in fungal studies, there is also some ambiguity in the classifications of fungi. For instance, *Malassezia* are dimorphic, with both yeast and mycelial phases, and

in the past had been placed in multiple genera (Dupuy et al. 2014). The authors claim that, while the taxonomy for this particular genus has largely been resolved, older studies may miss this information, and this issue may also occur for other fungal organisms. Within the last decade, there was a push to end the system of dual nomenclature, as this approach came to be seen as archaic, and a single name classification has since begun to be adopted (Hibbett and Taylor 2013). As fungal taxonomy continues to be expanded, NGS-based studies contribute greatly to the identification of new species, both with ITS-amplicon (Schoch et al. 2012) and shotgun metagenomic techniques (Donovan et al. 2018).

Complications in technical approaches and in classification have led to scarce investigations of the oral virome, but we can begin to draw a few conclusions from some of the recent work in this area. This is an important segment of microbiome research because not only can eukaryotic viruses affect the health of a host directly, but prokaryotic viruses can do so as well by altering the overall bacteriome composition and thus its function (Wylie, Weinstock, and Storch 2012). A study in Spain using single-virus genomics (SVG) and viral metagenomics in 15 saliva samples found 439 oral viruses, which they grouped into about 200 clusters that corresponded to genus-level classification (de la Cruz Peña et al. 2018). They saw that most viruses were not consistently predominant, and it was difficult to define a core group of salivary viruses, and instead there were variable interpersonal compositions in the oral virome. However, 26 of their 200 viral clusters shared many genes, and most of these were *Streptococcus* phages, which is a reasonable finding since, as we have seen, *Streptococcus* is typically among the most abundant genera of oral bacteria, if not the most abundant, in Western oral microbiomes. Another study, also in Spain, of 72 healthy adult oral viromes had similar findings (Pérez-Brocal and Moya 2018). They found very few ubiquitous viruses while most were found only in individual samples, and once again *Streptococcus* phages were common. However, they did suggest a small core of oral viruses and pointed to the presence of viral cores in other body sites seen in other studies, including the lung in healthy samples (Willner et al. 2009), the gut even after fecal

transplant (Broecker et al. 2017), and the skin (Oh et al. 2016). They also stress the specificity of this oral viral core to Western cultures, since virome research also suffers from the “WEIRD” bias mentioned above.

Protozoa and archaea are also components of the oral microbiome, though little has been said on either group. There do not appear to be any NGS-based explorations of oral protozoa, but instead they have been identified by microscopy techniques (Wantland et al. 1958; Feki et al. 1981; Chomicz et al. 2002; Cielecka et al. 2000). However, the presence of the 16S rRNA gene in archaea has led to the use of NGS techniques in some studies. All of the archaeal species thus far discovered in the oral cavity are methanogens (methane-producing organisms) of the phylum Euryarchaeota (Wade 2013). It has been shown that these archaea tend to be present at higher abundances in patients suffering from periodontitis (Matarazzo et al. 2011; Lepp et al. 2004). However, it has been suggested that there may be more archaeal diversity that has as yet gone undetected, either because conventional methods have precluded the detection of other archaea, because they occur at low prevalence and abundance, or because of a lack of diversity in the populations sampled (Horz 2015). Each of these issues could be addressed with further explorations of NGS-based studies among diverse populations.

1.7 Oral microbiome and oral diseases

The plant ecologist Robert Harding Whittaker, in defining terrestrial biomes in the 1970s, discussed gradients of environmental conditions ranging from favorable to extreme. He showed that both alpha and beta diversities decrease as biome conditions become more extreme (Whittaker and Others 1975). A parallel to this generalization has been seen with microbiome studies over the last decade if we consider that disease states equate to “extreme” environmental conditions within certain body sites. This frequently causes low alpha diversity (fewer distinct organisms), which leads to low beta diversity (uniqueness of an individual sample’s overall

composition) as certain organisms become better equipped to dominate their habitat. Common diseases of the oral cavity, like periodontitis and dental caries, provide explicit examples of this phenomenon, wherein the microbiome composition is strongly tied to the disease state. However, at this stage in the development of the microbiome field, it is not always clear whether changes in microbial compositions lead to disease, or vice versa. Nevertheless, it is certainly worth discussing the associations that have been observed to begin to postulate microbiome-related mechanisms of disease origin or progression.

The species of the “red complex” (*Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*) have historically been seen as the primary infective organisms implicated in periodontitis (S. S. Socransky et al. 1998), but this was determined by culture-based studies, which thus missed much of the bacterial diversity present in samples. NGS techniques have since revealed other organisms that are also associated with periodontitis (**Table 1.1**, **Figure 1.2**), such as the classes Clostridia, Negativicutes, and Erysipelotrichia (Griffen et al. 2012); the genera *Synergistes* (Vartoukian, Palmer, and Wade 2009), *Prevotella*, and *Fusobacterium* (Costalonga and Herzberg 2014); and the species *Filifactor alocis* (Griffen et al. 2012); as well as the archaeal species *Methanobrevibacter oralis*, *Methanobacterium curvum/congolense*, and *Methanosarcina mazeii* (Matarazzo et al. 2011; Lepp et al. 2004). Conversely, some organisms are associated with periodontal health, including the phylum Proteobacteria and the Firmicutes class Bacilli (Griffen et al. 2012), and the genera *Streptococcus*, *Actinomyces*, and *Granulicatella* (B. Liu et al. 2012).

Clearly, there are many organisms associated in one way or another with periodontitis, but that raises the question of which may actually be causative agents, and which are merely impacted by environmental alterations in the disease state. One study that employed metatranscriptomics techniques compared the expression profiles of 160,000 genes and showed conserved differences in metabolism, despite variation in microbiome composition, suggesting that, in a disease state, the organisms present in a sample

perform similar functions, even if species differ between samples (Jorth et al. 2014). This notion could be corroborated by another study, which proposed that methanogenic archaeal species develop syntrophic relationships by acting as “hydrogen sinks”, allowing for increased growth of pathogenic secondary fermenters. Members of the genus *Treponema* have a similar hydrogen-consuming activity, perhaps explaining their involvement in the “red complex”. Indeed, this study showed that abundances of *Treponema* and methanogenic archaea anti-correlated, suggesting that they may fill the same functional niche (Lepp et al. 2004).

In dental caries, alpha diversity has been shown to diminish as the disease progresses and the species *Streptococcus mutans* has been found at high levels at early stages of caries development but not at later stages while other species of *Streptococcus* are associated with dental health (Gross et al. 2010). It has been suggested that, while *S. mutans* is acidogenic and this may contribute to initial caries formation, other oral taxa are also acidogenic. The significant virulent factor in this situation is its ability to metabolize sucrose from a host’s diet into extracellular polysaccharides (EPS), which are necessary to produce cariogenic biofilms. Furthermore, adhesion between *S. mutans* and *Candida albicans* is promoted in this setting, with *C. albicans* providing additional acidogenesis (Koo and Bowen 2014).

Connections between the microbiome and cancer have also been explored recently. A number of species found in the oral cavity have been associated with oral cancer, including *Capnocytophaga gingivalis*, *Prevotella melaninogenica*, and *Streptococcus mitis* (Mager et al. 2005). **Table 1.1** lists a number of species that have been described as having associations with oral cancer, though there is the caveat that the samples in one of these studies were taken from tumor and non-tumor sites in the same patient, and thus the composition at both sites may in fact be affected by the disease, and it is unclear exactly what, if any, cancer-related treatment the patients have undergone (Pushalkar et al. 2012). Unfortunately, there do not appear to be many studies exploring this connection with modern techniques, nor is there much consensus among such

studies on the microbiome composition in the presence of oral cancer, but there are a number of hypotheses on the potential cancer-promoting action of microbiota. It has been suggested that some of the normal oral bacteria, including *Streptococcus salivarius*, *Streptococcus intermedius*, and *Streptococcus mitis*, can convert ethanol to the carcinogen acetaldehyde (L. Wang and Ganly 2014; Kurkivuori et al. 2007) or upregulate cytokines and other proinflammatory molecules, leading to chronic inflammation that may be involved in carcinogenesis (Meurman 2010), and that bacterial toxins may also affect cell signaling pathways or damage DNA (Lax 2005).

Esophageal cancer has also been explored, with associations seen with the periodontal pathogens *Tannerella forsythia* and *Porphyromonas gingivalis* (members of the "red complex") (Peters, Wu, Pei, et al. 2017). The study showed that the genus *Neisseria* was linked to a lower risk of esophageal cancer, as was the carotenoid biosynthesis pathway, to which a number of *Neisseria* species can potentially contribute.

1.8 Oral microbiome and non-oral diseases

The microbiome of the oral cavity is by no means an isolated biome, but it is instead part of a highly interconnected series of microbiomes across the human body, forming a sort of micro-biosphere. As the entry point of nearly all ingested material, and due to its high vascularity, the oral cavity has ample opportunity to influence activity at other body sites. So, it is no surprise that, in addition to diseases of the oral cavity, the oral microbiome has been implicated in a number of systemic diseases.

The mouth is a direct route to both the lungs and the digestive system, so an association between oral taxa and disorders like cystic fibrosis (CF) (Whiley et al. 2015) or colorectal cancer (CRC) (Kato et al. 2016; McCoy et al. 2013; Castellarin et al. 2012; Kostic et al. 2012) can perhaps be expected, given what we have already discussed. The pathogenicity of *Pseudomonas aeruginosa*, the

primary agent in biofilm formation in the lungs of CF patients, can be inhibited by oral commensal streptococcal species, particularly *Streptococcus oralis*, through the production of hydrogen peroxide, which can disrupt biofilm production, but this was only observed if these streptococci were primary colonizers before the introduction of *P. aeruginosa*. However, in a typical CF lung environment, these streptococcal species actually stimulate the production of *P. aeruginosa* virulence factors, including elastase and pyocyanin (Whiley et al. 2015).

Dysbiosis in the oral cavity resulting in periodontitis has been linked with oral, esophageal, gastric, lung, pancreatic, prostate, hematologic, and breast cancers (Fitzpatrick and Katz 2010; Michaud et al. 2017), amongst others. Hypotheses for these connections include: Production of carcinogenic molecules like nitrosamines by nitrate-reducing taxa (Abnet et al. 2005) or acetaldehyde by ethanol-metabolizing taxa (L. Wang and Ganly 2014; Kurkivuori et al. 2007), increased abundance of cancer-linked viruses like cytomegalovirus and Epstein–Barr virus (Chalabi et al. 2008; Slots, Sugar, and Kamma 2002), and, perhaps most prominently, increased proinflammatory markers stemming from immune reactions to periodontal disease like cytokines (Meurman 2010) and the receptor for advanced glycation end products (RAGE) (Tateno et al. 2009).

Multiple studies have linked *Fusobacterium nucleatum* with CRC (McCoy et al. 2013; Castellarin et al. 2012; Kostic et al. 2012), as this is an oral commensal species that is highly invasive and adherent (Y. W. Han et al. 2000) and appears in adenoma samples of CRC patients. In adenoma cases, it was correlated with local inflammation, TNF-α, and IL-10 (McCoy et al. 2013). However, a recent NGS-based study found no association between *Fusobacterium* and CRC, but instead saw associations with the genera *Lactobacillus* and *Rothia* (Kato et al. 2016). The authors suggest that the results of some other studies may actually be confounded by smoking, which has been shown to associate with *F. nucleatum* abundance in the mouth as well (Sergio Bizzarro et al. 2013; J-H Moon, Lee, and Lee 2015). They also posit that, while

Lactobacillus has been suggested as a probiotic when present in the gut microbiome (Dassi et al. 2018; Saxelin 2008), as seen in the previous section (see **Table 1.1**, **Figure 1.2**), this genus is associated with dental caries in the oral microbiome. Thus, *Lactobacillus* may not have a direct impact on colorectal carcinogenesis, but it could perhaps be an ancillary indicator of poor oral health, which we have seen is strongly linked to cancer.

Another study showed higher abundances of *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* in pancreatic cancer samples (Fan, Alekseyenko, et al. 2018), both of which are keystone pathogens in periodontitis (Costalonga and Herzberg 2014; Haubek 2010). As supporting evidence, they referenced a study that found that risk for pancreatic cancer was significantly increased in the presence of elevated serum antibodies to *P. gingivalis* (Michaud et al. 2013), and another showing that both *P. gingivalis* and *A. actinomycetemcomitans* have the potential to initiate Toll-like receptor (TLR) pathways, which has been shown to be a driver of pancreatic carcinogenesis (Zambirinis et al. 2015). However, another study seems to contradict the first, finding a greater abundance of *Leptotrichia* in pancreatic cancer samples compared to healthy controls, and lower abundance of *Porphyromonas*, as well as lower abundance of *Aggregatibacter* (though the latter was not statistically significant) (P. J. Torres et al. 2015). They suggest that a high ratio of *Leptotrichia* to *Porphyromonas* (LP ratio) is a biomarker for pancreatic cancer. In fact, they were able to reclassify one patient that was originally a healthy control but diagnosed with an unknown digestive disease. The individual's high LP ratio prompted a re-evaluation that led to diagnosis of pancreatic cancer in that patient.

Combining the evidence presented in both studies may lead to a solution for the incongruous findings. Both reference the link between *P. gingivalis* antibodies and pancreatic cancer, so this may be due to the high abundance of this species prior to onset of the disease, leading to antibody production and eventual diminishing abundances. High initial *P. gingivalis* abundance would then potentially be linked to periodontal disease, which may then lead to

various cancers as a result of systemic inflammation or any of the other mechanisms discussed here. Competition between *Leptotrichia* and *Porphyromonas* would explain their anti-correlation, so that as antibodies reduce the abundance of oral *P. gingivalis*, *Leptotrichia* is able to thrive. In fact, in the second study, some of the pancreatic cancer samples had low LP ratios, on par with non-pancreatic cancer LP ratios, so these may then have been early stage cases. This situation highlights a need for a greater focus on the temporal dynamics of the microbiome in these kinds of association studies to discover important factors across the onset, progression, and maintenance of disease states.

Some other systemic disorders are linked with periodontitis as well, like cardiovascular disease/atherosclerosis (Teles and Wang 2011; Chhibber-Goel et al. 2016). One study following over 3000 subjects for a 16-year period found that periodontitis with the loss of molars was linked with breast cancer (among other types), and premature death due to cancer and cardiovascular or gastrointestinal diseases (Söder et al. 2007). As with the cancers that are linked to periodontitis, the resulting increase in systemic inflammation is a primary explanation for the link with cardiovascular disease (Ali et al. 2011). This is largely due to the invasive nature of some of the associated taxa, like *P. gingivalis*, which also promotes invasiveness into host epithelial cells in species like *Prevotella intermedia*, an otherwise commensal oral species (Rangé et al. 2014). The proteins secreted by these organisms are implicated in their pathogenicity, such as gingipains from *P. gingivalis*, which aid in its biofilm formation during periodontitis (Haraguchi et al. 2014), and subsequently activate cytokine production (Jayaprakash, Khalaf, and Bengtsson 2014). Studies linking periodontitis and atherosclerosis have relied on findings of oral bacteria colonizing atherosclerotic plaques (Teles and Wang 2011; Chhibber-Goel et al. 2016) rather than on abundances of taxa in the oral cavity (**Table 1.2**, **Figure 1.2**). Although there appear to be fewer studies using NGS techniques of oral cavity samples to determine the associations between the oral microbiome composition and cardiovascular disease, the potential mechanisms of action by oral taxa make this another attractive area for investigation.

Rheumatoid arthritis (RA) has often been connected with periodontitis (Roszyk and Puszczewicz 2017; Dissick et al. 2010), once again implicating *P. gingivalis*, in this case for its production of gingipains and peptidylarginine deiminase, which enable protein citrullination, an important trigger for RA. However, recently it was shown that this direct connection may be erroneous (Konig et al. 2015). In fact, studies of RA using 16S sequencing and whole metagenome shotgun sequencing, respectively, found that *P. gingivalis* was either not associated with RA (Scher et al. 2012) or actually more abundant in healthy controls compared to RA samples (Xuan Zhang et al. 2015) while different organisms were associated with the disease (**Table 1.2**, **Figure 1.2**). This demonstrates how NGS studies can allow researchers to determine the veracity of previously held beliefs and potentially open new pathways for investigation.

Neurological disorders have also been associated with the oral microbiome. Perhaps the most complete study in this regard is the association of *P. gingivalis* with Alzheimer's disease (Dominy et al. 2019). The authors of this study not only identified this bacterium in the brains of Alzheimer's patients at levels that correlated with tau and ubiquitin aggregates (a hallmark of the disease), but also showed that *P. gingivalis* infection in mice resulted in brain colonization and increased the production of components of amyloid plaques. They went on to show that the gingipains proteases produced by *P. gingivalis* are neurotoxic and inhibit tau function. This suggests a direct connection between baring colonization of *P. gingivalis* and the origin or progression of Alzheimer's disease and also suggests that gingipain inhibitors could be used to treat neurodegeneration in this disease. It has also been shown that typical oral species of the phylum Spirochaetes, including multiple species of the genus *Treponema*, often make up amyloid plaques, and that these organisms are capable of producing additional amyloid-β and amyloid-β protein precursor (Miklossy 2016).

Dysbiosis of the oral microbiome is also implicated in disorders of the endocrine system. In this case, periodontitis appears to be a

potential result of diabetes, as opposed to a potential cause as in most of the diseases discussed above. In one longitudinal study, the prevalence of periodontitis was 60% in subjects with diabetes and 36% in subjects without diabetes (Mealey, Oates, and American Academy of Periodontology 2006). Some of the proposed mechanisms of the influence of diabetes on periodontal health include microangiopathy or alterations in the inflammatory response, collagen metabolism, or the glucose concentrations in gingival crevicular fluid (G. W. Taylor 2001). Nevertheless, it has been proposed that periodontitis can also compound the effects of diabetes by upregulating the production of inflammatory factors like TNF-α, which can act as insulin antagonists (Scannapieco, Dasanayake, and Chhun 2010). Surprisingly, one NGS study showed that subjects with diabetes actually had lower abundances of the typical "red complex" species seen in periodontitis infections, *Porphyromonas gingivalis, Tannerella forsythia,* and *Treponema denticola* (Casarin et al. 2013). The authors note that potential confounders in their study include a higher mean plaque index and greater age in the subjects with diabetes. However, it is also possible that any of the proposed mechanisms of diabetic influence on periodontal health could result in a variation of the typical dysbiotic composition in periodontitis, or that the samples were collected at a stage in the progression of the disease in which the common pathogens are found at lower abundances. In either case, this presents another interesting opportunity for larger scale longitudinal studies, to discover potential alternative pathways to periodontal disease, and thus a greater comprehension of its mechanisms and potential treatments or preventative measures.

1.9 Clinical potentials of the oral microbiome / Manipulations and perturbations of the oral microbiome

We have demonstrated here a few key examples of the wide-ranging action of the oral microbiome upon the human body. This opens vast possibilities for diagnosis and intervention, only some of which have begun to be explored, or even conceptualized. To fully

leverage this potential, we will need to continue to probe the compositions and actions of oral microbiomes, both at small-scale interactions within individuals in different states of health, as well as at broad scales among different populations. The oral microbiome has already shown potential as a diagnostic tool. One machine learning-based study, which collected 2424 publicly available metagenomes from eight studies, showed that the performance of disease predictions was improved when using strain-level features (not feasible with 16S sequencing), and suggested that disease phenotypes are linked to "non-core" microbial genes/factors which may be found in variable genomic regions that are specific to strains/subspecies (Pasolli et al. 2016). However, they caution the use of some potential biomarkers for diagnosing disease, such as the species *Streptococcus anginosus*, which actually associates with general dysbiosis rather than a particular disease. The authors still consider this work to be an early stage of modeling the healthy microbiome, so that it can be used to contrast states of dysbiosis associated with disease. Another attempt at using oral microbial abundances as biomarkers has shown some preliminary success in the early detection of colorectal cancer (CRC) lesions (Flemer et al. 2018). The fecal immune test (FIT) and the fecal occult blood test (FOBT) are typical non-invasive screening procedures used to detect CRC, but they suffer from poor sensitivity in detecting early lesions. The authors showed that a classification model based on oral microbiome samples had high specificity and greater sensitivity than standard tests in distinguishing CRC and polyp samples from healthy controls, with a further increase in sensitivity when combining with fecal microbiome samples. Thus, with further verification studies, the microbiome could be implemented to improve the rate of early detection of CRC.

In addition to a biomarker, the oral microbiome can be both a tool and a target for treating diseases. Commercially available probiotics, which contain live strains from bacterial genera, such as *Bifidobacterium*, *Lactobacillus*, and *Streptococcus*, have been shown to promote greater alpha diversity in the oral microbiome, though without making large-scale or permanent alterations to its

composition (Dassi et al. 2018). As mentioned in the previous section, *Lactobacillus* in particular is a common and effective probiotic in the context of the gut microbiome, but it has been associated with dental caries, and indirectly with CRC. However, it is also possible that this effect is dependent upon the environmental context, as discussed above in the case of *Streptococcus oralis* during cystic fibrosis infection or *Leptotrichia* during pancreatic cancer. Thus, *Lactobacillus* without environmental conditions suitable for dental caries may instead help to promote greater alpha diversity and better oral health. This is another case of preliminary results with interesting potential that require deeper longitudinal study.

Mechanisms have been proposed for the beneficial impact of other potential probiotics as well, to counteract the progression of periodontal disease and caries, including strains of *Streptococcus salivarius*, which has been shown to downregulate inflammatory responses and to stimulate beneficial pathways like type I and II interferon responses (Devine, Marsh, and Meade 2015). Other potential probiotics for this use are *Streptococcus dentisani* (López-López et al. 2017) and *Streptococcus* A12 (Huang et al. 2016), which can buffer the acidic pH produced within cariogenic biofilms through arginine metabolism. The Proteobacteria species *Bdellovibrio bacteriovorus* has been suggested as a tool for potentially targeting periodontal pathogens (Loozen et al. 2015). It feeds on Gram-negative bacteria by invading them, eventually leading to lysis of its prey. Ex vivo experiments on saliva and subgingival plaque showed that *B. bacteriovorus* was able to attack two important oral pathogens, *Fusobacterium nucleatum* and *Aggregatibacter actinomycetemcomitans*, though not certain other desired targets, nor was it specific only to these targets.

Pathways and products of oral microbiota may also be treatment targets. The proinflammatory cytokine IL-17 was shown to be the most strongly upregulated in diabetic subjects and was associated with periodontitis while treatment with anti-IL-17 antibodies was able to mitigate this effect (E. Xiao et al. 2017). In Alzheimer’s disease, inhibitors of the gingipains proteases secreted by

Porphyromonas gingivalis were able to reduce infection by this species in the brain as well as amyloid-β production and neuroinflammation (Dominy et al. 2019).

There is still a lot to learn about the functionality of the oral microbiome, and its potential reaction to treatments. One instance stems from the "WEIRD" bias that has left many of the world's populations out of this area of investigation. A study of "uncontacted" Amerindians in the Venezuelan Amazon, an isolated community with no known exposure to antibiotics, had bacteria carrying functional antibiotic resistance (AR) genes (Clemente et al. 2015). The authors suggest that "AR genes are likely poised for mobilization and enrichment upon exposure to pharmacological levels of antibiotics". The authors emphasize the need for characterizing remote populations "before globalization of modern practices affects potentially beneficial bacteria harbored in the human body". Thus, the context of microbiome studies must continue to be expanded before we can find the optimal approaches to treatment.

1.10 Conclusions and future outlook

Improvements in sequencing techniques and costs continue to propel the field of microbiome research forward, allowing for larger-scale approaches and wider understanding of its structure and function. However, over the last decade or so, much of the research has been exploratory and many investigators have taken their own approaches to perform experiments and analyses as the potentials and the limitations of these new techniques have been probed. Some amount of technical bias has undoubtedly permeated the results in this field as a result. As we begin to understand what NGS experiments show us, and as new approaches are developed, we must strive for standardized methods that will allow for reliable data comparison with minimal bias. The same is true for study designs and sample collection methods, as has been proposed (Woo and Lu 2019). The oral microbiome will benefit especially from concrete methodologies to ensure the proper context for a given study,

because, depending on the goal of study, results may be dependent on particular niches within this habitat.

Current evidence allows for at least some tentative generalizations about the structure of the oral microbiome. Despite the arguments for the focus on gradients of abundances across samples of a population and against the reliance on stomatotypes (or other relevantly termed clusters of microbiome samples), stomatotypes allow for researchers to obtain a broad perspective and can help guide further analyses. The *Neisseria*/*Haemophilus* and *Prevotella*/ *Veillonella* stomatotypes discussed above may represent relatively healthy compositions while the other stomatotypes that have been detected, which tend to be less consistent in terms of composition, may represent dysbiosis. The driver genera of the first two stomatotypes often appear to be associated with health while driver genera in other described stomatotypes are often associated with disease, like *Porphyromonas*, *Rothia*, and certain species of *Streptococcus*. Of course, there are exceptions, such as the associations that have been reported between periodontitis and *Prevotella* (Costalonga and Herzberg 2014) or between dental caries and *Neisseria* (Gross et al. 2010), though to qualify that statement, as mentioned above, there may be alterations in the abundances of many taxa as a result of the disease state, not as a cause of the disease. So, as in the case of *Prevotella* in periodontitis or *Neisseria* in dental caries, these taxa may simply become opportunistic in the environment created by the disease (*Neisseria* may simply be taking advantage of the "hydrogen sink" created by archaeal methanogens and/or *Treponema* species, since its species tend to be acidogenic (Knapp 1988)), and therefore could still be considered to be associated with general oral health. Periodontitis and dental caries are heavily researched diseases, but in seeking associations between the microbiome and any disorder, it is important to thoroughly explore the mechanisms of disease progression before labelling any particular organism as a causative agent. **Table 1.1** and **Table 1.2** provided in this review, which list associated taxa, should be taken as just that: A list of potential associations to be further explored.

As has been emphasized throughout this review, our current conceptions of the oral microbiome are largely biased toward the lifestyles in “WEIRD” nations (Western, educated, industrialized, rich, and democratic) (Henrich, Heine, and Norenzayan 2010), as most of the investigators and studied samples come from these populations. The few studies referenced here that have incorporated non-Westernized samples highlight the holes in our current knowledge of the potential compositions and structures of the oral microbiome (Pasolli et al. 2019) and its functionality in particular contexts (Clemente et al. 2015). As researchers integrate more diverse populations into the field (as well as expanding the focus to include other domains of life present in the microbiome), they will be able to continually generate more extensive databases of human oral-associated taxonomy and functionality, allowing for more comprehensive studies.

Its position as the gateway to some of the most vital functions for human life tightly connects the oral cavity to the rest of the body and gives it a powerful influence on many of the body’s processes. Dysbiosis in the oral microbiome is clearly linked to a number of local and systemic diseases, though some of the particular associations and mechanisms of action remain conjectural pending further study. Some of the treatment examples discussed here appear straightforward, even despite the need for deeper verification of efficacy. However, the human microbiome is a multi-dimensional interconnected micro-biosphere, and perturbations may have as yet unforeseen effects, whether undetected in the short term or undeterminable in the long term without the appropriate study designs. We have already mentioned examples where unrelated organisms can fill some of the same functional niches (see the discussion above of *Treponema* and methanogenic archaeal species acting as hydrogen sinks in periodontitis) and their competition can be implicated in disease. Microbiome research is far from complete, and without a deep understanding of its nature, we should exercise caution and patience before widely exploiting its medical potential. This will require many more large-scale and longitudinal studies, greater focus on the functional component of the microbiome, and a stronger characterization of the varying structures of each

microbiome of the body in different contexts, part of which will be to combat the WEIRD bias, which has thus far limited the scope of study, and the dearth of attention to the less prominent domains of the microbiome, like fungi, archaea, and viruses.

Chapter 2: Citizen science charts two major “stomatotypes” in the oral microbiome of adolescents and reveals links with habits and drinking water composition

Willis JR, González-Torres P, Pittis A, Bejarano LA, Cozzuto L, Andreu-Somavilla N, Alloza-Trabado M, Valentín A, Ksiezopolska E, Company C, Onywera H, Montfort M, Hermoso A, Iraola-Guzmán S, Saus E, Labeeuw A, Carolis C, Hecht J, Ponomarenko J & Gabaldón T. (2018). Citizen science charts two major ‘stomatotypes’ in the oral microbiome of adolescents and reveals links with habits and drinking water composition. *Microbiome*. 6(1): 218. doi: 10.1186/s40168-018-0592-3

Chapter 2: Citizen science charts two major "stomatotypes" in the oral microbiome of adolescents and reveals links with habits and drinking water composition

2.1 Abstract

Background: The oral cavity comprises a rich and diverse microbiome, which plays important roles in health and disease. Previous studies have mostly focused on adult populations or in very young children, whereas the adolescent oral microbiome remains poorly studied. Here, we used a citizen science approach and 16S profiling to assess the oral microbiome of 1500 adolescents around Spain and its relationships with lifestyle, diet, hygiene, and socioeconomic and environmental parameters. **Results**: Our results provide a detailed snapshot of the adolescent oral microbiome and how it varies with lifestyle and other factors. In addition to hygiene and dietary habits, we found that the composition of tap water was related to important changes in the abundance of several bacterial genera. This points to an important role of drinking water in shaping the oral microbiota, which has been so far poorly explored. Overall, the microbiome samples of our study can be clustered into two broad compositional patterns (stomatotypes), driven mostly by *Neisseria* and *Prevotella*, respectively. These patterns show striking similarities with those found in unrelated populations. **Conclusions**: We hypothesize that these stomatotypes represent two possible global optimal equilibria in the oral microbiome that reflect underlying constraints of the human oral niche. As such, they

should be found across a variety of geographical regions, lifestyles, and ages.

Keywords: Oral microbiome, Metagenomics, Stomatotypes, Tap water composition

2.2 Introduction

The oral cavity is among the most heavily colonized areas of the human body and harbors the second most diverse human microbiome (M. Kilian et al. 2016). Previous studies of the oral microbiome have estimated the presence of around 10^8 microbial cells per milliliter of saliva, and the presence of up to 700 distinct prokaryotic taxa, of which approximately one third cannot be cultured (H. Chen and Jiang 2014; E. Hajishengallis et al. 2017; He et al. 2015). The mouth is also the site where the most prevalent human diseases occur, including caries, gingivitis, and periodontitis (M. Kilian et al. 2016; Belstrøm, Constancias, et al. 2017; Kageyama et al. 2017). In addition, given the close connections of the oral cavity with the vascular system and the digestive and respiratory tracts, alterations of the mouth microbiota have been related with diseases that affect other body parts, such as diabetes or cardiovascular disease (M. Kilian et al. 2016; He et al. 2015). Understanding the composition of the oral microbiome across individuals, and how it relates with lifestyle habits such as diet or hygiene, is important to achieve a proactive management of oral health. The analysis of the microbiome through the next-generation sequencing of 16S amplicons (i.e., 16S metabarcoding) offers a cost-effective approach to assess the overall composition of an individual’s microbiome (Ji-Hoi Moon and Lee 2016; McLean 2014). Previous studies have assessed the oral microbiome in relation with factors such as biogeography, environment, age, or

ethnicity (Ji-Hoi Moon and Lee 2016; Takeshita et al. 2016, 2014), or have focused on the effect of smoking (J. Wu et al. 2016; Yu et al. 2017), diet (Berni Canani, Gilbert, and Nagler 2015; Cuervo et al. 2016; Jose, Padmanabhan, and Chitharanjan 2013; De Filippis et al. 2014), or hygiene habits (Klaus et al. 2016; Al-Mulla et al. 2010; Koopman et al. 2015). On the more clinical side, some studies have uncovered alterations of the oral microbiota in prevalent diseases of the oral cavity including periodontal disease (Richards et al. 2017) and caries (E. Hajishengallis et al. 2017; Costalonga and Herzberg 2014). In addition, previous studies suggest that intrinsic physiological parameters of the host such as enzymatic content of saliva relate to variations in the microbiome (Zaura et al. 2017). Although the mouth comprises several distinct niches, previous large-scale studies have mostly probed microbial composition of saliva. This fluid can gather bacteria and metabolites that originate from other oral niches, and appear to be representative of the overall oral microbiome (Yamashita and Takeshita 2017). Furthermore, considering that saliva tests offer an ideal non-invasive source for diagnosis, relationships of its microbial composition with the presence of several diseases such as cancer have been investigated (Patil and Patil 2011; C.-Z. Zhang et al. 2016; Galloway-Peña et al. 2017). Most previous studies have focused on adults, or very young infants, with studies on adolescents lagging behind. The largest dataset on adolescents so far corresponds to a longitudinal study of 107 individuals, including 27 monozygotic and 18 di-zygotic twin pairs (Stahringer et al. 2012). This study suggested that environment is the main determinant of the oral microbiome with differences between mono- or di-zygotic twins not being significant. Here, we used a citizen science approach and 16S metabarcoding to assess the composition of the microbiome of the oral cavity among teenagers in Spain. We studied its variation with more than 50 parameters including geographical location, gender, and urban environment, as well as several dietary and hygiene habits. Our

study showcases the use of a citizen science approach to generate hypotheses that can be further validated in subsequent studies.

2.3 Material and methods

2.3.1 Sample collection

All participants, and at least one of their parents or legal guardians for those under the age of 18, signed a consent form to use their saliva samples for microbiome research. This consent form and the purpose of this project received approval by the ethics committee of the Barcelona Biomedical Research Park (PRBB). The target population was teenagers in the third course of Spanish secondary compulsory education (ESO), ages 13–15 years old. Additionally, we also collected samples from teachers of the participating schools. Schools were selected among those which volunteered to cover a broad range of Spanish provinces, a similar number of schools in urban (towns or cities with more than 50,000 inhabitants) or rural (towns with less than 50,000 inhabitants and more than 50 km away from a large town) environments. Samples were collected during February to April in 2015. Participants were asked not to eat for 1 h prior to the sample collection. We tried to minimize variability as much as possible. To minimize sample collection variability, all donors received clear instructions on the procedure in person and sample collection was performed with the assistance of one researcher involved in the project, after a clear demonstration. All participants responded to a uniform questionnaire (see below). Before sample collection, saliva pH was measured using pH test strips (MColorpHast, Merck, range 5.0–10.0; 0.5 accuracy units). Although the use of pH test strips have been validated extensively (Cocco et al. 2017), we validated our chosen strips. For this, we compared values given by eight different researchers using these strips to a scale of solutions with different pH to the values provided by a PHmeter (SevenEasypH model, Mettler-Toledo (GmbH). The correlation was high ($R2 = 0.96$), with average absolute differences

between the value of the pH meter and that provided by the researcher being 0.33 which is within the range of the limit of detection of the method (0.5). Saliva samples were collected using a mouthwash and using a protocol that had been previously tested and compared with other procedures during a pilot phase of the project. Of note, this procedure is used in other oral microbiome studies and has been shown to produce consistent results with other sampling procedures (Y. Lim et al. 2017). The protocol used is as follows: Study participants rinsed their mouth with 15-mL sterile phosphate-buffered saline (PBS) for 1 min and subsequently returned the liquid into a 50-mL centrifuge plastic tube. The collected samples were centrifuged at 4500 g for 12 min at room temperature (r.t.) in an Eppendorf 5430 centrifuge equipped with an Eppendorf F-35-6-30 rotor. Pellets were resuspended with PBS, transferred to 1.5-ml eppendorf tubes and centrifuged at 4500 g for 5 min at r.t. using an Eppendorf FA-45-24-11-HS rotor. Supernatants were discarded, and pellets were frozen and kept at − 20 °C until the time of analysis.

2.3.2 DNA extraction and sequencing

DNA from samples was extracted individually using the ZR-96 Fungal/Bacterial DNA kit (Zymo research Ref D6006) following manufacturer's instructions. The extraction tubes were agitated twice in a 96-well plate using Tissue lyser II (Qiagen) at 30 Hz/s for 5 min at 4 °C. As a control for downstream procedures, we also used two DNA samples derived from bacterial mock communities obtained from the BEI Resources of the Human Microbiome Project: Each sample contained genomic DNA of ribosomal operons from 20 bacterial species. The HM-782D community contained an even number of ribosomal DNA per species (100,000 operons per species). The HM-783D community contained a variable number of operons, ranging from 1000 to 1000,000 per species.

DNA samples were diluted to 12.5 ng/µl and used to amplify the V3–V4 regions of 16S ribosomal RNA gene, using the following

universal primers in a limited cycle PCR:

V3-V4-Forward

(5′-
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACG
GGNGGCWGCAG-3′)

V3-V4-Reverse

(5′-
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTA
CHVGGGTATCTAATCC-3′)

The PCR was performed in 10-µl volume with 0.2-µM primer concentration. Cycling conditions were initial denaturation of 3 min at 95 °C followed by 20 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, ending with a final elongation step of 5 min at 72 °C. After this first PCR step, water was added to a total volume of 50 µl and reactions were purified using AMPure XP beads (Beckman Coulter) with a 0.9× ratio according to manufacturer's instructions. PCR products were eluted from the magnetic beads with 32 µl of Buffer EB (Qiagen) and 30 µl of the eluate were transferred to a fresh 96-well plate.

The above described primers contain overhangs allowing the addition of full-length Nextera adapters with barcodes for multiplex sequencing in a second PCR step, resulting in sequencing ready libraries with approximately 450 bp insert sizes. To do so, 5 µl of the first amplification were used as template for the second PCR with Nextera XT v2 adaptor primers in a final volume of 50 µl using the same PCR mix and thermal profile as for the first PCR but only 8 cycles. After the second PCR, 25 µl of the final product was used for purification and normalization with SequalPrep normalization kit (Invitrogen), according to manufacturer's protocol. Libraries were eluted in 20-µl volume and pooled for sequencing. Final pools were quantified by qPCR using Kapa library quantification kit for Illumina Platforms (Kapa Biosystems) on an ABI 7900HT real-time cycler (Applied Biosystems). Sequencing was performed in eight runs on an Illumina MiSeq with

2 × 300 bp reads using v3 chemistry with a loading concentration of 10 pM. In all cases, 15% of PhIX control libraries was spiked in to increase the diversity of the sequenced sample. Negative controls of the sample collection buffer, DNA extraction, and PCR amplification steps were routinely performed in parallel, using the same conditions and reagents. Our controls systematically provided no visible band or quantifiable DNA amounts by gel visualization or Bioanalyzer, whereas all of our samples provided clearly visible bands after 20 cycles. Four such controls were subjected to library preparation and sequenced. Expectedly, these sequenced non-template controls systematically yielded very few reads (a range of 30–880 reads per sample), in contrast to an average of 54,000 reads/library in sample-derived libraries.

2.3.3 Pre-processing of 16S rRNA sequence reads and operational taxonomic unit assignment

The specific pipeline and parameters were set using sequence reads from both 16S rRNA amplicon and whole genome sequencing of the described mock communities. In the final adopted pipeline, reads were checked for quality using FastQC (Andrews 2010). 16S amplicons were analyzed by Mothur v1.34.4 (Schloss et al. 2009) following instructions described in the author's website (https://www.mothur.org/wiki/MiSeq_SOP). Overlapping pairs of sequence reads were assembled, contigs with more than 4 ambiguities and shorter than 439 bp or larger than 466 bp were discarded, and the remaining contigs were aligned to the reference alignment provided by the SILVA database (Pruesse et al. 2007) (version 119) with a k-mer size of 8. Artifacts from the alignment and the contigs with more than 12 homo-polymers (the maximum number found in the reference alignment) were removed. The resulting alignment was simplified by removing the columns containing only gaps and by discarding duplicated sequences. The aligned sequences were then grouped allowing up to 4 mismatches and clusters with only one sequence were removed. Uchime

(embedded in the Mothur framework) was used to remove chimeras, and the resulting sequences were classified according to the taxonomy into the corresponding operational taxonomic units (OTUs). Undesired lineages such as chloroplast, mitochondria, archaea, eukaryota, and "unknown" were removed. Sequences were then grouped again into OTUs by using the cluster.split command and considering the genus level. Finally, OTUs mapping to the same genus were grouped together.

2.3.4 Microbiome composition profiling

The 16S rRNA OTU counts from the 1532 samples in this study for which we also had survey data were stored and analyzed using the R package Phyloseq (version 1.16.2) (McMurdie and Holmes 2013), which also has functions for filtering operational taxonomic units (OTUs), normalizing values, and various other calculations. One hundred eighty samples from 5 of the schools had to be removed due to an apparent batch effect during the sequencing procedure. This batch effect was detected in the initial quality assessment of the comparison of the data. In a diversity analysis these samples behaved very distinctly from the rest of the sample showing very low diversity values and corresponded to samples that had been processed and sequenced as a batch on the same day. Additionally, 33 samples were removed from the analyses because of errors with the sample identifiers, leaving a total of 1319 samples from 35 different schools from around Spain. Three hundred thirty-two different genera were identified in these samples. The 16S counts were normalized per sample, leaving the relative abundance of each genus within that sample, with all values between 0 and 100.

2.3.5 Diversity measures

We estimated alpha diversity as measured by Shannon Diversity Index and Simpson Diversity Index (Morris et al. 2014) with the estimate_richness function from the Phyloseq package v1.16.2. We

estimated beta diversity as the weighted and unweighted Unifrac distance between samples with the Unifrac function, as well as the Jensen-Shannon Divergence (JSD) with the JSD function, both from the Phyloseq package. In addition, we calculated the Bray-Curtis dissimilarity and Canberra index using the vegdist function in the vegan package (version 2.4.6) (Oksanen et al. 2017). Both unifrac calculations require a phylogenetic tree which indicates phylogenetic distances by branch lengths. We obtained the tree by following the procedure described by Callahan et al. (Ben J. Callahan et al. 2016), wherein sequences are aligned, then using the R package phangorn (version 2.4.0), we construct a neighbor-joining tree and then fit a maximum likelihood tree. The weighted unifrac distance adds weights to the branch lengths based on relative abundance, while the unweighted unifrac distance considers only the presence or absence of OTUs. For each of these alpha and beta diversity measures, we also divided samples into quartiles in order to label each sample as having low (1st quartile), average (2nd and 3rd quartiles), or high diversity (4th quartile).

2.3.6 Sample clustering

To cluster the samples in terms of their taxonomic composition (stomatotypes), we adapted the procedure described previously (Arumugam et al. 2011) for the determination of enterotypes, which we here refer to as stomatotypes. For this, we employed each of five beta diversity measures—Jensen-Shannon Divergence (JSD), weighted and unweighted UniFrac distance, Bray-Curtis dissimilarity, and Canberra index—to produce distance matrices for the genera of all samples and then Partitioning Around Medoids (PAM) clustering to group samples with similar overall oral microbiomes. Next, we used the Calinski-Harabasz (CH) index (Caliński and Harabasz 1974) to determine the optimal number of clusters, and we further verified this by calculating the average silhouette width of the samples, which is a measure of the separation of samples within one cluster from those of another cluster, as well as the prediction strength, another measure of the efficiency of clustering. The functions for these calculations come

from the R packages cluster v2.0.6 (https://cran.r-project.org/package=cluster), clusterSim v0.45-1 (https://cran.r-project.org/package=clusterSim), and fpc v2.1-11.1 (https://CRAN.R-project. =org/package=fpc). Clustering was validated using all five distance measures to ensure proper clustering, but analyses here are performed using the clustering based on JSD. As detailed in Bork's group tutorial (http://enterotype.embl.de/enterotypes.html), we used the R package ade4 v1.7-4 (https://cran.r-project.org/package=ade4) for visualization. We first excluded those genera that are potentially noisy, removing those for which the average relative abundance across all samples was lower than 0.01%. We then used Between Class Analysis (BCA) to determine the "drivers" for each stomatotype, which are the genera accounting for the greatest separation between samples of a given stomatotype from the other types. We used a Principal Coordinate Analysis (PCoA) to visualize the clustering of the samples within their respective stomatotypes. Furthemore, the adonis function in the vegan package was used to perform a PERMANOVA test on each beta diversity measure to ensure significant separation of stomatotypes.

2.3.7 Gradients of abundances

The gradients of abundances were displayed using the same coordinates in the PCoA plots described above, and points were colored based on abundances of the indicated taxa binned into every 10th percentile of those abundances. Shapes of points are determined by the stomatotype based on a given distance measure, typically the JSD measure in figures here.

2.3.8 Co-occurrence networks

To produce co-occurrence networks of genera within a given stomatotype, we use the R packages sna v2.4 (https://cran.r-project.org/package=sna) and network v1.13.0 (https://cran.r-project.org/package=network). We first calculated Pearson

correlations between pairs of genera within samples of a given stomatotype and used the Bonferroni correction to adjust the *p* values. Then, considering the 20 most common genera within the samples of a given stomatotype, we produce a network wherein edges are formed between only those genera that have a correlation coefficient greater than 0.25 or less than − 0.25 and an adjusted *p* value less than 0.05. Red edges indicate positive correlations, blue edges indicate negative correlations and edge width is proportional to the absolute value of the correlation coefficient. Vertex color is based on the phylum to which the given genus belongs.

2.3.9 Questionnaire and other metadata

Participants were asked to answer one questionnaire inquiring about aspects relevant to their hygiene and dietary habits. These questions were adapted from questionnaires available at the PhenX toolkit (consensus measures for Phenotypes and eXposures), which provides recommended standard data collection protocols for conducting biomedical research (Hendershot et al. 2015) and which has been recommended by the microbiome research community (Huttenhower et al. 2014). In addition, some of the questions were selected among those suggested by citizens themselves through the project's website. The final questionnaire is available at (**Additional file 2.2**). Data on average socioeconomic status of each participating high school was obtained as follows. We first assigned geographic coordinates to all schools based on their postal address, which were used to assign socioeconomic values from their districts using the GIS (Geographic Information System) software QGIS v.2.14 and based on the Census Tracts of 2001, of the Urban Vulnerability Atlas Database from the Spanish government (http://www.fomento.gob.es/MFOM/LANG_CASTELLANO/DIRECCIONES_GENERALES/ARQ_VIVIENDA/SUELO_Y_POLITICAS/OBSERVATORIO/Atlas_Vulnerabilidad_Urbana/). Data on tap water hardness was obtained from several national ionic composition studies (Vitoria et al. 2015; Maraver, Vitoria, Ferreira-Pêgo, et al. 2015; Maraver, Vitoria, Almerich-Silla, et al. 2015).

2.3.10 Statistical analyses

We obtained the Pearson correlation coefficient between abundances of pairs of genera, between genera and other continuous variables (i.e., questionnaire answers, pH), and between pairs of variables. We performed the Kruskal-Wallis rank sum test between categorical variables (i.e., questionnaire, stomatotype) and abundances or other continuous variables. In those cases where the Kruskal-Wallis test was statistically significant, the differential groups and the direction of their difference (greater or less than other groups) was determined by ANOVA using the aov and TukeyHSD functions from the base R package stats v3.4.1. We also performed chi-squared tests between categorical values as well as between those variables and the presence/absence of OTUs. In all cases, we applied the Bonferroni correction to adjust the *p* values by the number of comparisons. Correlation heatmaps, boxplots, and volcano plots were generated using ggplot2 v2.2.1 (https://cran.r-project.org/package=ggplot2), and association plots were generated using the assoc function from the R package vcd v1.4-3 (https://cran.r-project.org/package=vcd). In general, all of our statistical analyses considered all 1319 samples, except for the instances that are specifically mentioned in the text (i.e., by referring to a correlation affecting students), we did so with subsets of the samples, including students only (1297 of the 1319 samples) or those samples not drinking primarily from bottled water (814 of the 1319 samples). To assess the robustness of correlations with pH to stochastic variations within the precision range of the measurements, we performed a computational test, changing measured pH value of each saliva sample to a random number within the precision range (± 0.5). We repeated this 1000 times and measured whether reported significant correlations were still existing. For all reported correlations, they remained in 100% of the cases.

2.3.11 Distribution maps

We produced maps with distributions of various values using shape

files for Spain obtained from the GADM database of Global Administrative Areas (http://gadm.org/). We used the readShapeSpatial function from the R package maptools v0.9-2 (https://cran.r-project.org/package=maptools) which creates a Spatial DataFrame object that can be used to plot values in different regions of a map, and the boxed.labels function from the R package plotrix v3.6-6 (https://cran.r-project.org/package=plotrix) to include labels for regions on the figure.

2.4 Results

2.4.1 Data collection and analysis

One thousand five hundred fifty-five samples were collected from students (ages 13–15) and their teachers in 40 schools around Spain during Spring 2015 [see **Additional file 2.1**]. Sample collection was coupled to science communication activities aiming to raise awareness about the role of the microbiome in health and disease, the potential of sequencing and bioinformatics technologies, and the scientific career (see http://www.sacalalengua.org). Donors were asked to answer a questionnaire, including 54 questions [see **Additional file 2.2**], some of which proposed by citizens, about their health, and their dietary and hygiene habits. The pH of the donor's saliva was measured prior to sample collection. Samples were obtained using oral rinse, from which cells were collected and frozen (see "Online methods"). DNA extracted from the samples were subjected to 16S profiling of the V3–V4 regions, using Illumina MiSeq technology, and processed bioinformatically (see **Materials and methods**). Data from 1319 samples that passed all the quality filters were explored in terms of the relationships of the microbiome composition, the questionnaire results, and other metadata (see **Materials and methods**).

2.4.2 Oral microbiome diversity is structured into two major stomatotypes

Our analyses provide a snapshot of the microbial diversity in oral samples in young adolescents across Spain, and do so with unprecedented scale and resolution (**Figure 2.1**). Overall, we identified 332 operational taxonomic units (OTUs) at the genus level in our dataset. Thirty-two genera were common, appearing in 75% or more of the sampled individuals. This “core” set comprised typical oral bacteria. The top ten most abundant genera represented collectively 84.64% of the analyzed sequences and were present in 99.6% of the samples. *Streptococcus* was the most abundant genus in most (68%) samples and showed an average relative abundance of 22.3%, followed by *Prevotella* (11.9%), *Haemophilus* (11.4%), *Neisseria* (10.1%), and *Veillonella* (9%). This core community composition and distribution is consistent with previous studies of oral healthy microbiomes (H. Chen and Jiang 2014; Takeshita et al. 2016; Zaura et al. 2017, 2009; Human Microbiome Project Consortium 2012; Wade 2013). For instance, 20 of our 32 common genera are also common in a recent study of the oral microbiome of 2343 adults in Hisayama (Japan) (Takeshita et al. 2016). Similar to previous oral microbiome surveys (Ji-Hoi Moon and Lee 2016; Human Microbiome Project Consortium 2012), we found high alpha (within sample) diversity (mean Shannon diversity 2.5, see **Additional file 2.3**) and low beta (between samples) diversity (mean weighted UniFrac distance 0.118). Overall correlations among taxa across all samples revealed several clusters of co-occurring genera that hint to underlying ecological interactions (**Figure 2.2**, **Additional file 2.4**). For instance, strong co-occurrence links *Leptotrichia*, *Actinomyces*, and *Prevotella*, and this latter one with *Veillonella*, suggesting they may be ecologically related. Genera in this cluster tend to anti-correlate with *Haemophilus*, *Porphyromonas*, and *Gemella*. Previous studies have

shown that individual microbiomes from certain niches can be clustered into different types, such as the enterotypes of the gut microbiome (Arumugam et al. 2011). Using this approach on our data (see **Materials and methods**) results in two major clusters, which we here refer to as “stomatotypes” in analogy to the enterotypes of the gut microbiome. The two defined stomatotypes differ in their microbial composition and abundance covariations, and for which *Neisseria* (stomatotype 1) and *Prevotella* (stomatotype 2) are the genera driving most differences (**Figure 3**). Other differences include higher proportions of *Haemophilus* in stomatotype 1 and higher proportions of *Veillonella* and *Streptococcus* in stomatotype 2. Importantly, although the studies are performed with different methodologies and have largely different target populations, we noted a strong parallelism between our two stomatotypes and the defined “coinhabiting groups” in the abovementioned Hisayama study (Takeshita et al. 2016). Of note, other studies of the oral microbiome have found a different number of clusters. An analysis of the oral microbiome in 268 healthy young adults (18–32) classified the samples into five discrete clusters (Zaura et al. 2017), whereas another study of 161 healthy adults found three different clusters (De Filippis et al. 2014). Yet, many parallels can also be found between our stomatotypes, and those in these studies. In terms of the driving species, our stomatotypes 1 and 2 are similar, respectively, to MIC1.3 and MIC2 of the 268 adults study and to clusters 1 and 2 of the 161 adults study. These striking similarities between disparate studies suggest that these two major stomatotypes may be ubiquitous and define global equilibria in the human mouth microbiome. As we discuss below, these stomatotypes are not discrete, well separated entities, but rather represent two poles of a gradient of microbial compositions. The stomatotypes are driven mostly by certain abundant genera, but do not explain the variability found in many other genera. This is apparent when plotting the abundance of

different genera onto the principal coordinate analyses (**Figure 2.4**).

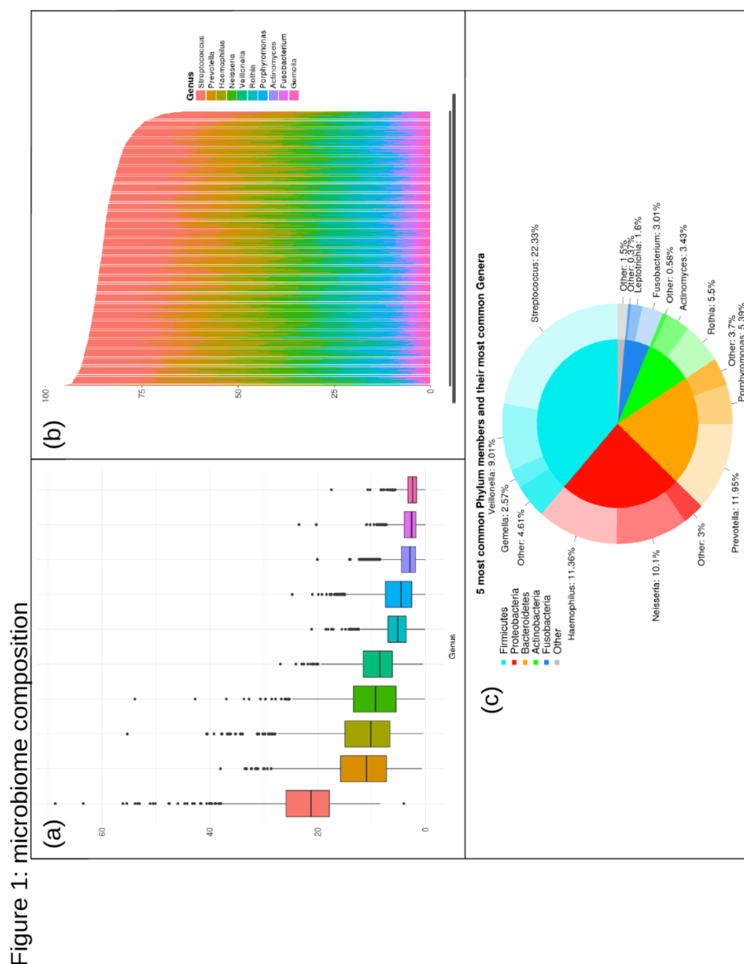

Figure 2.1: Microbiome composition. **a** Box plots of the relative abundances of the ten most common genera. **b** Stacked bars of relative abundances of the ten most common genera for all samples, showing the relative proportion of all samples made up of these ten genera. Stacked white bars are meaningless and appear due to lack of image resolution. **c** Donut chart showing the five most common phyla (inner ring) and the most common genera (outer ring) within each phylum with the average relative abundance per sample.

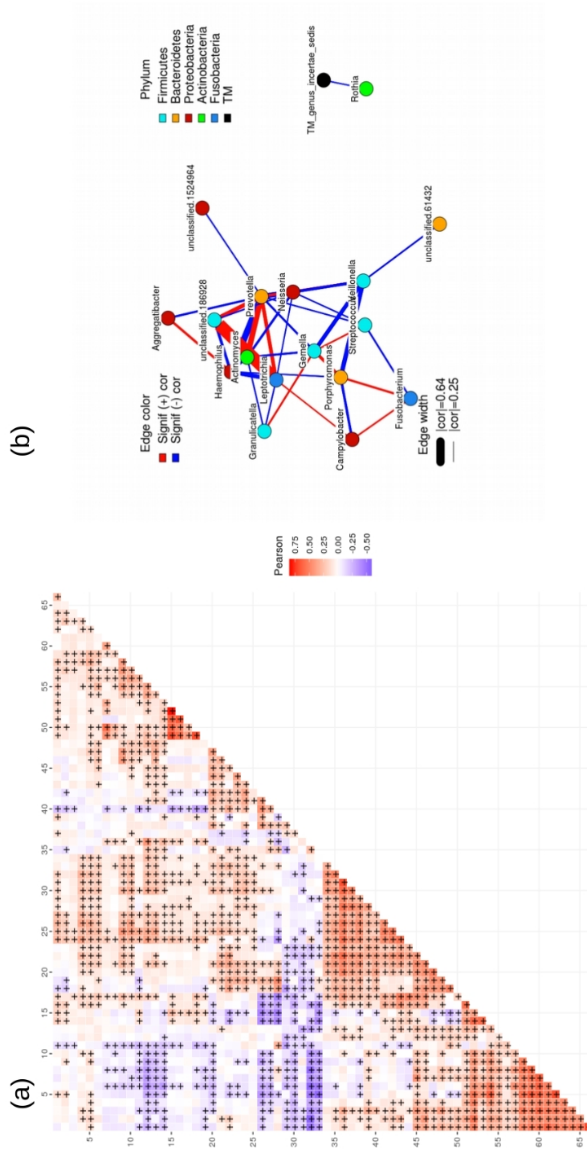

Figure 2.2: Correlations among genera (all samples). **a** Heatmap of correlations between relative abundances of genera. Color indicates Pearson correlation coefficient and "+" indicates a statistically significant correlation. While 332 different genera in total were detected, for the sake of visual representation, this figure shows only the 67 genera which were present in at least 1/3 of all samples (436). Correlation coefficient values for significant correlations can be found in

Additional file 2.4. The indexes of genera within **Additional file 2.4** are marked at every fifth position in the figure here so that names can be matched to the figure if so desired. **b** Co-occurrence network of the 20 most common genera. Edges indicate significant positive (red) or negative (blue) correlations between indicated genera. Edge width is proportional to Pearson correlation coefficient. Only displaying edges for coefficients of 0.25 or greater and − 0.25 or lower. The largest and smallest edge widths are shown with the corresponding absolute value of the correlation coefficient as it appears in the figure.

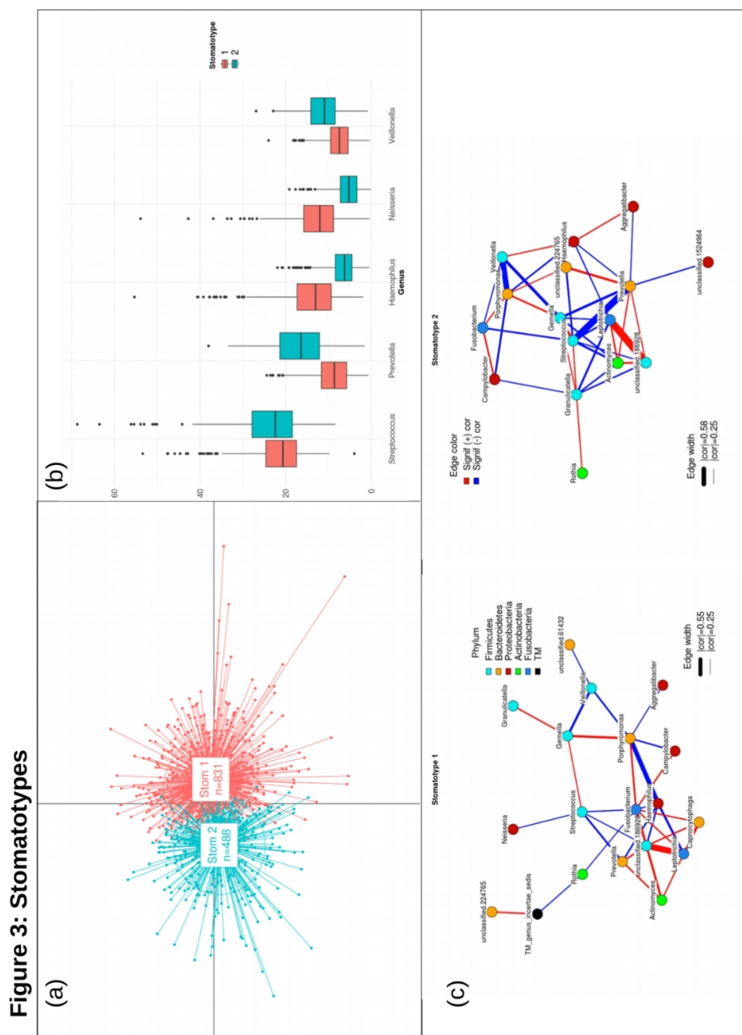

Figure 2.3: Stomatotypes. **a** Principal Coordinates Analysis (PCoA) of samples using a Jensen-Shannon Divergence (JSD). Shows that the samples cluster into 2 groups (stomatotypes). **b** Boxplots of relative abundances of the five most common genera in samples with stomatotype 1 (red) and stomatotype 2 (blue). Bonferroni-adjusted *p* values from Wilcoxon tests between samples of stomatotypes 1 and 2 for streptococcus is 1.1e−7, while the values for the other 4 genera here were all less than 2e−16. **c** Co-occurence networks of 20 most common genera within samples of Stomatotypes 1 and 2 separately. Edges indicate significant positive (red) or negative (blue) correlations between indicated genera. Edge width is proportional to Pearson correlation coefficient. Only displaying edges for coefficients of 0.25 or greater. The largest and smallest edge widths are shown with the corresponding absolute value of the correlation coefficient as it appears in the figure.

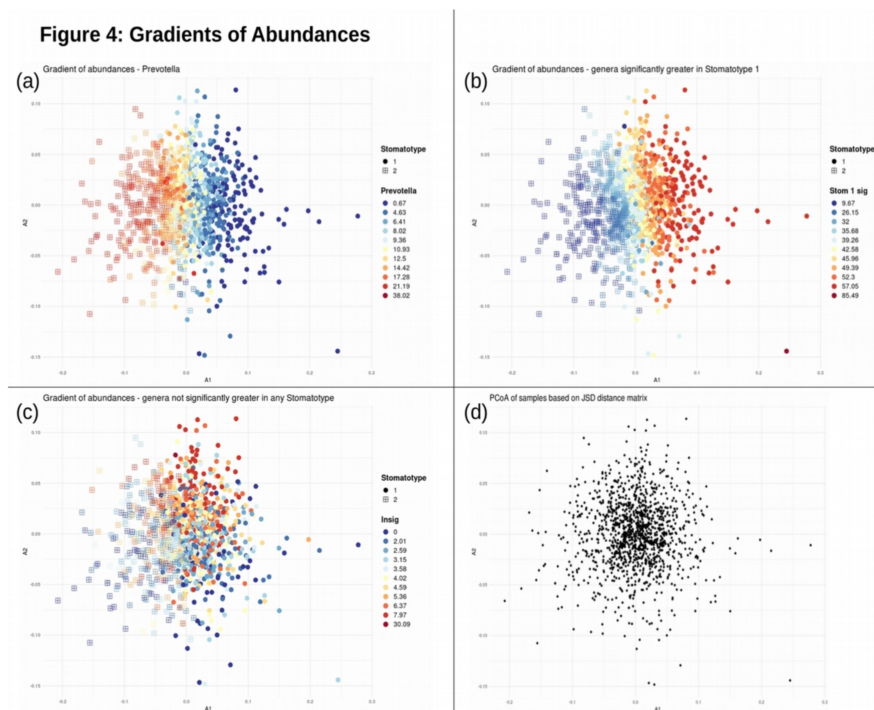

Figure 2.4: Gradients of abundances. Principal Coordinates Analysis (PCoA) of samples using a Jensen-Shannon Divergence (JSD) shows points with the same coordinates as in **Figure 2.3a**. Circles indicate samples of stomatotype 1 and squares indicate samples of stomatotype 2. Colors represent abundance of the

indicated genus (or the sum of abundances of indicated genera) for a given sample, where red is higher and blue is lower, with values indicated in the legend to the right. **a** Abundances for the genus *Prevotella*. **b** Sum of abundances of the 19 genera that were found to have significantly higher abundances in samples of stomatotype 1. **c** Sum of abundances of the 275 genera that did not have significantly higher abundances in samples of either stomatotype over the other. **d** Points without color or shape in order to display the spread of samples within the PCoA.

Studying co-occurrence patterns for each stomatotype separately reveals underlying bacterial communities that are shared or specific (**Figure 2.5**). In both stomatotypes, *Streptococcus* is positively correlated with *Gemella* and negatively with *Prevotella and Fusobacterium* (**Figure 3c**). This observation fits with recent studies describing oral plaque formation and evolution in dysbiotic processes that have led to the formulation of the “ecological plaque hypothesis” (M. Kilian et al. 2016; Jakubovics 2015; Takeshita et al. 2015; Mahajan et al. 2013). In this model, *Streptococcus*, *Gemella*, and *Neisseria* are among the pioneer colonizers that contribute to initial plaque formation. These genera are replaced in further evolution of the plaque by anaerobic species of several genera, including *Prevotella*, *Porphyromonas*, *Fusobacterum*, and *Veillonella*. Thus, the abundance covariations observed in both stomatotypes may partly reflect the underlying diversity of biofilm succession stages in our samples and would support the main axis of previously observed core community changes in dental plaque. Positive correlations between *Porphyromonas* and *Fusobacterium* and negative correlations between *Veillonella* and *Gemella* further support this model, while positive correlations between *Porphyromonas* and *Gemella* and negative correlations between *Porphyromonas* and *Veillonella* would not be explained by the current plaque succession model. Of note, several of the correlations mentioned in our study coincide with those found in previous studies (De Filippis et al. 2014; Zaura et al. 2017). Microbial

compositions in oral rinse samples can only be considered a proxy for plaque communities, as the procedure collects cells from different oral niches. However, earlier studies using similar collection protocols and including information on plaque status or dental health have found correlations between microbial composition of saliva, the amount of plaque and diseases such as periodontitis, or caries (Belstrøm, Constancias, et al. 2017; Takeshita et al. 2016; Belstrøm, Holmstrup, et al. 2017; J. Zhou et al. 2016).

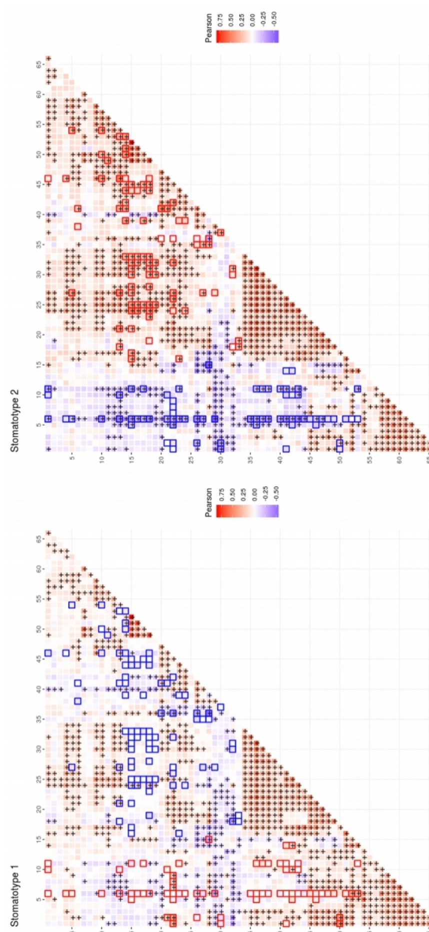

Figure 5: Correlations among Stomatotypes

Figure 2.5: Correlations among stomatotypes. Heatmap of correlations between relative abundances of genera in samples with stomatotype 1 (left) and stomatotype 2 (right). Color indicates Pearson correlation coefficient and "+" indicates a statistically significant correlation. Highlighted boxes indicate genera pairs for which the correlation coefficient in the given stomatotype is at least 0.2 greater (red) or lower (blue) than the correlation coefficient in the other stomatotype. While 332 different genera in total were detected, for the sake of visual representation, this figure, as in **Figure 2.3**, shows only the 67 genera which were present in at least 1/3 of all samples (436). Row and columns are ordered as in **Figure 2.2** and thus can also be compared with **Additional file 2.4** in the same manner mentioned in **Figure 2.2**.

Although many of the covariations between the two stomatotypes are similar, their strengths can be markedly different. In addition, some covariations appear specific for each stomatotype. For instance, in the case of stomatotype 1, we detected positive covariation of *Fusobacterium* and *Capnocytophaga*, both anaerobic bacteria implicated in dental plaque progression (Takeshita et al. 2015), while in stomatotype 2, we specifically detect antagonism between *Streptococcus* and *Actinomyces*, which are known to compete in the initial phases of dental plaque formation (Jakubovics 2015; Dige et al. 2009). Thus, the two stomatotypes may point to differences in the relative impact of underlying processes and microbial communities that differentially affect individuals in our study.

2.4.3 Lifestyle and social parameters

We next explored correlations between social parameters, questionnaire answers, and microbial composition [see **Additional files 2.5, 2.6, 2.7,** and **2.8**]. We found that living in rural or urban areas did not correlate with significant changes in the microbiome. This suggests that diets and lifestyles of students are similar in cities and rural areas in Spain, as confirmed by our questionnaire, which

only revealed significant differences in terms of a higher likelihood of having dogs for students living in the countryside. Socioeconomic status did correlate significantly with the abundance of some genera, positively with *Rhizobium* and negatively with *Bradyrhizobium*, *Acinetobacter*, and *Pseudomonas.* Here, some differences in dietary habits were found, with a lower socioeconomic status being correlated with higher consumption of coke and sweets among students. We found no large differences between oral microbiomes of males and females, with only two genera (*Actinomyces* and *Oribacterium*) showing significantly different abundances (both higher in males). Boys and girls had some different habits. While the former tended to drink more milk, coke, or energetic drinks, the latter chewed gum and brushed their teeth more often. Larger differences in the oral microbiome were found between students and their teachers. Teachers' microbiomes were enriched in *Alloscardovia*, *Parascardovia*, *Filifactor*, *Bulleidia*, *Mycoplasma*, *Phocaeicola*, *Hallella*, *Howardella*, *Anaeroglobus*, *Dialister*, *Desulfobulbus*, and *Campylobacter*, while those of students were enriched in *Actinomyces*, *Abiotrophia*, *Granulicatella*, *Rhizobium*, *Burkholderia*, and *Ralstonia*, with the latter two genera being absent from any of the teacher's samples. These large differences may be related to age but also to their understandably different lifestyle. The students were more often consuming sweets and chewing gum, while teachers were consuming significantly more coffee and alcohol, reported more dental health problems, and used flossing more frequently. Although not the focus of the study, some interesting correlations did emerge among the items in the questionnaire. For instance, smokers tend to consume more alcohol, and students who reported having a kissing partner were more likely to smoke, drink alcohol, or chew gum [see **Additional file 2.6**]. Interestingly, students with kissing partners had a higher number of taxa in their microbiomes, which also showed a significantly higher presence of *Treponema*.

Importantly, the reported consumption of alcohol among 314 students was associated with a higher presence of several bacterial genera including *Mycoplasma*, *Filifactor*, *Treponema*, and *Desulfobulbus*, among others [see **Additional file 2.7**]. Although 108 students declared smoking occasionally, we did not detect significant differences in their microbiomes. *Gemella* negatively correlated with the consumption of yogurt and milk. In addition, the consumption of milk was positively correlated with the abundances of *Actinomyces* and *Atopobium*.

2.4.4 Hygiene habits and saliva pH

Acidification plays an important role in oral health problems such as caries or periodontitis (E. Hajishengallis et al. 2017; Takahashi and Schachtele 1990). A pH level of less than 5.5 can put a person at risk of tooth enamel erosion, leading to the formation of cavities, while higher pH can reduce this risk. Measured oral pH in our samples had a median of 7.5 but showed a wide range [see **Additional file 2.9**]. Higher oral pH was positively correlated with the abundance of *Fusobacterium* and *Porphyromonas*, a bacterial genus known to grow optimally in alkaline environments, and able to increase the pH of its medium (Takahashi and Schachtele 1990). Other genera such as *Streptococcus* or *Veillonella*, among others, correlated negatively with saliva pH [see **Additional file 2.7**]. *Veillonella* species are known to increase their abundance in acidic environments derived from fermentation processes, such as those occurring in mature dental plaque (E. Hajishengallis et al. 2017). Importantly, no hygiene or dietary habit was shown to impact saliva pH in our study [see **Additional file 2.8**]. Admittedly, measurements of saliva pH using pH strips—a limitation imposed in part by our citizen science approach—lack the precision provided by a pH meter (see “**Materials and methods**” section). However, all our detected correlations were robust to stochastic variations within the precision range of the measurement, as shown by 1000 randomization tests (see “**Materials and methods**” section).

Our questionnaire included several questions on oral hygiene and dental devices. Hygiene habits usually showed high correlations among themselves, so that people who brush their teeth more often tended to use fluoride supplements and floss and were more likely to wash their hands before eating and/or after using the bathroom. Additionally, people using braces were more often brushing their teeth, and those reporting past nerve extractions drank more alcohol. According to our data, differences in type and frequency of oral hygiene do have measurable effects in the oral microbiome. Frequency of brushing teeth correlated negatively with the relative abundance of *Gemella*, *Streptobacillus*, *Granulicatella*, and *Porphyromonas*. It is known that caries is generally associated with an increase of *Streptococcus*, but also of *Granulicatella*, and *Gemella* (Costalonga and Herzberg 2014)—although in the latter case, this varies with age (Lif Holgerson et al. 2015)—supporting the effect of brushing against primary dental plaque. In contrast, flossing or using supplemental fluoride mouthwash did not seem to significantly impact the oral microbiome. The presence of dental implants did not show any correlation with oral microbiome changes, but wearing orthodontic braces did correlate positively with the abundance of many genera. These included several anaerobic or facultatively anaerobic genera such as *Corynebacterium*, *Bifidobacterium*, *Parascardovia*, *Olsenella*, *Capnocytophaga*, *Lactobacillus*, *Dialister*, *Schwartzia*, *Selenomonas*, and *Cardiobacterium.* This suggests that such orthodontic devices and their surfaces may promote the proliferation of specific biofilm communities. Most of these genera comprise anaerobic Gram negative species or Gram positives associated to acidic fermentations, which are generally associated to mature biofilm acidification, as well as caries and periodontal disease (E. Hajishengallis et al. 2017). *Selenomonas* has been described as one of the most abundant taxa during orthodontic braces treatment and has been linked to common oral diseases such as gingivitis (Koopman et al. 2015).

2.4.5 Tap water influences the oral microbiome

Unexpectedly, we found no significant differences between the lifestyles of students with the two oral stomatotypes, suggesting our data have not sufficiently captured the key factors underlying these different microbial communities. Notably, however, the two stomatotypes, and some genera, were geographically widespread but showed distinct abundance patterns, which suggest some environmental influence. The patterns were sometimes reminiscent of maps of certain public water quality parameters, such as alkalinity or water hardness, which differ significantly across regions in Spain (**Figure 2.6**). In addition, the mouth is constantly exposed to tap water, which is consumed for drinking, cooking, and hygiene. Hence, we decided to investigate this factor in more detail and linked our samples to the chemical composition of tap water of the nearest town, as reported in recent studies (Vitoria et al. 2015; Maraver, Vitoria, Ferreira-Pêgo, et al. 2015; Maraver, Vitoria, Almerich-Silla, et al. 2015). For this analysis, we removed individuals that declared drinking bottled water. No strong correlation was found between the two stomatotypes and any of the 17 water parameters investigated. However, we found that most considered water quality parameters are associated with alterations in the composition of several genera (**Figure 2.7**, **Additional file 2.10**). *Porphyromonas* was positively associated with the presence of fluoride (F) and sulfate (SO4) in tap water. A group of genera including, among others, *Veillonella*, *Ralstonia*, *Rhizobium*, *Rhodococcus*, and *Pseudomonas* negatively correlated with several of the following parameters: water hardness, alkalinity, conductivity, and the presence of SO_4, magnesium (Mg), sodium (Na), calcium (Ca), chloride (Cl), and the amount of dry matter after boiling. Other genera correlated positively with several of these same variables, including *Porphyromonas* and *Flavobacterium*. *Ralstonia* abundance was also negatively affected by nearly all

other water variables, and it was the genus whose abundance changed the most with tap water quality, followed by *Rhizobium*, *Veillonella*, and *Pseudomonas*. These results suggest that tap water composition may be an important, poorly studied factor shaping the oral microbiome.

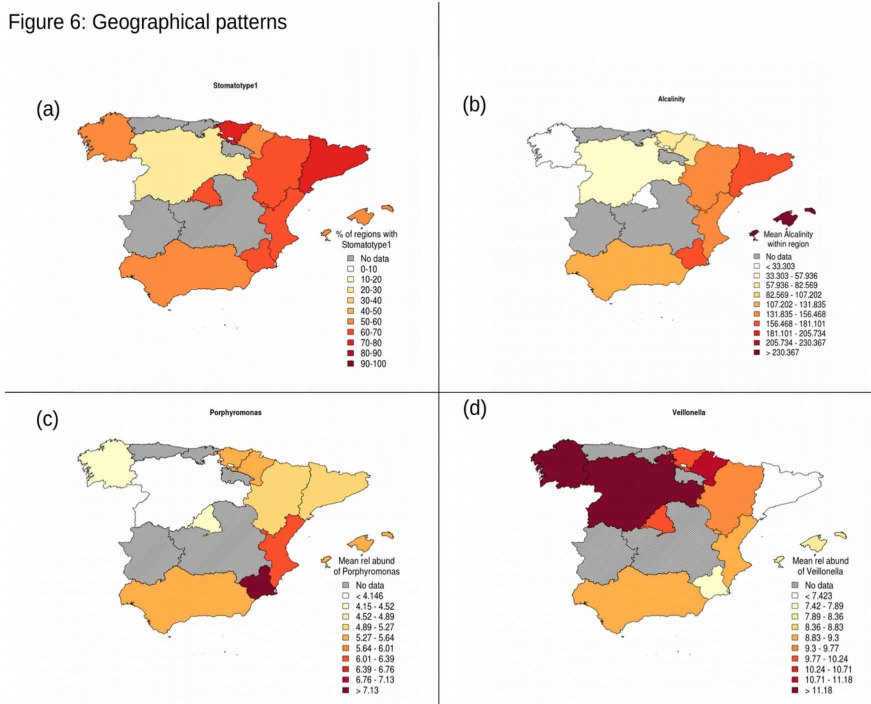

Figure 2.6: Geographical patterns. Maps show that most of the measured ion levels follow a similar pattern to the proportion of stomatotype 1 samples. *Porphyromonas* had a significantly higher abundance in stomatotype 1 samples, while *Veillonella* had a significantly higher abundance in stomatotype 2 samples. Region names can be seen in **Additional file 2.1**. **a** Percentage of samples from each region that have stomatotype 1. **b** Mean alkalinity level per sample in each region (an example of one of the tap water measurements compared in this study). **c** Mean abundance of *Porphyromonas* per sample in each region. **d** Mean abundance of *Veillonella* per sample in each region.

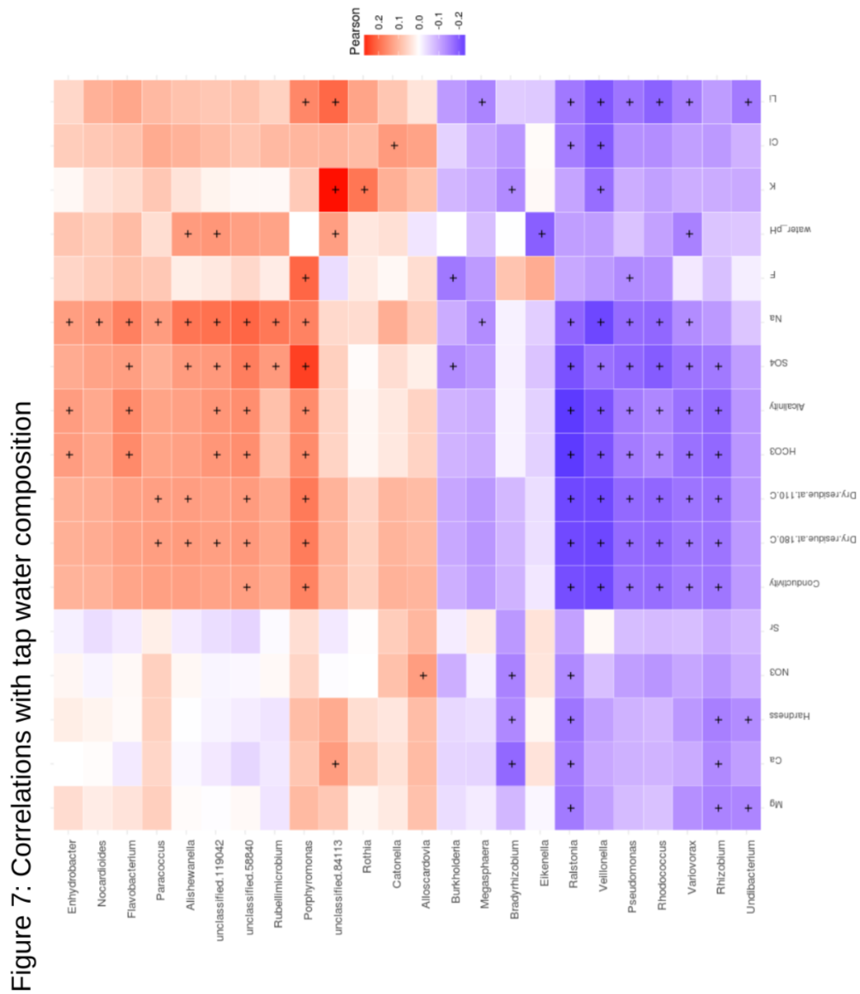

Figure 2.7: Correlations with tap water composition. Heatmap of correlations between relative abundances of genera with measurements of various components of tap water. Samples that primarily drank from bottled water (505 out of 1319) at home were excluded here. Color indicates Pearson correlation coefficients and “+” indicates a statistically significant correlation.

2.5 Discussion

Our study provides a comprehensive survey of the oral microbiome in Spanish adolescents, a target group that remains poorly explored. The citizen science approach has allowed us to address questions raised by citizens, train them in the use and interpretation of the data, and open a dialog with society on technologies and scientific questions of growing relevance. Although a citizen-based approach faces important limitations as compared to clinical studies, such as the difficulty to comprehensively evaluate clinical parameters by experts, it enables access to a large number of samples and of a different kind of those usually targeted by other studies. The high number of samples, the narrow range of geographical areas and ages under study, and the richness of collected metadata provide us an unprecedented level of resolution to study the adolescent oral microbiome. The insights gained from our study have served to generate working hypotheses regarding the composition and variability of the oral microbiome of adolescents that can be tested in future, more conventional studies. The core microbiome comprised typical oral bacteria that are commonly identified as abundant in similar oral microbiome surveys (Zaura et al. 2009). All genera discussed in the paper with the exception of *Rubellimicrobium* and *Undibacterium* have been previously identified in oral samples. Although the issue of contamination is a common theme in microbiome analyses, 20 amplification cycles and cell-rich starting materials such as oral samples are predicted to be minimally affected (Salter et al. 2014). In accordance with this, all of our negative controls provided no measurable results and a negligible number of reads when forced into library preparation and sequencing (see the "**Materials and samples**" section). However, we cannot discard the possibility that some of the low abundance genera identified are not stable components of the oral cavity but result from sporadic colonization from the close environment of the donor (i.e., food, air, or water).

Overall, we see that the oral microbiome of Spanish adolescents is impacted by dietary, hygiene, and other lifestyle habits. Differences observed point to a differential impact of habits on the oral

microbiome of adolescents. For instance, frequent teeth brushing was shown to affect the relative proportion of oral genera more than flossing, or the use of fluoride supplements.

Similarly, consumption of alcohol among adolescents seemed to impact the oral microbiome more than smoking. In contrast, we did not find many differences between genders or rural versus urban environments. Interestingly, some variables such as body mass index, which is generally associated to alterations in the gut microbiome, and it has been associated to changes in the oral microbiome in adults (Shillitoe et al. 2012), seemed to have a minor impact on the mouth microbiome of adolescents in our sample. Some of these differences may relate to the fact that some habits, such as smoking or some dietary habits, may have just been recently established, or the habit is more sporadic in adolescents, and the effects in the microbiome will only be apparent after a prolonged period of sustained habit. In addition, the oral microbiome of adolescents may have specificities as a transition phase from childhood to adulthood. Adolescence is a stage with major hormonal and habit changes, which likely impact the oral microbial community. In fact, this period of life is associated with a sharp increase in the incidence and severity of gingivitis (Mombelli et al. 1989), which may be related to underlying oral microbiome changes. This highlights the importance of increasing our knowledge of the adolescent oral microbiome, as well as to undertake longitudinal studies over adolescent to adulthood phases of life. Altogether, the chemical composition of tap water was found to be the investigated factor with the highest impact on the composition of the oral microbiome. Although the presence of the most abundant genera of the oral microbiome such as *Streptococcus*, *Prevotella*, or *Haemophilus* (the top three in our samples) were not significantly affected by tap water, some genera among the ten most abundant were affected, including *Veillonella*, *Porphyromonas, and Gemella.* Our results thus raise the question of the role of drinking water in shaping the oral microbiome, suggesting a potentially important role. Previous studies have analyzed the relationship between the presence of fluoride and the incidence of caries (Iheozor-Ejiofor et al. 2015), but the overall

impact on the human oral microbiota of this and other factors remain unexplored. In this regard, experiments in mice have shown that the composition of tap water can be related with changes in the gut microbiome (Dias et al. 2018) and have an incidence in the progression of diseases such as diabetes (Wolf et al. 2014). Further research is needed to follow up the potential role of tap water in shaping the human oral microbiome.

We found that the oral microbiome of the studied population can be broadly classified into two different stomatotypes. Although the time since last tooth brushing was not controlled in our study, we do not think this would drive overall observed differences regarding stomatotypes as all students in one class were sampled at the same time and we found that differences in stomatotypes were not driven by school class. Importantly, our two defined stomatotypes show notable overlap with the two “coinhabiting” groups of bacteria identified in another large study (Takeshita et al. 2016). Considering that the two studies use different profiling approaches (V1V2 regions in ion torrent vs V3V4 regions in MiSeq), and they target broadly different populations with markedly different genetic backgrounds and lifestyles (adults in Japan vs adolescents in Spain), the similarities are striking. The two studies coincide in defining higher proportions of *Neisseria*, *Haemophilus*, and *Porphyromonas*, in one of the types (stomatotype 1, coinhabiting group 2), and those of *Prevotella*, and *Veillonella* in the other (stomatotype 2, coinhabiting group 1). That the two disparate studies agree in the two broadly defined groups strongly suggests that these two stomatotypes define two possible equilibria of oral microbial communities which are globally present. In addition, that the two stomatotypes are similarly identified in adult and adolescent datasets suggests that, despite important differences, oral microbiomes from these two age groups are similar at a broad level. This reinforces the idea that the two stomatotypes define global equilibria of microbial communities, despite a possibly large underlying diversity. We propose naming these stomatotypes *Neisseria-Haemophilus* (stomatotype 1) and *Prevotella-Veillonella* (stomatotype 2) based on the four most abundant genera among those driving their differences. Although other studies have defined

higher numbers of clusters in the oral microbiome (De Filippis et al. 2014; Zaura et al. 2017), some of these clusters show clear similarities with the two stomatotypes found in this study.

We hypothesize that these two main stomatotypes are ubiquitous in humans and that they can be found across geographical regions, ethnic groups, and lifestyles, pointing to inherently deep relationships between the human oral niches and the bacterial communities that colonize them. Further support of this hypothesis with broader studies in other populations and geographical regions is needed. This finding also opens the question of the stability of these two stomatotypes and how lifestyle may promote shifts between the two equilibria. It is unclear whether differences in the number of clusters found across studies are due to differences in the studied populations or to variations in the applied methods. In addition, some authors have warned about the necessity to consider variations among samples as a gradient rather than as discrete clusters (Knights et al. 2014). We agree with this view and consider that stomatotypes represent trends in a continuous space of variation. As shown here, stomatotypes are appropriate to describe trends of change in the underlying microbial communities, which hint to shifts in the balance between driver genera. However, the two stomatotypes do contain a significant amount of variability and a gradient of variation, sometimes unrelated to the stomatotypes, is observed for the most abundant genera. In addition, that the described stomatotypes are common and globally distributed does not preclude the possibility that further, clearly distinct, stomatotypes may be found in other populations. Particularly, as the mentioned studies represent mostly healthy populations, further stomatotypes may be present that are associated to specific lifestyles or health conditions, which may represent alternative equilibria, or dysbiotic alterations from the two described stomatotypes. Certainly, further studies including broader samples and specific sampling from different niches within the oral cavity will help us describe in more detail the oral microbial ecosystem and its interactions.

2.6 Conclusions

The core oral microbiome described in this study is composed of genera commonly identified in other oral microbiome studies. We have shown that a number of diet and hygiene factors are associated with alterations in the composition of the oral microbiome, though one caveat is that, since the bulk of the sample set is from adolescents, some habits may be too recently developed to have already had a strong impact. The factor with the highest impact was the chemical composition of tap water from the hometowns of the donors. Indeed, most of the 17 ionic measurements showed significant correlations with a number of common genera such as *Veillonella* and *Porphyromonas*. This points to an important role of tap water in shaping the oral microbiome, which has been overlooked in previous studies.

We show that the samples can be clustered into two distinct groups which we call stomatotypes. The structures of these stomatotypes show notable similarities to the two clusters presented in another oral microbiome study of Japanese adults, despite differences in the technical approaches to the metagenomic analyses and highly distinct populations. Here, we propose the hypothesis that these two stomatotypes (the *Neisseria-Haemophilus* and *Prevotella-Veillonella* stomatotypes) represent global equilibria of oral microbial communities.

Additional files

Supplementary information contained in the additional files referenced throughout this chapter accompanies the publication at https://doi.org/10.1186/s40168-018-0592-3.

Acknowledgements

We are thankful to all students, teachers, and other citizens that participated in the "Saca La Lengua" project by contributing samples and sharing ideas. Only with their effort are studies like this possible. The authors acknowledge CRG Genomics Core Facility, CRG Bioinformatics Core Facility, CRG biomolecular

screening and protein technologies unit, CRG communication and public relationships department, and UCT ICTS High Performance Computing unit for providing access to the computing facilities. CRG authors acknowledge the Spanish Ministry for Economy, Industry and Competitiveness (MEIC) for the EMBL partnership, and Centro de Excelencia Severo Ochoa.

Funding

The project was financed by CRG through Genomics and Bioinformatics Core Facilities funds, and byEduCaixa programme through funds from the Fundación Bancaria "La Caixa," with the participation of the Center for Research into Environmental Epidemiology (CREAL), and the "Center d'Excellència Severo Ochoa 2013-2017" programme (SEV-2012-02-08) of the Ministry of Economy and Competitiveness. Eppendorf, Illumina, and ThermoFisher sponsored the research by donating some materials and reagents. David Harris Onywera was supported by a grant from the CRG-Novartis-Africa Mobility Program. TG group acknowledges support of the Spanish Ministry of Economy and Competitiveness grant BFU2015-67107 cofounded by European Regional Development Fund (ERDF); of the CERCA Programme/Generalitat de Catalunya; from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. H2020-MSCA-ITN-2014-642095.

Availability of data and materials

The fastq files for the 1319 samples used for the analyses in this study were uploaded to the Sequence Read Archive (SRA) with the BioProject accession number PRJNA427101 and can be found here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA427101.

The R script (SLL_pyroseq.r) used to perform the analyses, the OTU table, the taxonomy table, and a table of the questionnaire responses can be found here:

https://github.com/Gabaldonlab/ngs_public.

Author's contributions

TG designed and supervised the project and wrote the first draft of the manuscript. AL organized the citizen science aspects and the sample collection. LAB collected samples. EK, SI, and ES performed and helped optimize the DNA extraction and amplification procedures. NAS, CCa, JH, and MAT performed DNA extraction, library construction, and sequencing. LC, CCo, MM, AH, HO, and JP performed quality check and trimming of the raw sequencing data and benchmarked and ran the taxonomic assignment pipeline. AV linked samples to socio-economic status values from geo-localized databases. JRW, PGT, and AAP performed statistical analyses, and JRW and PGT helped write the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All participants, and at least one of their parents or legal guardians for those under the age of 18, signed a consent form to use their saliva samples for microbiome research. This consent form and the purpose of this project received approval by the ethics committee of the Barcelona Biomedical Research Park (PRBB). Data is anonymous.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 3: Oral microbiome in Down Syndrome and its implications on oral health

Willis JR, Iraola-Guzmán S, Saus E, Ksiezopolska E, Cozzuto L, Bejarano LA, Andreu-Somavilla N, Alloza-Trabado M, Puig-Sola A, Blanco A, Broglio E, Carolis C, Hecht J, Ponomarenko J, & Gabaldón T. (2020). Oral microbiome in Down Syndrome and its implications on oral health. *J Oral Microbiol*. 13(1). doi: 10.1080/20002297.2020.1865690

Chapter 3: Oral microbiome in Down Syndrome and its implications on oral health

3.1 Abstract

Background: The oral cavity harbors an abundant and diverse microbial community (i.e. the microbiome), whose composition and roles in health and disease have been the focus of intense research. Down syndrome (DS) is associated with particular characteristics in the oral cavity, and with a lower incidence of caries and higher incidence of periodontitis and gingivitis compared to control populations. However, the overall composition of the oral microbiome in DS and how it varies with diverse factors like host age or the pH within the mouth are still poorly understood. **Methods**: Using a Citizen-Science approach in collaboration with DS associations in Spain, we performed 16S rRNA metabarcoding and high-throughput sequencing, combined with culture and proteomics-based identification of fungi to survey the bacterial and fungal oral microbiome in 27 DS persons (age range 7–55) and control samples matched by geographical distribution, age range, and gender. **Results**: We found that DS is associated with low salivary pH and less diverse oral microbiomes, which were characterized by lower levels of *Alloprevotella, Atopobium,* Candidatus Saccharimonas, and higher amounts of *Kingella, Staphylococcus, Gemella, Cardiobacterium, Rothia, Actinobacillus*, and greater prevalence of *Candida*. **Conclusion**: Altogether, our study provides a first global snapshot of the oral microbiome in DS. Future studies are required to establish whether the observed differences are related to differential pathology in the oral cavity in DS.

Keywords: Oral microbiome, down syndrome, oral mycobiome, *Candida*

3.2 Introduction

Down syndrome (DS), also known as trisomy of chromosome 21, is the most common genetic cause of mental disability worldwide (Lukowski, Milojevich, and Eales 2019). The estimated incidence of DS is between 1 in 1,000 and 1 in 1,100 live births worldwide, according to the World Health Organization. DS is generally caused by maternal nondisjunction errors during meiosis, which results in chromosome 21 trisomy in all cells of the body, wherein advanced maternal age is the main risk factor (Sherman et al. 2007). DS is characterized by variability in cognitive development and distinct physical features causing unique health conditions, including congenital heart disease, immune system alterations, premature dementia, Alzheimer's disease, and many other symptoms related to premature aging. Indeed, current clinical and experimental findings support the concept that DS may be considered a premature aging disorder (Franceschi et al. 2018). Consistent with this, DS presents with premature immune system senescence, increased plasmatic levels of inflammatory markers resembling the chronic increase in proinflammatory status observed during aging, as well as oxidative stress due to mitochondrial dysfunction (Franceschi et al. 2018). Furthermore, consistent with premature aging, two markers of biological age, DNA methylation and quantification of circulating N-glycan species (GlycoAgeTest), show differences in DS compared with control individuals (Horvath et al. 2015; Borelli et al. 2015).

The composition of the gut microbiome has been suggested to be a powerful marker for distinguishing between biological and chronological age (Rampelli et al. 2020; A. Liu et al. 2020; Maffei et al. 2017; Biagi et al. 2016), and, therefore, the characterization of the gut microbiome in DS populations may be of interest (Thevaranjan et al. 2018). Development of new technologies and the application of metagenomic analyses have enabled the characterization of the human microbiome at different body sites (NIH HMP Working Group et al. 2009; Human Microbiome Project Consortium 2012). The profiling of the gut microbiome of 17 individuals with DS and matching controls showed a similar

structure of the gut microbiome, with alterations in only two genera: *Parasporobacterium* and *Sutterella* (Biagi et al. 2014).

The microbiome of the oral cavity has been related to several diseases, including not only common oral diseases such as gingivitis or periodontitis, but also systemic ones (**Chapter 1**, (Willis and Gabaldón 2020)). The study of the oral microbiome in DS is of particular interest due to the many specific features of the oral cavity associated with this syndrome. These include, among others, different saliva composition, poor occlusal correlation, high frenum insertion, early mucogingival problems and advanced tongue position. In addition, due to genetic abnormalities in their immune system and environmental factors, DS patients are more susceptible to infections. In particular, it has been largely reported that persons with DS have an increased prevalence of periodontal disease (Cichon, Crawford, and Grimm 1998; A. Khocht et al. 2012; Kornman 2008; Barr-Agholme et al. 1992; Meskin, Farsht, and Anderson 1968; Sakellari, Arapostathis, and Konstantinidis 2005). First observations pointed to an excessive inflammatory response of the gums followed by manifestations of aggressive and/or early-onset periodontitis, characterized by high plaque formation and increased presence of pathogenic species *(Aggregatibacter actinomycetemcomitans, Porphyromonas gingivalis, Prevotella intermedia, Eikenella corrodens, Tannerella forsythia, Fusobacterium nucleatum, Treponema denticola*, and *Campylobacter rectus*) when compared with age-matched control groups and age-matched groups with intellectual disability (A. Amano et al. 2001). The early periodontal colonizers in DS are facilitated by a combination of lower salivary flow rate, limited antibody production in the saliva, and neutrophils with impaired chemotaxis which prevents the immune cells from reaching target pathogens (Ahmed Khocht 2011). Several studies have focused on identifying factors influencing the onset of aggressive variants of oral diseases that often affect youths and adults with DS, such as destructive forms of periodontitis (Agholme, Dahllöf, and Modéer 1999; Reuland-Bosma and van Dijk 1986). More recently, some authors have suggested that the composition of oral biofilms, independently of the immunological alterations in DS, has a critical

role in periodontal development (Martinez-Martinez et al. 2013). Although recent studies have failed to find a specific combination of pathogens causing periodontitis in DS patients, other than the ones causing periodontitis in the general population, it appears that these might be established earlier in young adult DS, which is associated with early-onset and more aggressive forms of the disease (Atsuo Amano et al. 2008). Despite the impaired immune responses and comparatively poor oral health in DS, a lower or similar incidence of dental caries in DS has been seen compared to non-DS (Deps et al. 2015; Moreira et al. 2016). This is potentially due to the relatively late eruption of teeth in DS, microdontia, more missing teeth and greater dental spacing (Cheng, Yiu, and Leung 2011; Vigild 1986), and so it would also be interesting to explore whether specific oral microbiotas are also related to this relatively lower incidence of caries in DS. Most previous studies have been based on identification by conventional PCR of already described individual species. Hence, we still lack a comprehensive understanding of the global oral microbiome composition in DS.

With the aim of shedding light on the composition of the oral microbiota in DS, and how it differs from similar non-affected populations, we used 16S rRNA metabarcoding coupled to culture and proteomics-based identification of fungi to characterize the bacterial and fungal components of the oral microbiome of 27 DS volunteers and their relatives, and compared it with non-affected volunteers. To this end, we collected and processed saliva samples from different locations across Spain in the context of the second edition of the citizen-science project 'Saca la lengua' (SLL2) (www.sacalalengua.org, **Chapter 2**, (Willis et al. 2018)). We expected to find differing abundances of organisms that may explain the greater incidence of periodontitis and lower incidence of dental caries in DS. Furthermore, considering the various signs of premature aging across the body in DS, we attempted to analyze potential signs within the oral microbiome. Overall, our study provides a first global overview of the species present in the oral cavity of DS. Future studies would benefit from an increment of sample size and the characterization of relevant variables for the study, such as the presence of comorbidities that may be shaping

differences in the composition of the microbiome of the oral cavity.

3.2 Materials and methods

3.2.1 Sample collection

The target population of this study was individuals with DS and their relatives, which were contacted with the collaborations of local associations of DS families. We collected 27 oral rinse samples from individuals with DS (ages 7–33) in the context of the second edition of the 'Stick out Your Tongue' citizen science project (SLL2, see http://www.sacalalengua.org, **Chapter 2**, (Willis et al. 2018)), and in close collaboration with DS family associations in Spain. Sample collection was coupled to science communication activities with DS individuals and their relatives, aiming to raise awareness about the microbiome, its role in health and disease, and its potential particularities in DS. The SLL2 project questionnaire about health and lifestyle was adapted with the help of DS associations and was answered jointly by DS participants and their relatives [see metadata file at the following github link: https://github.com/Gabaldonlab/ngs_public/blob/master/SLL2/SLL2.metadata.xlsx]. There were 20 relatives that participated, 18 of which were parents, and two of which were siblings. The siblings were 10 and 28 years old (their DS siblings were 7 and 22 years old, respectively), while the parents ranged in age from 44 to 77 years old.

All participants signed an informed consent form allowing the use of their saliva samples for microbiological research. For participants under the age of 18, the consent form was also signed by one of the parents or a legal guardian. This project was approved by the ethics committee of the Barcelona Biomedical Research Park (PRBB). Samples were collected from January to November 2017. Participants were asked not to ingest any food or beverage (except water) for 1 h before collecting the sample. All donors received clear indications about the sample collection procedure in person, and the collection of the samples was carried out with the assistance

of a researcher involved in the project, following a demonstration. All participants responded to a uniform questionnaire (see below), which was adapted for DS in collaboration with DS partner associations, so that DS participants could decide for themselves whether or not to participate, and be able to answer most questions on their own. Before collection of the oral rinse, the pH of the saliva was measured using pH test strips (MColorpHast, Merck, range 5.0–10.0; 0.5 accuracy units), the accuracy of which has been previously validated (**Chapter 2**, (Willis et al. 2018)). Saliva samples were collected using a mouthwash as described earlier (**Chapter 2**, (Willis et al. 2018)). In brief, the protocol is as follows: participants rinsed their mouth with 15 mL of sterile phosphate-buffered saline (PBS) solution, for 1 min. Then, they returned the liquid into a 50 mL tube. The samples were then centrifuged at 4,500 g for 12 min at room temperature (r.t.) in an Eppendorf 5430 centrifuge equipped with an Eppendorf F-35-6-30 rotor. The supernatant was discarded and the pellets were resuspended with the remaining PBS, transferred to 1.5 ml tubes and centrifuged at 4,500 g for an additional 5 min at r.t. using an Eppendorf FA-45-24-11-HS rotor. Supernatants were discarded, and pellets were frozen and stored at − 80°C until further analysis.

3.2.2 DNA extraction and sequencing

Sample DNA was extracted using the ZR-96 Fungal/Bacterial DNA kit (Zymo research Ref D6006), following the manufacturer’s instructions. The extraction tubes were agitated twice in a 96-well plate using Tissue lyser II (Qiagen) at 30 Hz/s for 5 min at 4°C. We included as controls of library preparation and MiSeq sequencing processes two DNA samples derived from bacterial mock communities from the BEI Resources of the Human Microbiome Project: each sample comprised genomic DNA of ribosomal operons from 20 bacterial species. The ‘HM-782D’ community contained an even number of ribosomal DNA per species (100,000 operons per species). The ‘HM-783D’ community contained a variable number of operons, ranging from 1,000 to 1,000,000 per species.

Obtained DNA was diluted to 12.5 ng/µl and used to amplify the variable regions known as V3–V4 of the *16S ribosomal RNA* gene, using a pool of modified universal primers in a limited cycle PCR:

V3-V4-Forward

(5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3′)

V3-V4-Reverse

(5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3′)

Low sequence diversity or unbalanced base composition in template DNA can negatively affect the sequence output, quality, and error rate due to problems in cluster identification in MiSeq sequencing. To prevent this unbalanced base composition with all libraries having the same initial sequence, we performed the first PCR with a mix of four forward and four reverse primers, shifting the sequencing phases by adding a varying number of bases (from 0 to 3 N bases) as spacers. The PCR was carried out using KAPA HiFi HotStart ReadyMix (Roche) in a total volume of 10-µl with 0.2-µM of the primers. Cycling conditions were: 3 min at 95°C followed by 20 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, with a final step of 5 min at 72°C. PCR products were purified using AMPure XP beads (Beckman Coulter), with a 0.9 x ratio and eluted with 32 µl of Buffer EB (Qiagen), then 30 µl of the eluate were transferred to a fresh 96-well plate.

Next, for the purpose of multiplex sequencing, we performed a second PCR in which full-length Nextera adapters (Illumina) with barcodes were added to the overhangs of the primers used in the first PCR. Thus, products of the second PCR step were sequencing ready libraries with an insert size of approximately 450 bp. Specifically, 5 µl of the first PCR was used as a template for a second PCR with Nextera XT v2 adaptor primers (Illumina) in a final volume of 50 µl using the same mix and conditions as the first

PCR, but limited to eight cycles. Twenty-five µl of the final product of this second PCR was used for purification and normalization with the SequalPrep normalization kit (Invitrogen), using the manufacturer's protocol. Libraries were eluted in a volume of 20 µl and pooled for sequencing. Pools were quantified by qPCR using the Kapa library quantification kit for Illumina (Kapa Biosystems) on an ABI 7900HT real-time cycler (Applied Biosystems). Libraries were sequenced on an Illumina MiSeq with 2 × 300 bp reads using v3 chemistry with a loading concentration of 15 pM. To increase the diversity of the sequenced sample, 10% of PhIX control libraries were spiked. Negative controls with the same conditions and reagents but with sterile water instead of samples were made for the buffer, DNA extraction, and PCR amplification steps. The controls provided no visible band or quantifiable amount of DNA by gel visualization or Bioanalyzer, whereas all samples resulted in clearly visible bands after 20 cycles. Twelve such controls were subjected to library preparation and sequenced. Expectedly, these sequenced non-template controls systematically yielded very few reads (a range of 155–1,005 reads per sample), in contrast to an average of ~64,000 reads/sample in sample-derived libraries.

3.2.3 Fungal composition analysis

Attempts to identify fungi through ITS amplification failed in many samples that nevertheless were positive for a culture-based approach. This likely related to a low presence of fungi in oral rinse samples and the difficulty to break the fungal cell wall to access the DNA in comparison to bacteria. We therefore used traditionally culture-based methods to enrich the possible fungal species present in the samples and identify them. First, we optimized the experimental procedures for fungal composition detection with a small subset of control oral rinse samples, testing the following growing conditions: i) starting with fresh versus frozen oral rinse samples, ii) using of different antibiotics concentrations, iii) growth temperature (25°C, 30°C, 37°C), iv) plating method (spread plate method versus pour plate method), and iv) duration of incubation (2, 4 or 7 days). We observed no significant differences in the

number of grown colonies when starting with fresh versus frozen samples or when using different incubation temperatures, while the number of grown colonies when using the spread plate method was somewhat higher as compared to the pour plate method, and longer incubation times resulted in a higher number of grown colonies (data now shown). Thus, the preliminary results obtained helped us to design the following working protocol to carry out the analysis of the oral fungal composition with the study samples. First, we resuspended the frozen pellets in 100 µl of sterile PBS, taking 10 µl of those to be diluted in 40 µl of PBS (the remaining 90 µl were used for DNA extraction), from which we plated 30 µl (6% of the original sample) onto a YPD sterile plate with chloramphenicol and ampicillin (100 µg/ml each). After 7 days of incubation at 30°C, we counted the number of colonies, their phenotypes, and the presence of bacteria. A total of 10 colonies were randomly selected per sample (or all plate colonies if there were fewer than 10) and were re-grown under the same conditions in a fresh plate for 24 h. We used MALDI-TOF analysis for fungal identification and, in the instance of inconclusive results after two or three rounds of MALDI-TOF, we performed colony PCR to amplify the Internal Transcribed Spacer (ITS) hypervariable region of the ribosomal gene 5.8S (fungal marker) and further Sanger sequencing. Data are summarised in a table (**Table 3.1**).

Table 3.1: Analysis of colonies grown onto YPD + antibiotics plates. The table summarises the number (n) and frequency (%) of samples which formed colonies for **a**) yeasts, – with **b**) indicating the mean and the range of the number of colonies for the yeast positive samples – **c**) molds, **d**) bacteria, and **e-q**) identified fungal species as determined by MALDI-TOF per group: subjects with DS (Down Syndrome) and matched control individuals.

Colony analysis	**DS (n = 26)**	**CONTROLS (n = 332)**
a) Yeast	14 (53.85)	85 (25.60)
b) # yeast colonies, mean (min-max)	67.29 (1-477)	32.14 (1-578)

c) Mold	1 (3.85)	57 (17.17)
d) Bacteria	12 (46.15)	188 (56.62)
e) *Candida albicans*	8 (30.77)	63 (18.98)
f) *Candida parapsilosis*	4 (15.38)	2 (0.60)
g) *Candida dubliniensis*	4 (15.38)	5 (1.51)
h) *Candida lusitaniae*	1 (3.85)	1 (0.30)
i) *Trichosporon spp.*	1 (3.85)	0 (0)
j) *Debaryomyces hansenii*	0 (0)	3 (0.90)
k) *Candida guilliermondii*	0 (0)	6 (1.81)
l) *Candida intermedia*	0 (0)	2 (0.60)
m) *Candida lusitaniae*	0 (0)	1 (0.30)
n) *Candida zeylanoides*	0 (0)	1 (0.30)
o) *Rhodotorula mucilaginosa*	0 (0)	1 (0.30)
p) *Candida glabrata*	0 (0)	1 (0.30)
q) *Candida spp.*	0 (0)	2 (0.60)

MALDI-TOF analysis was performed with a MALDI Biotyper in the Centre for Omics Sciences (COS) in EureCat (Centre Tecnològic de Catalunya, Reus, Spain). Total proteins of colonies were extracted in our laboratory following their standard protocols and then samples were sent to COS for the analysis. In brief, we picked fresh grown colonies into 1.5 ml tubes, added 300 µl of milliQ water and mixed thoroughly. Next, we added 900 µl of 100% ethanol to the sample, mixed thoroughly again, and centrifuged the

tubes for 2 min at 13,000 rpm in a benchtop centrifuge to remove the supernatant (the centrifugation step was repeated twice). To increase the efficiency of identification, we dried pellets for some minutes at room temperature. After that, we resuspended the pellet in 70% aqueous formic acid (covering the pellet), added an equal volume of 100% acetonitrile, mixed carefully, and centrifuged for 2 min at 13,000 rpm in a benchtop centrifuge. Finally, we transferred the supernatant with the fungal protein extract into a fresh tube. Samples were frozen at −20°C and sent to COS for analysis in a MALDI Biotyper (Bruker Daltonik MALDI Biotyper). Samples were deposited in duplicate on a Polished Steel Target Plate (Bruker), coated with the matrix α-Cyano-4-hydroxycinnamic (HCCA) and analyzed with MALDI-TOF/TOF (MALDI ultrafleXtreme, Bruker, Germany). Spectrum identification was performed using the real-time classification software MALDI Biotyper (Bruker, Germany). Thus, for each colony sample we obtained a score with the following meaning according to the manufacturer’s recommended score identification: i) 2.3–3.0, highly probable species identification; ii) 2.0–2.299, secure genus identification, probable species identification; iii) 1.7–1.999, probable genus identification; iv) < 1.7, not reliable information. Samples identified with scores lower than 2.0 were manually re-analyzed, and its spectrum was classified using the Biotyper Offline Classification software. We considered that the results were consistent with species identification when the best match had a score >2.0 and the second best match was at least >1.7 with at least the same genus as the first one. The rest of the samples were re-grown and the whole process was repeated. If after two more experiments with MALDI-TOF analysis the identification of the fungal species was still inconclusive, we performed ITS amplification and sequencing.

For ITS-Sequencing, we performed a colony PCR from fresh colonies (replated 12–24 h before) with DongSheng Biotech (DSBio) Taq mix (#2012) and 20 pmol of each primer (ITS1: 5′-TCCGTAGGTGAACCTGCGG-3′; NL4: 5′-GGTCCGTGTTTCAAGACGG-3′) in a total volume of 40 µl, and the following PCR program: 5 min at 94°C, then 30 cycles of 30 s

at 94°C, 30 s at 55°C, 1.5 min at 72°C, and a final extension of 5 min at 72°C. PCR products were purified with the QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions. Samples were eluted in 33 ul of elution buffer, and sent for Sanger sequencing to Eurofins Genomics sequencing service, following the SUPREMERUN recommendations. The resulting sequences were evaluated with the webtool Blast (Altschul et al. 1990). Only samples with a high score (>95%) of identity were considered as correctly identified. Additionally, for some samples, we were able to identify the predominant species responsible for bacterial growth on the YPD plates with antibiotics.

3.2.4 Pre-processing of 16S rRNA sequence reads and taxonomy assignment

Sequence reads from fastq files were filtered using the 'dada2' R package (version 1.10.1) (Benjamin J. Callahan et al. 2016) to produce counts of amplicon sequence variants (ASVs). Low-quality reads were first removed by applying the *filterAndTrim* function with the following parameters: forward and reverse reads were trimmed to lengths of 275 and 230 nucleotides, respectively (truncLen = c(275,230)); the leading 10 nucleotides were trimmed in both reads (trimLeft = c(10,10)); reads with maximum expected errors greater than five in both reads were discarded (maxEE = c(5,5)); all other parameters used the default values. The remainder of the pipeline followed the suggestions in the tutorial from the authors of the tool (("DADA2 Pipeline Tutorial (1.16)" n.d.), https://benjjneb.github.io/dada2/tutorial.html). Taxonomy was assigned using the dada2-formatted database of SILVA version 132 (B. Callahan 2018). A phylogenetic tree for use in UniFrac distance calculations was generated by following a protocol (Ben J. Callahan et al. 2016) that uses the 'DECIPHER' (version 2.10.2) (Wright 2016) and 'phangorn' (version 2.5.5) (Schliep 2011) R packages. After processing reads with the dada2 pipeline, only those samples with at least 5,000 reads were retained. At the end of this process there were a total of 1,648 samples, though only a portion of those

were used for the analyses of this study, as described below in the section ‘Statistical analyses’.

For analyses regarding the abundances of taxa, a centered log-ratio transformation was applied to the ASV counts. Zeros were first replaced with the ‘count zero multiplicative’ method in the *cmultRepl* function from the ‘zCompositions’ R package (version 1.3.4) (Palarea-Albaladejo and Martín-Fernández 2015). Then centered log ratios were calculated using the *codaSeq.clr* function from the ‘CoDaSeq’ R package (version 0.99.5) (Gloor et al. 2016; Gloor and Reid 2016).

3.2.5 Diversity measures

Alpha diversity measures were calculated using the *estimate_richness* function from the ‘phyloseq’ R package (version 1.30.0) (McMurdie and Holmes 2013). For beta diversity measures, both the weighted and unweighted UniFrac distances, which weight dissimilarity between samples by phylogenetic distances between taxa, were calculated using the *UniFrac* function from the ‘phyloseq’ package. The weighted UniFrac distance gives additional weight based on taxa abundances. Bray-Curtis and Jaccard distances were calculated using the *vegdist* function from the ‘vegan’ R package (version 2.5–6) (Oksanen et al. 2017). As the Jaccard distance is based on the presence or absence of taxa, the *decostand* function, also from the ‘vegan’ package, was applied to the ASV counts table, using the method ‘pa’ for presence/absence, before the *vegdist* function was applied. The Aitchison distance was calculated using the *aDist* function from the ‘robCompositions’ R package (version 2.2.1) (Templ, Hron, and Filzmoser 2011).

3.2.6 Statistical analyses

When running statistical tests, we first randomly selected representative-matched non-DS samples as controls 100 times to ensure consistency in the results. These same 100 sub-samples were used for each of the relevant tests, and were matched for

geographical location, age, and gender by the following process. Of the 1,621 non-DS samples in the SLL2 dataset, we removed those samples with any other chronic disorder, leaving 1,335 samples. The DS samples came from four autonomous communities in Spain (Andalucia, Catalunya, Galicia, and the Basque Country), so from the non-DS controls, we first randomly selected two times the proportion of DS samples from each of those locations (i.e. 2 x 3/27, 14/27, 1/27, and 9/27, respectively). To ensure a comparable age range, we determined rough age brackets of youth (under 20), adult (20–60), and senior (60 and over), and randomly selected from the geographically matched samples the same proportions of each age group from DS samples (i.e. 12/27 youths, 15/27 adults, 0/27 seniors). Among the DS samples, there were 12 females and 15 males. Thus, a given sub-sampling was finally rejected and reselected if the proportions of males and females were not similar to that of DS samples (i.e. (12±2)/27 females and (15±2)/27 males). Among all of the 100 sub-samplings, a total of 332 samples were used as matched controls.

For each of these sub-samples, a number of statistical tests were run with the DS and matched controls together. First, we performed a permutational multivariate analysis of variance (permanova) based on each of the five distance metrics mentioned above using the *adonis* function from the 'vegan' package. The model included the following fixed effects: DS/non-DS, gender, age, and population of the city/town from which the sample came (as a generalized proxy of both location and lifestyle).

Then, to determine differential abundances of taxa and variation in other variables like alpha diversity and pH, we performed a linear model using the function *lm* from the base R package 'stats' (version 3.6.3) (R Core Team 2020), again using the same fixed effects as for the permanova test. The abundance values used for these tests were the centered log ratios of the ASV counts, as described above. The *Anova* function from the 'car' R package (version 3.0–7) (Fox and Weisberg 2019) was used to calculate type-II anova tables, from which p-values were taken for each fixed effect in the model. These p-values were corrected for multiple

testing with the *p.adjust* function from the 'stats' package, using the 'fdr' method.

3.3 Results

3.3.1 Increased abundance of periodontal pathogens in DS

A number of genera consistently showed significant differences in abundance between DS and controls among the 100 sub-samples (**Table 3.2**, see **Materials and methods** section for explanation of this process) and all of these organisms have potential implications in the pathogenesis of periodontitis and dental caries, as will be explored in the discussion section. The genera found at higher abundance in DS included *Kingella, Gemella, Cardiobacterium, Staphylococcus, Rothia*, and *Actinobacillus* (**Table 3.2**, **Figure 3.1b**, **Supplementary Figure 3.1**). In addition, reads that could not be classified at even the phylum level were found at greater abundance in DS. While the relevance of these to DS is questionable, this observation suggests that rare and perhaps understudied organisms were more abundant in these DS samples. On the other hand, the genera *Alloprevotella* (and its phylum *Bacteroidetes), Atopobium*, and *Candidatus_Saccharimonas* (and its phylum *Patescibacteria*) were found at lower abundance in DS (**Table 3.2**, **Figure 3.1b**, **Supplementary Figure 3.1**). There were no taxa from the genus to the phylum level that differed significantly between the 20 relatives of DS individuals and the same sub-samplings of matched controls.

Table 3.2: Significance of differentially abundant taxa and other variables between DS and matched controls. Columns indicate, in this order, the taxonomic level or the type of variable considered, the organism name or the variable name, the tendency of the difference in DS (↗: higher in DS, ↘: lower in DS, permanova results are not directional), the mean adjusted p-value of the statistical comparison between DS and matched controls, and the numbers of matched

controls sub-samples for which the test is significant (out of the 100 matched control groups that were sampled from the total SLL2 dataset). Rows are ordered by mean adjusted p-value within each variable group.

Taxonomic level/Variable	Organism/ Variable	Tendency in DS	Mean adjusted p value	Number of significant sub-sample tests
Genus	*Kingella*	↗	8.34e-6	100
	Alloprevotella	↘	0.00165	100
	Atopobium	↘	0.014	96
	Unclassified Phylum	↗	0.0247	87
	Staphylococcus	↗	0.0253	83
	Gemella	↗	0.039	77
	Cardiobacterium	↗	0.0397	74
	Candidatus Saccharimonas	↘	0.0422	70
	Rothia	↗	0.0428	70
	Actinobacillus	↗	0.0451	66
Phylum	Patescibacteria	↘	4.72e-5	100
	Bacteroidetes	↘	0.0022	99
Fungi	*Candida parapsilosis*	↗	0.0282	79
	Candida dubliniensis	↗	0.0897	2
	Yeast detected (Yes/No)	↗	0.119	18
Sialochemistry	pH	↘	0.0109	97

Alpha diversity	Shannon diversity index	↘	0.0411	72
Permanova (Beta diversity)	Aitchison	-	0.001	100
	Jaccard	-	0.00101	100
	Weighted UniFrac	-	0.00136	100
	Bray-Curtis	-	0.00139	100
	Unweighted UniFrac	-	0.00149	100

3.3.2 Lower pH and species diversity in DS

DS samples consistently had lower pH and lower alpha diversity than the matched controls, as calculated by the Shannon diversity index (**Figure 3.1b**). There was no significant difference in either alpha diversity or pH between the 20 relatives of DS individuals and the same sub-samplings of matched controls. Overall microbiome composition differed significantly in all 100 sub-samples between DS and matched controls based on a permanova test on each of the five distance metrics mentioned in the methods section, each of which calculates the distance between given samples using different criteria (**Table 3.2**).

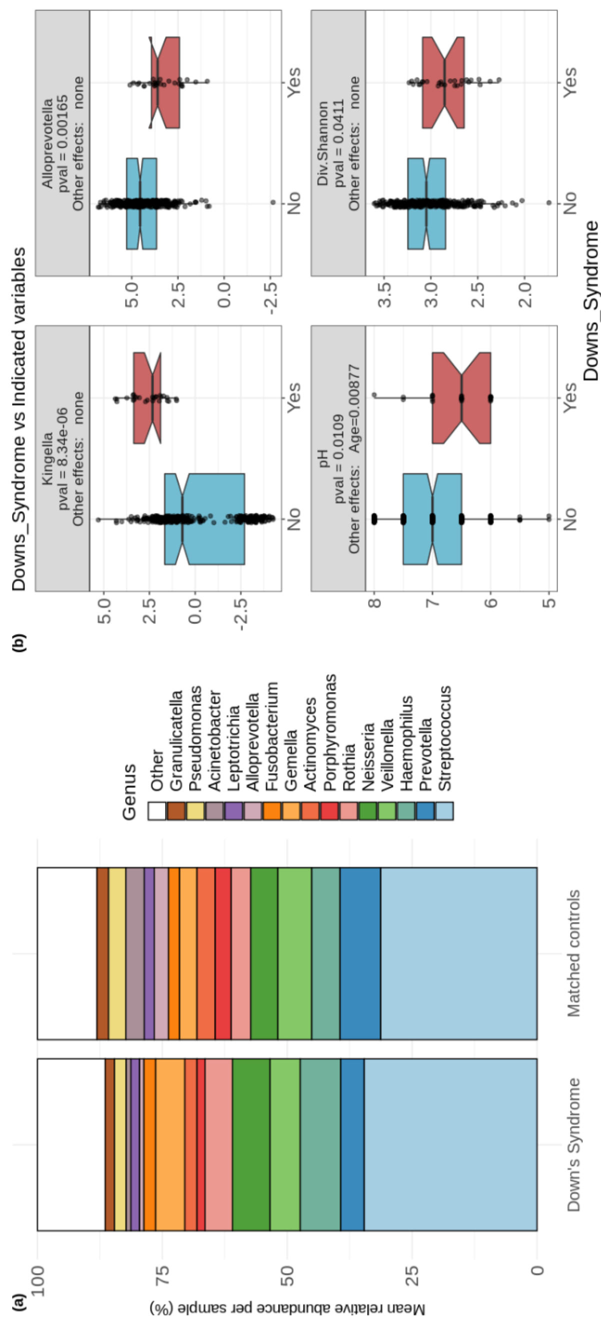

Figure 3.1: Down Syndrome differs in factors affecting oral health. (**a**) Mean relative abundances of 15 of the most abundant genera in DS samples and

matched controls. The remaining genera are grouped together and colored in white. (**b**) The two most significantly differentially abundant genera are shown (*Kingella* and *Alloprevotella*), as well as the salivary pH and alpha diversity as calculated by the Shannon diversity index.

3.3.3 Opportunistic pathogenic *Candida* species were more prevalent in DS

We assessed the presence of yeast in the oral microbiome through a culture-based approach coupled with proteomics identification (**Table 3.1**, see **Materials and methods**). This information was available for 26 of the 27 DS samples, and all of the matched control samples. Of the 26 DS samples, four were positive for *Candida parapsilosis*, while just two of the 332 matched controls used in all sub-samples were positive for *C. parapsilosis* ($p = 0.00082$ for multinomial log-linear model including all DS and matched control samples, mean adjusted p among 100 sub-samples = 0.0282, significant in 79 sub-sample tests, **Table 3.2**). *Candida dubliniensis* was also found at a greater proportion in DS samples: four out of 26 as compared to five out of 332 matched controls ($p = 0.00082$), though this association was not consistently significant among 100 sub-samples (mean adjusted $p = 0.0897$, significant in two sub-sample tests, **Table 3.2**). Yeasts in general were also proportionately more prevalent among DS samples, found in 14 of 26 samples as compared to 85 of 332 ($p = 0.0161$), though again this association was not consistently significant (mean adjusted $p = 0.119$, significant in 18 sub-sample tests, **Table 3.2**). There was no significant difference in the prevalence of any yeast between the 20 relatives of DS individuals and the same sub-samplings of matched controls.

3.3.4 No evidence for premature aging in the oral microbiome of DS

Given the above-mentioned notion that the microbiome composition might reflect premature aging in DS, we looked for potential

signatures of this process in the oral microbiome following two main approaches: first we assessed whether DS samples were more similar to those from non-DS older individuals than to those of non-DS younger individuals, and second, we checked whether changes across age were more drastic within DS samples than within non-DS samples. For the first approach we grouped all samples with no chronic disorders from the full SLL2 dataset, which included 1,335 samples ranging in subject age from 7 to 85 years old. These were placed into four age groups: child (under 13 years old), teen (13 to 19 years old), adult (20–59 years old), and seniors (60 years or older). We then compared the overall composition of DS samples to the samples in each of these age groups using five different distance metrics (Aitchison, weighted and unweighted UniFrac, Bray-Curtis, and Jaccard) (**Figure 3.2a**). Overall, we observed no apparent ‘premature aging’ effect in the DS samples in the sense that they would be more similar to samples from individuals older than themselves. Generally, DS samples showed the lowest difference with children, suggesting this is the most similar group to DS samples, while teens were typically the most distant, depending on the metric used. But none of the metrics showed that DS samples were most similar to either adults or seniors, which would be the expected result in the case of premature aging of the oral microbiome in DS. For the sake of clarity, in DS we had one child (age 7), 11 teens, and 15 adults (the maximum age was 33). We also did the same comparison including only the adult DS samples, as well as only the child and teen samples together, but in both instances the results were essentially the same as when including all 27 DS samples (**Supplementary Figure 3.2a-b**). In a similar approach, we grouped the non-DS samples into smaller age bins of 10 years, with the exception of 10–20 years old, which was made into two groups (10–15 and 15–20) since the majority of samples were teenagers of age 15–16, but again the same conclusions could be drawn (**Supplementary Figure 3.2c**).

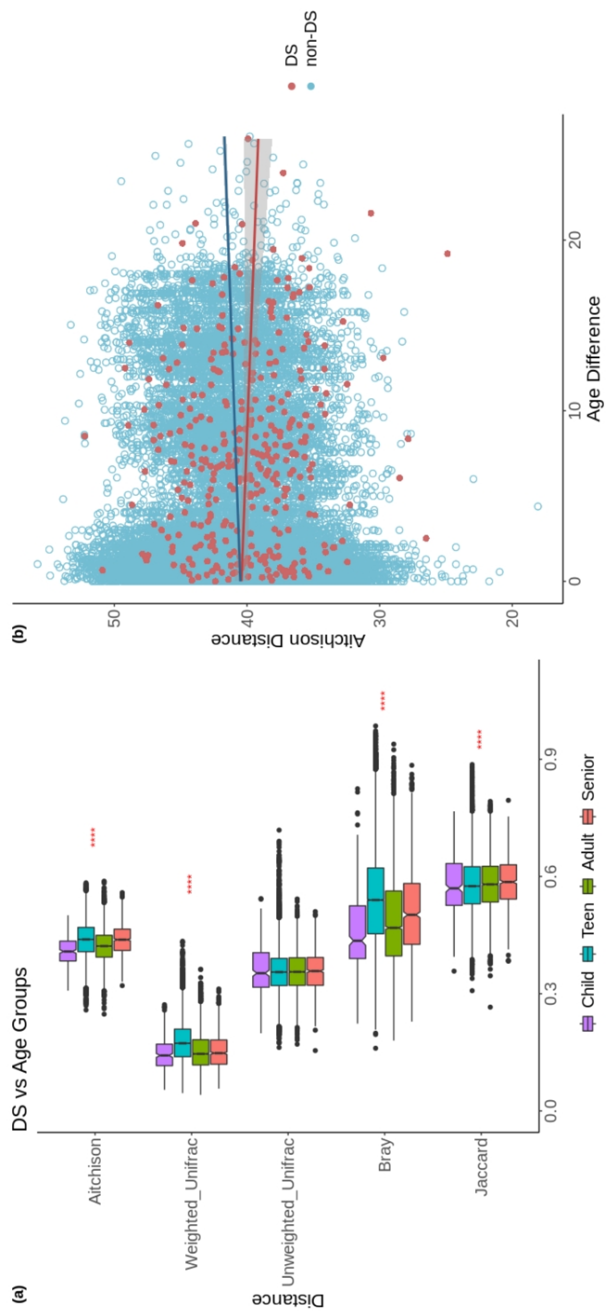

Figure 3.2: Comparisons across age. (**a**) Distributions of distances between DS

samples and non-DS samples of each age group based on five distance metrics. Red stars represent p-values for Kruskal-Wallis tests to compare the means of each age group within the comparison of a given distance metric. The representation of p-values are as follows: 0 ‘****’ 0.0001 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘ ’ Not significant. (**b**) Scatterplot of age differences between pairs of samples vs Aitchison distance between those samples. Red points represent DS samples compared to other DS samples, blue points represent non-DS samples compared to other non-DS samples, with corresponding lines of the best fit.

For the second approach, in order to determine whether changes in the microbiome across age were more extreme in DS, we looked for correlations between the difference in age between any pair of samples and the compositional distance between those samples, as calculated by the Aitchison distance. In this case, a premature aging effect might be inferred from a stronger positive correlation in DS samples between age difference and compositional distance than that seen in non-DS controls. However, as with the first approach, we did not see evidence of this effect in the DS samples, which actually showed a slightly negative correlation in these values which was not statistically significant (Pearson $r = -0.065$, $p = 0.08$) (**Figure 3.2b**). We compared this to non-DS samples of all ages ($n = 1{,}335$), as well as just those matched controls up to a maximum age of 33 ($n = 262$), in order to see the effect on the full range of ages, and just within the same age range as that of the DS samples, respectively. The full age range showed a statistically significant positive correlation between age difference and compositional distance (Pearson $r = 0.101$, $p = 0$). The matched controls up to age 33 showed a similar trend, though with a weaker correlation than that seen in the full age range (Pearson $r = 0.0499$, $p = 4.9e\text{-}39$). So, while there is evidence that there is a greater difference in overall composition as age difference increases between two non-DS samples, there is no evidence to suggest that this difference is more extreme in DS samples as a result of faster aging. In fact, this trend does not appear to occur at all in DS samples, though this may be

the result of the relatively limited number of DS samples compared to non-DS samples. There were no differences between DS and matched control samples in the trajectory of abundances of any particular genus across age. However, there were differences for reads unclassified at the phylum level, which show the opposite trend between DS and non-DS, with a decrease across age in non-DS samples and an increase in DS samples (**Supplementary Figure 3.3**).

3.4 Discussion

Our study provides a first snapshot of the oral microbiome in DS and how it compares to non-DS individuals. The results indicate a significant shift in the overall composition of the oral microbiome between DS and non-DS individuals, with significantly lower pH and species diversity, as well as a larger presence of periodontal pathogens and *Candida* in DS. Although many DS symptoms relate to premature aging, we did not find such a signature in the composition of the oral microbiome.

Regarding the differences in microbiome composition between DS and non-DS individuals, the five distance metrics for which we ran the permanova test each measures the distance between samples using different criteria, and all five of these calculations showed statistically significant differences between the sample groups. The Bray-Curtis dissimilarity is most heavily influenced by dominant taxa in the samples, while the Jaccard index is based on the presence or absence of taxa, and thus primarily measures differences in rare taxa. The weighted and unweighted UniFrac distances use similar considerations to those of Bray-Curtis and Jaccard, respectively, but add additional weights based on phylogenetic distances between taxa. The Aitchison distance is based on centered log-ratio values, and thus is robust to changes in variation of abundances despite potentially different relative abundances of taxa that result from the compositional nature of the

data. These results indicate strong differences between DS and non-DS across all levels of the oral microbiome, from rare taxa to the most dominant.

We do not have data on the incidence of any oral diseases in our samples, such as periodontitis or dental caries, as they are a subset of a larger exploratory study without an initial focus on oral disease (SLL2 – from 'Saca La Lengua' in Spanish, see http://www.sacalalengua.org, **Chapter 2**, (Willis et al. 2018)). Nevertheless, our results do provide a platform from which to reflect on findings in the literature on oral health in both the general population and in DS. We found significant differences between DS and matched control samples in the abundances of a number of key genera of bacteria that are complicit in the pathogenesis of some oral diseases, particularly periodontitis. This follows with the increased incidence of periodontal disease in DS (Cichon, Crawford, and Grimm 1998; A. Khocht et al. 2012; Kornman 2008; Barr-Agholme et al. 1992; Meskin, Farsht, and Anderson 1968; Sakellari, Arapostathis, and Konstantinidis 2005), which has been posited to result from a number of factors, including diminished salivary flow leading to a reduced immune response in the oral cavity (Ahmed Khocht 2011; Areias et al. 2012; Chaushu et al. 2002; Domingues et al. 2017), as well as difficulties in dental treatment (Cheng, Yiu, and Leung 2011; Down's Heart Group n.d.; Pilcher 1998).

However, we find a less straightforward connection to dental caries, wherein differential abundances of particular taxa suggest a non-caries environment, while the lower alpha diversity and low salivary pH suggest the potential occurrence of caries in DS samples. The literature has generally shown either lower incidence of caries in DS or no significant difference compared to non-DS (Deps et al. 2015; Moreira et al. 2016). This has been explained by the relatively late eruption of teeth in DS, microdontia, more missing teeth and greater dental spacing (Cheng, Yiu, and Leung 2011; Vigild 1986). Many of the studies that, in the context of the differential abundances in our data, would suggest a low incidence of dental caries in our DS samples, were conducted with samples from young children (Aas et

al. 2008; Lif Holgerson et al. 2015; Torlakovic et al. 2012; Johansson et al. 2016; L. Xu et al. 2018; Espinoza et al. 2018; Yanhui Li et al. 2016), and so are likely to represent early stages of cariogenesis. It may be that DS has a low incidence of dental caries because, despite typically worse oral hygiene, the unique dentition does not allow for optimal growth of many of the early plaque colonizers that initiate caries, while promoting the growth of organisms associated with the lack of caries. Nonetheless, the combination of poor oral health with the increase in acidogenic organisms in our DS samples and a low alpha diversity still may suggest a potentially cariogenic environment in DS.

The genera *Kingella* and *Cardiobacterium*, more abundant in DS samples in this study, form part of the HACEK group (made up of *Haemophilus*, *Aggregatibacter*, *Cardiobacterium*, *Eikenella*, and *Kingella*), which are the primary pathogens in infective endocarditis (Chambers et al. 2013). This is a disease strongly linked to periodontal disease (Dhotre, Davane, and Nagoba 2017; Carinci et al. 2018; Ninomiya et al. 2020), and one which can also be linked to individuals with DS (Down's Heart Group n.d.). Increases in *Kingella* (C. Chen 1996; Ruhl et al. 2014; Coelho et al. 2014; F. E. Dewhirst et al. 1993) and *Cardiobacterium* (Lourenço et al. 2014; X. Y. Han and Falsen 2005) in the oral cavity have been associated with the pathogenesis of periodontitis, regardless of their involvement in endocarditis. On the other hand, *Kingella* has been found at lower abundance in samples with dental caries than in healthy controls (Aas et al. 2008; Lif Holgerson et al. 2015; Torlakovic et al. 2012), though it is acidogenic (F. E. Dewhirst et al. 1993), a trait which may promote caries. Thus, the interpretation of the connection between *Kingella* and caries in the context of DS may be a bit muddled, but our data do follow the trend in the literature of low caries incidence in DS.

The other HACEK genera of *Haemophilus*, *Aggregatibacter*, and *Eikenella* were not differentially abundant in our dataset, but the genus *Actinobacillus* was more abundant in DS. This may be interesting in that *Aggregatibacter* is a relatively recently described genus made up of two former organisms of *Haemophilus* and the

former *Actinobacillus actinomycetemcomitans* (Nørskov-Lauritsen and Kilian 2006), the latter of which (now labeled *Aggregatibacter actinomycetemcomitans*) is implicated in the pathogenesis of an aggressive form of periodontitis, acting to prime a site for growth of other periodontal pathogens (Fine, Patil, and Velusamy 2019). Thus, while *Aggregatibacter* itself is not differentially abundant in these DS samples, the increase of the closely related *Actinobacillus* may represent a similar effect upon oral health in DS.

The literature is less certain on the roles of some other organisms in periodontitis. DS samples also showed higher abundance of the genus *Staphylococcus*. The species *Staphylococcus aureus* is of particular interest, as it has been found to be both associated with periodontal health (Vieira Colombo et al. 2016) and periodontal disease (Fritschi, Albert-Kiszely, and Persson 2008; A. V. Colombo et al. 2013; Loberto et al. 2004). Another study found that there was no difference in abundance between healthy and diseased sites of individuals with periodontitis, and suggests that *Staphylococcus* is not involved directly in the pathogenesis, but rather acts synergistically to promote the growth of other pathogens (dos Santos et al. 2014), a notion corroborated in another study saying that virulence factors of *S. aureus* may work to form biofilms along with other periodontal pathogens (Kim and Lee 2015). In addition, *S. aureus* is the world's most prominent pathogen of infective endocarditis (Rajani and Klein 2020), a disease which may be linked to bacteremia resulting from dental procedures or even tooth brushing in individuals with poor periodontal health (Kinane et al. 2005; Lockhart et al. 2008, 2009). The conjunction of DS with increased prevalence of periodontal disease and infective endocarditis illuminates the potential roles of *Staphylococcus* and the organisms of the HACEK group which we have seen at increased abundance in DS samples in this study.

Gemella and *Rothia*, also found at higher abundances in DS here, have had similar links to these DS-associated health concerns, as seen with *Staphylococcus*. Both *Gemella* (P. S. Kumar et al. 2005) and *Rothia* (Kistler et al. 2013; Griffen et al. 2012) have been associated with periodontal health. Conversely, *Gemella* (Al-

Jebouri 2016) and *Rothia* (Ramanan et al. 2014) have been linked to periodontal disease in other studies. Furthermore, both *Gemella* (Al Soub et al. 2003; Akiyama et al. 2001) and *Rothia* (Ricaurte et al. 2001; Fridman et al. 2016) have been shown to occasionally cause endocarditis. One study (A. P. V. Colombo et al. 2012) showed that the species *Gemella sanguinis, Gemella haemolysans*, and *Rothia dentocariosa* were prevalent in infected sites of individuals with chronic periodontitis. They remained prevalent in infected sites of individuals that did not respond well to treatment. *G. sanguinis* even increased in these sites, but the species *G. haemolysans* and *R. dentocariosa* actually increased in individuals that responded well to treatment (the species associated with therapeutic success also included *Cardiobacterium hominis* and *Kingella oralis*, from two other genera that were higher in DS samples in our study) (A. P. V. Colombo et al. 2012). It is difficult to conclude from this whether these species would then be associated with periodontal health, or if rather they were involved in particular stages of periodontal disease and, after treatment which worsened conditions for other organisms, were then primed to thrive in the now healthier host environment. For example, it has been posited that *R. dentocariosa* and *K. oralis*, among other organisms, are important for biofilm formation that may be used by other pathogens, due to the ability to form strong adhesive interactions (Ruhl et al. 2014). *R. dentocariosa* has also been shown to induce the production of tumor necrosis factor α (TNF-α) in the oral cavity (Kataoka et al. 2014), leading to periodontal inflammation. Therefore, these species may be important in the initial pathogenesis of periodontitis, and then after a successful treatment that removes other competitors that have taken advantage of their biofilms, they are able to prosper. Regardless, the generally poorer oral health in DS may create the conditions necessary at least for the initial stages of periodontal disease, allowing for increased abundances of those organisms we have seen in this study. The increased *Rothia* in DS here also follows the results of the literature regarding dental caries, which show that *Rothia* is typically associated with low caries incidence (L. Xu et al. 2018; J. L. Baker et al. 2019), potentially due in part to nitrate reduction and the production of enterobactin, which can

reduce the growth of some acidogenic and cariogenic organisms (Uranga et al. 2020). *Gemella* has been conflictingly associated both with dental caries (Yanhui Li et al. 2016) and with relative health (Aas et al. 2008). However, due its high degree of autoaggregation, which can allow for the initiation of plaque (Shen, Samaranayake, and Yip 2005), the species *Gemella morbillorum* may only be prominent in the initial stages of cariogenesis.

Apart from these relatively well-studied organisms, there was also an increased abundance in DS samples of reads that could not be classified at the phylum level. That would suggest that there is an increase in some rare and/or understudied bacteria, for which a *16S rRNA* gene sequence is not available. No conclusive argument can be made as to the impact or cause of this increase in DS samples without knowing what they are, but for future studies it may be worthwhile to take note of any rare organism with an increased presence in DS samples. In this regard, whole-genome shotgun studies of the oral microbiome in DS will be helpful.

DS samples had lower abundances of the genera *Alloprevotella* (and its phylum *Bacteroidetes* as a whole), *Atopobium*, and *Candidatus_Saccharimonas* (as well as its phylum *Patescibacteria*). As with some of the other genera mentioned already, *Alloprevotella* and *Atopobium* have been shown in some instances to be associated with periodontitis (Lourenço et al. 2014; Casey Chen et al. 2018; Camelo-Castillo et al. 2015), and in others to not be significantly different between periodontitis and health (Wolff et al. 2017; Yuchao Li et al. 2020). Respectively, these studies showed that, while neither was different between periodontitis and health, subgingival *Alloprevotella* was associated with the early stages of rheumatoid arthritis, and *Atopobium* was significantly associated with gingival squamous cell carcinoma. However, *Alloprevotella* has been shown to be involved in nitrate reduction in saliva as an initial step in the circulation of nitrate between the saliva and digestive tract (Espinoza et al. 2018). This nitrate/nitrite cycle is part of the host defense reaction against periodontal disease (Qu et al. 2016), since salivary glands boost the immune response as a reaction to periodontitis (Henskens et al. 1996). Thus, the lower

abundance of *Alloprevotella* in DS samples could be a result of the decreased salivary flow in DS, which may inhibit the overall nitrate availability to organisms like *Alloprevotella*, diminishing its ability both to thrive and to contribute to the host immune response in the pathogenesis of periodontitis. One study on the production of nitric oxide (NO) in the airways of cystic fibrosis patients, a process which also creates an anti-inflammatory effect, found that increasing dietary nitrate led to greater exhaled NO as compared to placebo treatments (Kerley et al. 2016). The NO is produced by the same pathway of nitrate reduction by oral and airway bacteria, so increasing the nitrate intake in DS individuals may help to promote the growth of *Alloprevotella* and other nitrate reducers to boost the immune response, though this would require further study in the context of DS. Both *Alloprevotella* (Johansson et al. 2016; L. Xu et al. 2018; Espinoza et al. 2018; Uchida-Fukuhara et al. 2020) and *Atopobium* (Aas et al. 2008; L. Xu et al. 2018; Kianoush et al. 2014) have been shown to be associated with increased incidence of dental caries, so their low abundances in DS here again support the low caries in DS seen in the literature. The lower abundance of the genus *Candidatus_Saccharimonas*, of the fairly recently described *Patescibacteria* phylum (Rinke et al. 2013), does not seem to show the same connection from the literature, as this genus (Espinoza et al. 2018) and its family, *Saccharimonadaceae* (Schoilew et al. 2019), were shown to be found at lower abundances in individuals with dental caries than in healthy controls. This is the only taxon in this study that, on its own, would suggest greater incidence of dental caries in DS.

The alpha diversity was lower in DS samples than matched controls in this study, meaning the compositions of the DS samples were dominated more by particular organisms than those of the controls. The lower diversity is actually more reminiscent of the pathogenesis of dental caries than that of periodontitis. It has been shown that alpha diversity is lower in severe early childhood caries (Y. Li et al. 2007; Hurley et al. 2019) and that it decreases as dental caries progresses over time (Gross et al. 2010). An explanation that has been posited for this was called the ‘ecological plaque hypothesis’ (P. D. Marsh 1994), which suggests that acidogenic bacteria lower

the pH and the diversity is reduced as other species intolerant of the acidic environment are inhibited. Conversely, alpha diversity has been shown to be higher in periodontitis (LaMonte et al. 2018; Abusleme et al. 2013) and increases with increasing severity of the disease (Genco et al. 2019), as well as with particular periodontitis indicators, like periodontal pocket depth (PPD), bleeding on probing (BOP), and mean plaque index (Takeshita et al. 2016). As we do not have data related to specific aspects of the dental health in the samples of this study, we can only speak to the trend in our data that suggests generally worse oral health in DS samples than in matched controls, whether the characteristics be reminiscent of the pathogenesis of periodontitis or dental caries. Furthermore, the low pH that was seen in DS samples here, is important to cariogenesis according to the ecological plaque hypothesis, and has also been seen in periodontitis (Baliga, Muglikar, and Kale 2013; Prasad et al. 2019). Salivary pH has also been shown to be lower in DS than in non-DS (Davidovich et al. 2010; Siqueira and Nicolau 2002).

Yeasts were generally found to be more prevalent in DS, consistent with earlier reports (Maranhão et al. 2020). *Candida* species that most prominently differed with respect to matched controls were *C. parapsilosis* and, to a less significant extent, *C. dubliniensis*. Both of these species are opportunistic pathogens in the oral cavity that have been associated with periodontitis (M. I. Brusca et al. 2010; Urzúa et al. 2008; Jewtuchowicz et al. 2008; McManus et al. 2012) and dental caries (Naidu and Reginald 2016; Lozano Moraga et al. 2017; Al-Ahmad et al. 2016). One study, however, from the Basque Country in Spain, actually found lower levels of both *C. parapsilosis* and *C. dubliniensis* in periodontitis as compared to controls (De-La-Torre et al. 2018). They suggested that this may be due to geographical differences with the other studies, as the pathogenesis of some fungal infections has been linked to geographical location (X.-B. Zhang et al. 2012; Krom, Kidwai, and Ten Cate 2014). Nevertheless, location cannot explain the differences from that study to ours, as all of our samples came from Spain, and nine of the 27 came specifically from the Basque Country. Thus, despite the findings of potential geographical aetiologies for fungal infections, DS may have a stronger impact on

the growth of these *Candida* species (assuming at least a periodontitis-like environment in our DS samples, for the sake of comparison, which the evidence does suggest). For instance, studies have found negative correlations between salivary flow rates and the number of *Candida* colony forming units (CFUs) (Nadig et al. 2017; S. R. Torres et al. 2002), so the low salivary flow in DS may promote oral *Candida* growth. Oral *C. dubliniensis* has generally been associated with immunocompromised individuals, and the hindered immune responses in the oral cavity in DS have already been touched upon here. *C. parapsilosis* has been shown to be acidogenic and able to induce salivary proteolysis via the production of secreted aspartyl proteinases (T. Wu and Samaranayake 1999) and to demineralize tooth enamel (Caroline de Abreu Brandi et al. 2016). Thus, the fungal component of our data suggests an environment suitable to both periodontitis and cariogenesis in DS.

Between the 20 relatives of DS individuals and the matched controls used with the analyses of the DS samples, there were no differences in either abundances of particular bacterial taxa, alpha diversity, pH, or the prevalence of yeast species. Of these 20 relatives, 18 were parents of DS individuals and two were siblings. One of the sibling pairs were of the ages 7 and 10 years old, and the other pair was 22 and 28 years old. So, while we do not have information regarding the living situations of the participants of the study, we assume that most of the relatives live in the same households as the DS participants. Thus, the relevance of not finding any differences between relatives and the matched controls is that we can say with greater confidence that the differences found in DS samples are related to this disorder, and are likely not trends that are shared within a household or due to particular lifestyle factors that are specific to individual households or families.

Despite findings that have shown various effects of premature aging in DS (Franceschi et al. 2018; Horvath et al. 2015; Borelli et al. 2015), there was no such effect apparent from the oral microbiome in DS in this study. DS samples were actually most similar to those of non-DS children and least similar to those of non-DS teens.

Furthermore, while non-DS samples showed significantly greater differences in overall composition as age differences between samples increased, the DS samples did not show any trend. This may have been due to a lack of statistical power compared to the tests in the non-DS samples, or may suggest that there is even an 'anti-aging' effect on the oral microbiomes of DS individuals, in that the evolution of the composition of the oral microbiome occurs at a slower rate in DS than in healthy populations (though this claim would require more extensive analysis with a larger sample size and a longitudinal study design). A hypothesis for such a phenomenon might be that the oral microbiome composition in DS individuals is less dynamic than in non-DS individuals due to the limited salivary flow, unique dentition, and hindered immune responses in the oral cavity, creating a relatively static environment. If so, this idea may even help to explain the comparative lack of cariogenesis and some cariogenic organisms in DS, despite an apparently cariogenic environment. The confluence of factors may promote the formation of periodontal biofilms that can be maintained, but may be less favorable to the dental plaques that erode the hard tissues of teeth in dental caries.

3.5 Conclusions

This study has shown that there are significant differences in the overall composition of the oral microbiome between DS and non-DS individuals from across Spain. Differences in the abundances of particular organisms, both bacterial and fungal, follow what has been seen in the literature for increased incidence of periodontitis and decreased incidence of dental caries, and both of these tendencies have typically been seen in DS individuals. Nonetheless, the low pH and alpha diversity seen in DS samples do present a potentially cariogenic environment, but it may be that the unique dentition typically seen in DS impedes the early colonizers of dental plaques. Although there are various physiological signs of premature aging in DS, we have found no evidence of this

phenomenon within the oral microbiome. There was nothing to suggest either that the oral microbiome in DS is more similar to older non-DS individuals, or that changes in the microbiome across age were more extreme in DS. Rather, there appears to be a relatively static microbial environment, potentially due to the limited salivary flow, unique dentition, and poor immune responses, though this hypothesis would require further exploration in a larger longitudinal study. Taken together, these results provide a glimpse into the distinctive oral microbiome in Down Syndrome and allow for a deeper understanding of the oral health trends therein.

Supplementary information

Supplementary information referenced in this chapter accompanies the publication at https://doi.org/10.1080/20002297.2020.1865690.

Chapter 4: Citizen-science based study of the oral microbiome in cystic fibrosis and matched controls reveals major differences in diversity and abundance of bacterial and fungal species

Willis JR, Saus E, Iraola-Guzmán S, Cabello-Yeves E, Ksiezopolska E, Cozzuto L, Bejarano LA, Andreu-Somavilla N, Alloza-Trabado M, Blanco A, Puig-Sola A, Broglio E, Carolis C, Ponomarenko J, Hecht J, & Gabaldón T. (2021). Citizen-science based study of the oral microbiome in cystic fibrosis and matched controls reveals major differences in diversity and abundance of bacterial and fungal species. *J Oral Microbiol*. 13(1). doi: 10.1080/20002297.2021.1897328

Chapter 4: Citizen-science based study of the oral microbiome in cystic fibrosis and matched controls reveals major differences in diversity and abundance of bacterial and fungal species

4.1 Abstract

Background: Cystic fibrosis (CF) is an autosomal genetic disease, associated with the production of excessively thick mucosa and with life-threatening chronic lung infections. The microbiota of the oral cavity can act as a reservoir or as a barrier for infectious microorganisms that can colonize the lungs. However, the specific composition of the oral microbiome in CF is poorly understood. **Methods**: In collaboration with CF associations in Spain, we collected oral rinse samples from 31 CF persons (age range 7-47) and matched controls, and then performed 16S rRNA metabarcoding and high-throughput sequencing, combined with culture and proteomics-based identification of fungi to survey the bacterial and fungal oral microbiome. **Results**: We found that CF is associated with less diverse oral microbiomes, which were characterized by higher prevalence of Candida albicans and differential abundances of a number of bacterial taxa that have implications in both the connection to lung infections in CF, as well as potential oral health concerns, particularly periodontitis and dental caries. **Conclusion**: Overall, our study provides a first global snapshot of the oral microbiome in CF. Future studies are required to establish the relationships between the composition of the oral and lung microbiomes in CF.

Keywords: Oral microbiome, Cystic fibrosis, oral mycobiome, *Candida*, *Pseudomonas*

4.2 Introduction

Cystic fibrosis (CF) is a severe autosomal recessive genetic disease caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene (Zielenski et al. 1991). CF is one of the most common rare genetic disorders, particularly in the Caucasian population, affecting one in 2000–3000 newborns in the European Union (Bassett, Boguski, and Hieter 1996). The CFTR protein acts as a chloride channel that transports ions across the apical membrane of epithelial cells throughout the body (Saint-Criq and Gray 2017). This channel is involved in the production of several secretions including sweat, digestive fluids, and mucus. If the channel is impaired, these secretions increase their thickness, mostly affecting the function of the lungs, but also other organs such as the pancreas, liver, kidneys, and intestine. Lung infections are common and often develop into chronic and severe, life-threatening forms due to a deficient mucociliary clearance of the thick mucus (Tilley et al. 2015). Several bacterial and fungal species are commonly associated with chronic respiratory infection of the lower airways in CF, including *Aspergillus fumigatus, Candida albicans, Haemophilus influenzae*, *Staphylococcus aureus, Burkholderia cepacia complex, Stenotrophomonas maltophilia*, *Achromobacter xylosoxidans* and *Pseudomonas aeruginosa*, with *P. aeruginosa* playing a major role in the morbidity and mortality of patients (C. J. Taylor et al. 1990; Marshall et al. 2015). Hence, *P. aeruginosa* is a key pathogen in CF lung disease and has been found to be involved in the progressive obstructive pulmonary disease and bronchiectasis resulting from chronic endobronchial infection (Rosenfeld et al. 2012). These infections start early in life and progressively increase with patient age, with *P. aeruginosa* present in the lungs of up to 80% of patients over the age of 18 years (Saiman, Siegel, and Cystic Fibrosis Foundation Consensus Conference on Infection Control Participants 2003).

Despite the fact that *P. aeruginosa* is one of the most widespread and destructive opportunistic pathogens, it does not colonize the airways alone. Microbes commonly present in the oral cavity are also present in sputum from CF patients (Rivas Caldas et al. 2015)].

In fact, the oral cavity has been proposed as a reservoir of bacteria, both commensal and pathogenic, that can colonize the lower airways due to micro-aspirations (Boutin et al. 2015; Gomes-Filho, Passos, and Seixas da Cruz 2010). A model has been proposed comparing the biodiversity of the respiratory tract to island biogeography (Whiteson, Bailey, et al. 2014), in which the mouth and throat, much like the mainland, are sources of relatively high species diversity, whereas the diversity of the airways decreases with the distance from the mouth, just as distant islands display more specific subsets of the mainland's diversity. An investigation of this notion showed that samples taken from distal lung sites were more distinct from the upper respiratory tract than proximal sites (Dickson et al. 2015). Thus, there might be a strong connection between the oral microbiome and many of the pulmonary pathogens acting in CF.

Beyond the mouth's microbiota, the physiology of the oral cavity is also a factor in CF. Both salivary pH and the pH of the airways in CF are typically lower than those of non-CF individuals (Pawlaczyk-Kamieńska, Borysewicz-Lewicka, and Batura-Gabryel 2019; Tate et al. 2002), due in part to malfunction of the CFTR protein, leading to defective secretion of the buffer molecule bicarbonate (Kunzelmann, Schreiber, and Hadorn 2017). The airway microbiota in CF have also shown a decrease in alpha diversity, a measure of the number of organisms present in a sample, and a value that decreases more with diminishing lung function (Cuthbertson et al. 2020; Blainey et al. 2012; Coburn et al. 2015; M. J. Cox et al. 2010; Zemanick et al. 2017). Taken together, these factors have important implications on the microbiome and oral health in general, though there is some debate in the literature. The acidic environment may leave CF individuals more susceptible to both dental caries and periodontitis (Baliga, Muglikar, and Kale 2013; Prasad et al. 2019). The combination of low pH and low alpha diversity may predispose CF individuals to dental caries in particular, as according to the 'ecological plaque hypothesis,' acidogenic bacteria foster an acidic environment and the diversity drops as many species intolerant to the change are unable to grow (P. D. Marsh 1994). Periodontitis, on the other hand, has been

shown to present with greater alpha diversity in the oral cavity (LaMonte et al. 2018; Abusleme et al. 2013; Genco et al. 2019; Takeshita et al. 2016). While some studies suggest a greater risk for dental caries in CF (Pawlaczyk-Kamieńska, Borysewicz-Lewicka, and Batura-Gabryel 2019; Catalán et al. 2011), a systematic literature review found that CF generally had lower incidence of caries, or no difference (Pawlaczyk-Kamieńska et al. 2019). The CFTR protein's role of pH regulation is vital in odontoblasts, cells which secrete dentin during tooth development, and ameloblasts, cells which deposit enamel, and in fact, abnormal enamel mineralization has been seen in CF as a result of the defective CFTR protein (Arquitt, Boyd, and Wright 2002; Bronckers et al. 2010). Conversely, it has been suggested that the CFTR protein may actually promote periodontitis, and indeed a study of gingival biopsies showed greater and more widespread *CFTR* expression in patients with periodontitis compared to healthy controls (Ajonuma et al. 2010), suggesting that a mutation in the *CFTR* gene could predispose CF individuals to have better periodontal health. The review of oral health in CF also suggested that treatments for CF, such as antibiotics and inhaled anti-inflammatory medications, may protect CF patients from the colonization of early caries and periodontitis pathogens (Pawlaczyk-Kamieńska et al. 2019). One study found that airway pH increased in CF patients after antibiotic treatment (Tate et al. 2002), and there have been instances of higher pH in CF as compared to controls, which have been attributed to treatment with supplementary pancreatic enzymes (Herman, Kowalczyk-Zając, and Pytrus 2017). The evidence suggests that the combination of factors typical to CF may inhibit cariogenic and periodontal pathogens in the oral cavity, but still produces an environment with a low pH and low alpha diversity that promotes damage to the outer layers of the teeth.

Here we present the first metabarcoding study, to our knowledge, of the oral microbiome in cystic fibrosis with a comparison to matched control samples. We compare the overall composition of oral bacteria between the groups and explore the differential abundances of bacterial taxa, as well as the presence or absence of yeast and mold species. We also calculate co-occurrence networks to study

the underlying ecologies present in the CF and non-CF groups and further analyze the results alongside metadata collected through a citizen science approach.

4.3 Materials and methods

4.3.1 Sample collection

We contacted CF individuals and their relatives, through local associations of CF families integrated in the Spanish Federation for CF (https://fibrosisquistica.org/). Additionally, we analyzed samples from matched groups of individuals (see below) from a larger study targeting the general population (www.sacalalengua.org). All participants signed an informed consent form allowing the use of their saliva samples for microbiological research. For participants under the age of 18, the consent form was also signed by one of the parents or a legal guardian. This project was approved by the ethics committee of the Barcelona Biomedical Research Park (PRBB). Samples were collected from January to November 2017. Participants were asked not to ingest any food or beverage (except water) for 1 h before collecting the sample. All donors received clear indications about the sample collection procedure in person, and the collection of the samples was carried out with the assistance of a researcher involved in the project, following a demonstration. All participants responded to a uniform questionnaire (see below), which was adapted for CF in collaboration with CF partner associations. Before collection of the oral rinse, the pH of the saliva was measured using pH test strips (MColorpHast, Merck, range 5.0–10.0; 0.5 accuracy units), the accuracy of which have been previously validated (**Chapter 2**, (Willis et al. 2018)). Saliva samples were collected using a mouthwash as described earlier (**Chapter 2**, (Willis et al. 2018)). In brief, the protocol is as follows: participants rinsed their mouth with 15 mL of sterile phosphate-buffered saline (PBS) solution, for 1 min. Then, they returned the liquid into a 50 mL tube. The samples were then centrifuged at 4,500 g for 12 min at room temperature (r.t.) in an Eppendorf 5430

centrifuge equipped with an Eppendorf F-35-6-30 rotor. The supernatant was discarded and the pellets were resuspended with the remaining PBS, transferred to 1.5 ml tubes and centrifuged at 4,500 g for an additional 5 min at r.t. using an Eppendorf FA-45-24-11-HS rotor. Supernatants were discarded, and pellets were frozen and stored at − 80°C until further analysis.

A total of 31 oral rinse samples were collected from individuals with CF (ages 7–47) during the second edition of ‘Stick out your tongue’, a citizen science project in Spain (SLL2 – from ‘Saca La Lengua’ in Spanish, see http://www.sacalalengua.org, **Chapter 2**, (Willis et al. 2018)), and in collaboration with CF family associations in Spain. Citizen science aims to involve the public in large-scale investigations to obtain a wide range of data as well as to increase public understanding of relatively esoteric scientific undertakings (National Academies of Sciences, Engineering, and Medicine et al. 2019; Gura 2013). Sample collection was coupled with science communication activities with CF individuals and their relatives, aiming to raise awareness about the microbiome, its role on health and disease, and its potential particularities in CF. The SLL2 project questionnaire about health and lifestyle was adapted to CF with the help of CF associations.

In order to determine the effects of familial relations in the context of this study, we also collected samples from some relatives. Some of the 31 CF individuals were related to others also with CF, including 8 siblings and 2 individuals that were partners. There were 36 other relatives without CF that also participated in the study, which primarily consisted of parents, but also included 2 other partners, 4 siblings, and 1 grandmother. The ages of the siblings with CF ranged from 7 to 41 years old, while the siblings without CF ranged from 22 to 32 years old. The parents ranged from 41 to 71 years old.

4.3.2 DNA extraction and sequencing

The DNA extraction and amplification and sequencing of the V3-V4 region of the 16S ribosomal RNA gene were performed as

previously described (**Chapter 3**, (Willis et al. 2020)). Briefly, for sample DNA extraction we used the ZR-96 Fungal/Bacterial DNA kit (Zymo research Ref D6006), following manufacturer's instructions. Two DNA samples derived from bacterial mock communities from the BEI Resources of the Human Microbiome Project were included as controls: 'HM-782D' and 'HM-783D'. V3-V4 16S primers, PCR conditions, and library preparation for further Illumina MiSeq sequencing in multiplex with 2×300 bp reads using v3 chemistry were done following our previous protocols already described (**Chapter 3**, (Willis et al. 2020)). We also included negative controls for the DNA extraction and PCR amplification steps, which provided no visible band or quantifiable amount of DNA by gel visualization or Bioanalyzer, whereas all samples resulted in clearly visible bands after 20 cycles. Twelve such controls were subjected to library preparation and sequenced. Expectedly, these sequenced non-template controls systematically yielded very few reads (a range of 155–1005 reads per sample), in contrast to an average of ~64,000 reads/sample in sample-derived libraries.

4.3.3 Fungal composition analysis

To assess the fungal composition in our samples, we used traditional culture-based methods to enrich for possible fungal species instead of fungal metagenomics technologies (such as ITS amplification), which produced unsatisfactory results in this type of samples in previous studies (**Chapter 3**, (Willis et al. 2020)). In this way, we could overcome the limitations of the low presence of fungi in oral rinse samples and the difficulty to break the fungal cell wall to access the DNA in comparison to bacteria. We used previously optimized protocols in our group (**Chapter 3**, (Willis et al. 2020)), which mainly consists of plating 6% of the original sample onto a YPD sterile plate with antibiotics (100 µg/ml of chloramphenicol and 100 µg/ml of ampicillin). After 7 days of incubation at 30°C, we assessed the number of colonies, their phenotypes, and the presence of bacteria. A maximum of 10 colonies were randomly selected per sample and re-grown onto a

fresh plate for 24 h in the same conditions. We used MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization-Time Of Flight) mass spectrometry analysis for fungal identification of each colony and, if the results were inconclusive after two attempts, we complemented this analysis with colony PCR to amplify the Internal Transcribed Spacer (ITS) hypervariable region of the 5.8S ribosomal gene (fungal marker) and further Sanger sequencing.

MALDI-TOF analysis was performed with a MALDI Biotyper (Bruker Daltonik MALDI Biotyper) in the Centre for Omics Sciences (COS) in EureCat (Centre Tecnològic de Catalunya, Reus, Spain), following previously described protocols (**Chapter 3**, (Willis et al. 2020)): total proteins from the fungal colonies were extracted in our lab following standard protocols and then samples were sent to COS for the analysis. Samples were deposited in duplicate on a Polished Steel Target Plate (Bruker), coated with the matrix α-Cyano-4-hydroxycinnamic (HCCA) and analyzed with MALDI-TOF/TOF (MALDI ultrafleXtreme, Bruker, Germany). Spectrum identification was performed using the real-time classification software MALDI Biotyper (Bruker, Germany). Thus, for each colony sample, we obtained two alternative scores, which were interpreted as follows: i) 2.3–3.0, highly probable species identification; ii) 2.0–2.299, secure genus identification, probable species identification; iii) 1.7–1.999, probable genus identification; iv) <1.7, not reliable information. Samples identified with scores lower than 2.0 were manually re-analyzed, and its spectrum was classified using the Biotyper Offline Classification software. We considered that the results were consistent with species identification when the best match had a score >2.0 and the second best match was at least >1.7 with at least the same genus as the first one. The remaining samples were re-grown, and the whole process was repeated. If after two more experiments with MALDI-TOF analysis the identification of the fungal species was still inconclusive, we performed ITS amplification and sequencing.

For ITS-Sequencing, we performed a colony PCR from fresh colonies, directly using biomass material from a recently replated colony (maximum of 12–24 hours old). To do this colony PCR we

used DongSheng Biotech (DSBio) Taq mix (#2012) and 20 pmol of each primer (ITS1: 5′-TCCGTAGGTGAACCTGCGG-3′; NL4: 5′-GGTCCGTGTTTCAAGACGG-3′) in a total volume of 40 µl, and the following conditions: 5 min at 94°C, then 30 cycles of 30 s at 94°C, 30 s at 55°C, 1.5 min at 72°C, and a final extension of 5 min at 72°C. PCR products were purified with QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions. Samples were sent for Sanger Sequencing to Eurofins Genomics sequencing service, following SUPREMERUN recommendations. The resulting sequences were evaluated with the webtool Blast (Altschul et al. 1990). Only samples with a high score (>95%) of identity were considered as correctly identified.

In summary, yeast colonies analyzed from a total of 75 samples could be properly identified by MALDI-TOF. For the rest of the samples (17) further analysis with ITS amplification and Sanger sequencing was required: 11 of them could be identified with certainty (percentage of identity >95% in Blast). Additionally, for some samples, we were able to identify the predominant species responsible for bacterial growth on the YPD plates with antibiotics.

4.3.4 Pre-processing of 16S rRNA sequence reads and taxonomy assignment

Sequence reads from fastq files were filtered using the 'dada2' R package (version 1.10.1) (Benjamin J. Callahan et al. 2016) to produce counts of amplicon sequence variants (ASVs). Low quality reads were first removed by applying the *filterAndTrim* function with the following parameters: forward and reverse reads were trimmed to lengths of 275 and 230 nucleotides, respectively (truncLen = c(275,230)); the leading 10 nucleotides were trimmed in both reads (trimLeft = c(10,10)); reads with maximum expected errors greater than 5 in both reads were discarded (maxEE = c(5,5)); all other parameters used the default values. The remainder of the pipeline followed the suggestions in the tutorial from the authors of the tool (("DADA2 Pipeline Tutorial (1.16)" n.d.), https://benjjneb.github.io/dada2/tutorial.html). Taxonomy was

assigned using the dada2-formatted database of SILVA version 132 (B. Callahan 2018). A phylogenetic tree for use in UniFrac distance calculations was generated by following a protocol (Ben J. Callahan et al. 2016) that uses the ‘DECIPHER’ (version 2.10.2) (Wright 2016) and ‘phangorn’ (version 2.5.5) (Schliep 2011) R packages. After processing reads with the dada2 pipeline, only those samples with at least 1,000 reads were retained. At the end of this process, there were a total of 1,648 samples, though only a portion of those were used for the analyses of this study, as described below in the section ‘Statistical analyses’.

For analyses regarding the abundances of taxa, a centered log ratio transformation was applied to the ASV counts. Zeros were first replaced with the ‘count zero multiplicative’ method in the *cmultRepl* function from the ‘zCompositions’ R package (version 1.3.4) (Palarea-Albaladejo and Martín-Fernández 2015). Then centered log ratios were calculated using the *codaSeq.clr* function from the ‘CoDaSeq’ R package (Gloor et al. 2016; Gloor and Reid 2016).

4.3.5 Diversity measures

Alpha diversity measures were calculated using the *estimate_richness* function from the ‘phyloseq’ R package (version 1.30.0) (McMurdie and Holmes 2013). For beta diversity measures, both the weighted and unweighted UniFrac distances, which weight dissimilarity between samples by phylogenetic distances between taxa, were calculated using the *UniFrac* function from the ‘phyloseq’ package. The weighted UniFrac distances give additional weight based on taxa abundances. Bray-Curtis and Jaccard distances were calculated using the *vegdist* function from the ‘vegan’ R package (version 2.5–6) (Oksanen et al. 2017). As the Jaccard distance is based on the presence or absence of taxa, the *decostand* function, also from the ‘vegan’ package, was applied to the ASV counts table, using the method ‘pa’ for presence/absence, before the *vegdist* function was applied. The Aitchison distance was calculated using the *aDist* function from the ‘robCompositions’ R

package (version 2.2.1) (Templ, Hron, and Filzmoser 2011).

4.3.6 Statistical analyses

When running statistical tests, we first randomly selected representative matched non-CF samples as controls 100 times to ensure consistency in the results. These same 100 sub-samples were used for each of the relevant tests and were matched for geographical location, age, and gender by the following process. Of the 1,617 non-CF samples in the SLL2 dataset, we removed those samples with any other chronic disorder, leaving 1,335 samples. The CF samples came from seven autonomous communities in Spain (Andalucia, Aragón, Cantabria, Catalunya, Madrid, Galicia, and the Basque Country), so from the healthy controls, we first randomly selected two times the proportion of CF samples from each of those locations (i.e. 2x 4/31, 1/31, 11/31, 6/31, 6/31, 2/31, and 1/31, respectively). To ensure a comparable age range, we determined rough age brackets of youth (under 20), adult (20–60), and senior (60 and over), and randomly selected from the geographically matched samples the same proportions of each age group from CF samples (i.e. 10/31 youths, 21/31 adults, 0/31 seniors). Among the CF samples, there were 14 females and 17 males. Thus, a given sub-sampling was finally rejected and reselected if the proportions of males and females were not similar to that of CF samples (i.e. (14±2)/31 females and (17±2)/31 males). From all of the 100 sub-samplings, a total of 352 samples were used as matched controls.

For each of these sub-samples, a number of statistical tests were run with the CF and matched controls together. First, we performed a permutational multivariate analysis of variance (permanova) based on each of the five distance metrics mentioned above using the *adonis* function from the 'vegan' package. The model included the following fixed effects: CF/non-CF, reported use of antibiotics, gender, age, and population of the city/town from which the sample came (as a generalized proxy of both location and lifestyle). There were 21 CF samples that reported antibiotic use and 10 that did not.

Then, to determine differential abundances of taxa and variation in other variables like alpha diversity and pH, we performed a linear model using the function *lm* from the base R package ‘stats’ (version 3.6.3) (R Core Team 2020), again using the same fixed effects as for the permanova test. The abundance values used for these tests were the centered log ratios of the ASV counts, as described above. The *Anova* function from the ‘car’ R package (version 3.0–7) (Fox and Weisberg 2019) was used to calculate type-II anova tables, from which p-values were taken for each fixed effect in the model. These p-values were corrected for multiple testing with the *p.adjust* function from the ‘stats’ package, using the ‘fdr’ method.

4.3.7 Inferred co-occurrence networks

To produce co-occurrence networks, we first filtered out very rare taxa to avoid spurious associations in taxa that do not appear regularly, by using the *filterTaxonMatrix* function from the ‘seqtime’ R package (version 0.1.1) (Faust et al. 2020). We retained those taxa that had at least 15 counts in at least 20 of the 383 samples that included the CF and matched control samples. Then, we calculated the networks for the CF samples and each of the 100 matched control sub-samplings using the *spice.easi* function from the ‘SpiecEasi’ package (version 1.0.7) (Kurtz et al. 2015). To produce the network figures, we used a tutorial from the authors of the ‘SpiecEasi’ package as a guide (Faust 2017). To calculate the Hamming distances between networks, we used the *netdist* function from the ‘nettools’ package (version 1.1.0) (Filosi, Visintainer, and Riccadonna 2017).

4.4 Results

4.4.1 Increased abundances of airway pathogens and decreased abundances of some periodontal

pathogens in CF

Our analyses identified 26 bacterial genera that consistently differed significantly in abundance between CF and controls among the 100 sub-samples (**Table 4.1**, see **Materials and methods** section for explanation of this process) and many of these organisms have implications in the pathogenesis of CF and a number of oral health conditions, which will be examined in the discussion section. The genera found at higher abundance in CF as compared to matched non-CF sets included *Chryseobacterium, Microbacterium, Brevundimonas*, Blvii28 wastewater-sludge group, *Stenotrophomonas, Streptococcus, Rothia, Staphylococcus, Delftia, Comamonas, Scardovia, Desulfobulbus*, genera of the family Clostridiales_vadinBB60_group, *Mobiluncus, Sphingobacterium*, and *Mogibacterium* (**Table 4.1**, **Figure 4.1b**, **Supplementary Figure 4.1**). On the other hand, the genera *Peptostreptococcus*, genera of the Clostridiales family Family_XIII, *Alloprevotella, Treponema, Aggregatibacter, Parvimonas, Bergeyella*, genera of the order Saccharimonadales, and *Fusobacterium* were found at lower abundance in CF, as well as, to a lesser extent, *Haemophilus* (**Table 4.1**, **Figure 4.1b**, **Supplementary Figure 4.1**). At the phylum level, Firmicutes and Actinobacteria, and unclassified sequences, were found at higher abundances in CF, while Spirochaetes, Fusobacteria and Patescibacteria presented lower abundances in CF (**Supplementary Figure 4.2**). There were no taxa from the genus to the phylum level that differed significantly between the non-CF relatives of CF individuals and the same 100 sub-samplings of matched controls. All these differences were not affected by the fixed effects controlled for in the calculations (antibiotic use, gender, age, and population).

Table 4.1: Significance of differentially abundant taxa and other variables between CF and matched controls. Columns indicate, in this order, the taxonomic level or the type of variable considered, the organism name or the variable name, the tendency of the difference in CF (↗: higher in CF, ↘: lower in CF, permanova results are not directional), the mean adjusted p-value of the statistical comparison between CF and matched controls, and the numbers of matched

controls sub-samples for which the test is significant. Rows are ordered first by the tendency in CF samples, with organisms/variables that were greater in CF first, and then by mean adjusted p-value within each variable group.

Taxonomic level/Variable	Organism/ Variable	Tendency in CF	Mean adjusted p value	Significant sub-sample tests
Genus	*Chryseobacterium*	↗	0.00199	100
	Microbacterium	↗	0.00231	100
	Brevundimonas	↗	0.00384	100
	Blvii28 wastewater-sludge group	↗	0.00847	98
	Stenotrophomonas	↗	0.00975	95
	Streptococcus	↗	0.0121	97
	Rothia	↗	0.0129	95
	Staphylococcus	↗	0.0130	92
	Delftia	↗	0.0138	93
	Comamonas	↗	0.0158	91
	Unclassified at Phylum level	↗	0.0173	94
	Scardovia	↗	0.0207	90
	Desulfobulbus	↗	0.0257	87
	Clostridiales vadinBB60_group unclassified genus	↗	0.0311	84

	Mobiluncus	↗	0.0318	83
	Sphingobacterium	↗	0.0325	82
	Mogibacterium	↗	0.0376	82
	Peptostreptococcus	↘	0.00853	97
	Family_XIII unclassified genus	↘	0.0106	97
	Alloprevotella	↘	0.0119	94
	Treponema	↘	0.0221	88
	Aggregatibacter	↘	0.0238	85
	Parvimonas	↘	0.0258	82
	Bergeyella	↘	0.0300	85
	Saccharimonadales unclassified genus	↘	0.0337	83
	Fusobacterium	↘	0.0436	69
	Haemophilus	↘	0.0611	51
Phylum	Firmicutes	↗	0.00575	100
	Actinobacteria	↗	0.00998	97
	Spirochaetes	↘	0.0157	93
	Fusobacteria	↘	0.0259	89
	Patescibacteria	↘	0.0796	49
Fungi	*Candida albicans*	↗	0.0310	82
	Yeast detected (Yes/No)	↗	0.0765	61

Sialochemistry	pH	↘	0.0249	86
Alpha diversity	Shannon diversity index	↘	0.00199	100
	Faith's phylogenetic diversity	↘	0.00001	100
	Species richness	↘	0.00001	100
Permanova (Beta diversity)	Aitchison	-	0.001	100
	Jaccard	-	0.001	100
	Unweighted UniFrac	-	0.001	100
	Bray-Curtis	-	0.00118	100
	Weighted UniFrac	-	0.0258	90

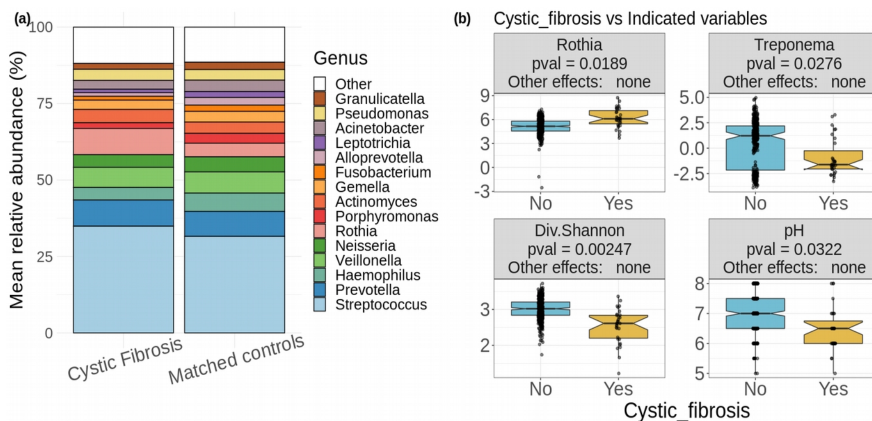

Figure 4.1: Cystic fibrosis differs in factors affecting both oral and lung health. (a) Mean relative abundances of 15 most abundant genera in CF samples

and matched controls. The remaining genera are grouped together and colored in white. (b) Two of the significantly differentially abundant genera are shown (centered log ratio values of *Rothia* and *Treponema*), as well as alpha diversity as calculated by the Shannon diversity index and the salivary pH. In the header of each boxplot, 'Other effects' refers to the significance of the fixed effects included in the calculations (antibiotic use, gender, age, and population). The label 'none' indicates that none of these had a significant effect on average for the given taxon or variable.

4.4.2 Candida albicans *more prevalent in CF*

We used a culture-based approach paired with proteomics to determine the presence of yeast and mold species in our oral rinse samples (**Table 4.2**, see **Materials and methods** section). *Candida albicans* was consistently significantly more prevalent CF samples than matched controls (**Table 4.1**, mean adjusted p among 100 sub-samples = 0.0310, significant in 82 of 100 sub-sample tests). *C. albicans* was detected in 17 of the 31 CF samples, while it was found in 50 of the 352 matched controls that were used in all of the sub-sampling tests (p = 0.0000875 for multinomial log-linear model including all CF and matched controls). We also detected yeast, regardless of species, more frequently in CF samples than matched controls, though less consistently than *C. albicans* in particular. Yeasts were found in 18 of 31 CF samples and in 74 of 352 matched controls (**Table 4.1**, mean adjusted p among 100 sub-samples = 0.0765, significant in 61 sub-sample tests, p = 0.000636 for multinomial log-linear model including all CF and matched control samples). These differences were not significantly affected by the fixed effects included in the calculations (antibiotic use, gender, age, and population).

Table 4.2: Analysis of colonies grown onto YPD + antibiotics plates. The table summarises the number (n) and frequency (%) of samples which formed colonies

for a) yeasts, – with b) indicating the mean and the range of the number of colonies for the yeast positive samples – c) mold, d) bacteria, and e-p) identified fungal species as determined by MALDI-TOF per group: subjects with CF (Cystic Fibrosis) and matched control individuals.

Colony analysis	**CF (n=31)**	**CONTROLS (n=352)**
a) Yeast	18 (58.06)	74 (21.02)
b) # yeast colonies, mean (min-max)	106.44 (1-519)	29.80 (1-578)
c) Mold	7 (22.58)	48 (13.64)
d) Bacteria	19 (61.29)	203 (57.67)
e) *Candida albicans*	17 (54.84)	50 (14.20)
f) *Candida guilliermondii*	2 (6.45)	4 (1.14)
g) *Candida glabrata*	1 (3.23)	2 (0.57)
h) *Candida parapsilosis*	1 (3.23)	5 (1.42)
i) *Debaryomyces hansenii*	1 (3.23)	3 (0.85)
j) *Candida dubliniensis*	0 (0)	4 (1.14)
k) *Rhodotorula mucilaginosa*	0 (0)	2 (0.57)
l) *Candida lusitaniae*	0 (0)	1 (0.28)
m) *Candida tropicalis*	0 (0)	1 (0.28)
n) *Candida zeylanoides*	0 (0)	1 (0.28)
o) *Candida intermedia*	0 (0)	1 (0.28)
p) *Candida spp.,*	0 (0)	2 (0.57)

4.4.3 Lower alpha diversity and salivary pH in CF

CF samples consistently had lower alpha diversity than the matched controls, as determined by the Shannon diversity index, Faith's phylogenetic diversity, which incorporates weights based on phylogenetic distances between taxa, and species richness, which is a count of the unique taxa identified in each sample. The CF samples also consistently had lower pH than matched controls (**Figure 4.1b**). There was no difference in these alpha diversity values or in pH between the non-CF relatives of CF individuals and the same 100 sub-samplings of matched controls. The fixed effects included in the calculations (antibiotic use, gender, age, and population) did not have significant effects in any of these differences. There was also a significant difference in the overall composition of CF samples compared to matched controls based on a permanova test of the five distance metrics described in the methods, each of which measures distances in a different way (**Table 4.1**).

4.4.4 Networks of co-occurring taxa differ significantly between CF and controls

We inferred taxon co-occurrence networks among the 31 CF samples as well as each of the 100 sub-samples to explore underlying differences in the ecology of the oral microbiome in these conditions (**Figure 4.2a-b**). Using a Hamming distance calculation, which measures the degree to which the connections within two networks differ from each other, we found that the networks of the 100 matched control sub-samples were significantly more similar to each other than they were to the CF network (Kruskal–Wallis p-value = 2.2e-16, **Figure 4.2c**). The mean Hamming distance between matched control networks was 0.018 ± 0.005, while the mean distance between the CF network and those of the matched controls was 0.025 ± 0.004.

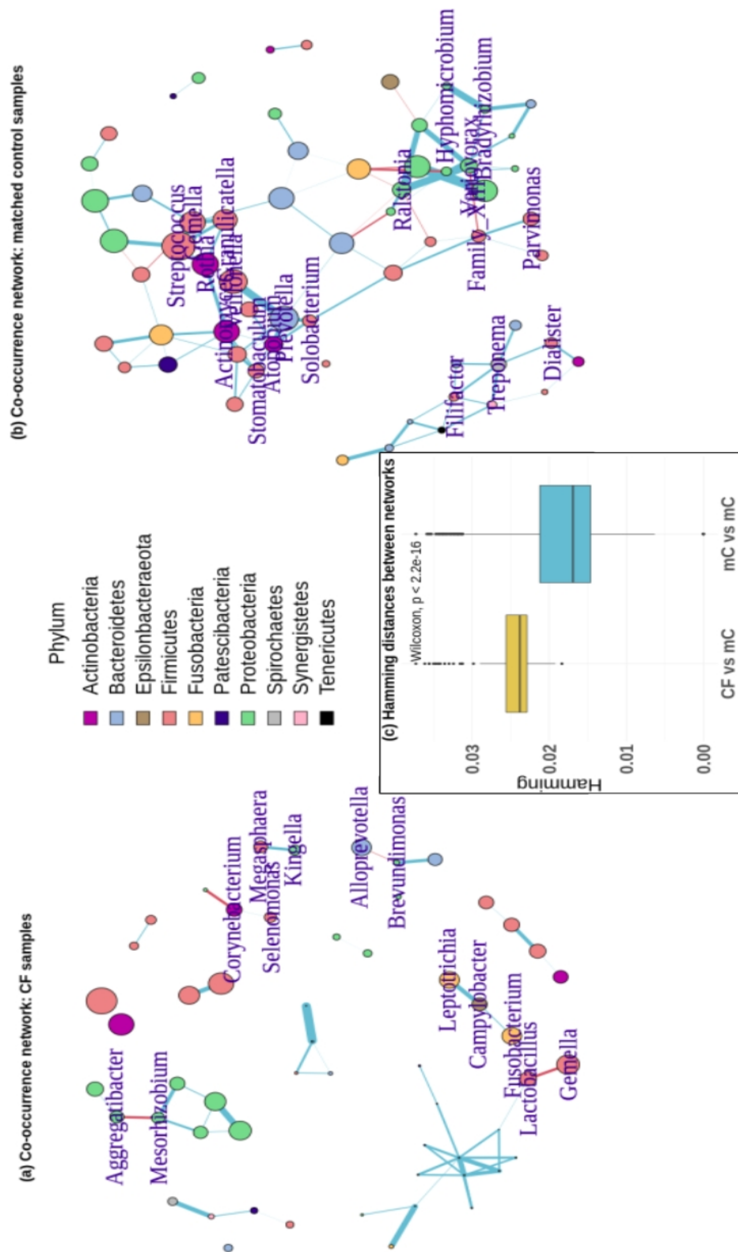

Figure 4.2: Networks of co-occurring taxa differ significantly between cystic fibrosis and matched control samples. Co-occurrence networks of taxa in (a)

the 31 CF samples and (b) all 352 matched control samples. Only those vertices representing taxa mentioned in the text are labeled, as these were the relevant connections that differed between CF and controls. Vertices are colored by the phylum to which they belong. Vertex sizes are proportional to the abundances of those taxa. Edges are colored according to the trend of the association between indicated taxa, where blue edges are positive and red edges are negative. Edge widths are proportional to the strength of the associations. For the matched control network in (b), this comes from a network calculated for all 352 controls together, merely for the sake of visualization. All statistics included in the text are based on the networks from the 100 matched control sub-samples. (c) Distributions of Hamming distances between the CF network and the 100 matched control networks (CF vs mC – yellow) and between each of the matched control networks (mC vs mC – blue). The p-value indicates the significance of the difference between these distributions.

There were a number of specific correlations between taxa that differed between the CF network and those of the matched controls, which have potential implications for the oral cavity as a reservoir of microorganisms for the lower airways, as well as for oral health conditions. Some connections that occurred exclusively in the CF network included negative associations between *Alloprevotella* and *Brevundimonas, Lactobacillus* and both *Fusobacterium* and *Gemella*, and *Aggregatibacter* and *Mesorhizobium*, as well as positive associations between *Campylobacter* and *Leptotrichia, Megasphaera* and *Kingella*, and *Corynebacterium* and *Selenomonas*. These connections are summarized in **Table 4.3**, in the 'CF networks' section. There were also some connections that

did not occur in the CF network, but did occur consistently in the matched control networks. These are also highlighted in **Table 4.3**, in the 'Matched control networks' section, which also shows the number of times a given association was significant among the 100 control networks.

Table 4.3: Associations of co-occurring taxa that differ between CF and matched control networks. The column 'Association' lists the co-occurring taxa. The column 'Direction of correlation' indicates the direction of that association. The final column shows the number of matched control networks (of the 100 sub-samplings that were performed) in which the correlation was significant.

	Association	**Direction of correlation**	**Significant controls**
CF networks	*Alloprevotella* + *Brevundimonas*	Negative	0
	Lactobacillus + *Fusobacterium*	Negative	0
	Lactobacillus + *Gemella*	Negative	0
	Aggregatibacter + *Mesorhizobium*	Negative	0
	Campylobacter + *Leptotrichia*	Positive	0
	Megasphaera + *Kingella*	Positive	0
	Corynebacterium + *Selenomonas*	Positive	0
Matched control	*Ralstonia* + *Variovorax*	Positive	96
	Streptococcus + *Gemella*	Positive	92

networks	*Prevotella + Veillonella*	Positive	92
	Prevotella + Solobacterium	Positive	61
	Prevotella + Atopobium	Positive	50
	Bradyrhizobium + Hyphomicrobium	Positive	86
	Gemella + Granulicatella	Positive	72
	Actinomyces + Stomatobaculum	Positive	62
	Actinomyces + Rothia	Positive	58
	Actinomyces + Atopobium	Positive	46
	Rothia + Granulicatella	Positive	55
	F.Family_XIII.UCG + Parvimonas	Positive	50
	Treponema + Filifactor	Positive	47
	Treponema + Dialister	Positive	33

4.5 Discussion

Our results provide further evidence for the relationship between the microbiota of the oral cavity and the lower respiratory tract and suggest a potential reservoir function of the oral microbiome (Boutin et al. 2015; Gomes-Filho, Passos, and Seixas da Cruz 2010; Whiteson, Bailey, et al. 2014; Dickson et al. 2015). We found a number of organisms that have been associated with lung infections in CF at higher abundance and prevalence in the oral rinse samples of CF individuals as compared to matched controls (**Table 4.1**), including the bacterial genera *Chryseobacterium* (Lambiase et al. 2007; Coenye et al. 2002; Fernández-Olmos et al. 2012),

Microbacterium (Sharma et al. 2012, 2013), *Brevundimonas* (Fernández-Olmos et al. 2012; Menuet et al. 2008; Carter 2010), *Stenotrophomonas* (Coburn et al. 2015; Zemanick et al. 2017; Razvi et al. 2009; Hauser et al. 2011; Talmaciu et al. 2000), *Streptococcus* (Zemanick et al. 2017; Paganin et al. 2015; Maeda et al. 2011; García-Castillo et al. 2007; Whiteson, Meinardi, et al. 2014), *Rothia* (Paganin et al. 2015; Whiteson, Meinardi, et al. 2014; Y. W. Lim et al. 2013), *Staphylococcus* (Zemanick et al. 2017; Razvi et al. 2009; Paganin et al. 2015; Sánchez-Bautista et al. 2019), *Delftia* (Fernández-Olmos et al. 2012; Carter 2010; de Dios Caballero et al. 2016; Tabatabaei, Dastbarsar, and Moslehi 2019; Filkins et al. 2012), *Comamonas* (Coburn et al. 2015; Coenye et al. 2002; Carter 2010; Filkins et al. 2012), *Scardovia* (Carter 2010; Filkins et al. 2012; Soret et al. 2020), *Mobiluncus* (Filkins et al. 2012; Worlitzsch et al. 2009), *Sphingobacterium* (Fernández-Olmos et al. 2012; Zhao et al. 2012), *Mogibacterium* (Filkins et al. 2012; Coffey et al. 2019), and the fungal species *Candida albicans* (Paganin et al. 2015; Bakare et al. 2003; Valenza et al. 2008; Chotirmall et al. 2010; Lepesqueur et al. 2020). In addition, we found significant differences in the overall composition of the oral microbiome based on CF using a permanova test on five different distance metrics, each of which focuses on different aspects of the composition. Taken together, this highlights the strong differences across the entirety of the oral microbiomes of CF and non-CF individuals. Although we did not find a significant difference in the oral cavity between CF and non-CF in the abundance of *Pseudomonas*, the genus of the primary infective agent in CF, *P. aeruginosa*, the differences in these other taxa highlight the dramatic shift in the microbial equilibrium in CF. There is, of course, the possibility that the increased abundances of these organisms results from colonization of bacteria originating in the lower airways, but we feel that it is more likely that the mouth acts as a source of bacteria for the lungs, which become more susceptible to infection under the conditions of CF. For one, all of the organisms mentioned here are detected in both CF and non-CF oral cavities, which would preclude their appearance as a result of CF lung infections. Moreover, the absence of some of the primary infectors in CF in both sample

groups, such as *Burkholderia* or *Achromobacter*, and the equivalent abundances of *Pseudomonas* would support the directionality we have suggested.

In fact, *P. aeruginosa* may be one of the few pathogens prominent in CF lung infections for which the abundance in the oral cavity is not an indicative biomarker, though it interacts with a number of oral species in the lung. Along with some of the other most significant CF pathogens like *Staphylococcus* and *Stenotrophomonas*, *Pseudomonas* is actually a late colonizer of the lungs. These taxa dominate the lung microbiome in CF after the age of 6, while the lungs of CF patients under the age of 2 are primarily composed of oral commensal species from the genera *Streptococcus, Prevotella* and *Veillonella* (Zemanick et al. 2017). When *Streptococcus* colonizes before *Pseudomonas*, the oral species *S. salivarius* is able to inhibit the growth of gram-negative bacteria, such as *P. aeruginosa*, by the production of lactic acid which disrupts their cell membranes (Scoffield and Wu 2015; Whiley et al. 2015). This effect may inhibit the initial spread of some CF pathogens to the lower airways, but it primarily acts within the mouth and throat (Boutin and Dalpke 2017), and so other mechanisms allow growth in the lungs as the patient ages. Even as some Streptococcal species inhibit *Pseudomonas* in the upper airways, others can promote its growth in the lower airways. One model that has been proposed for the physiology of CF lung infections suggests that 2,3-butanediol produced by oral and airway *Streptococcus* is abundant due to increased fermentation under more anaerobic conditions. The 2,3-butanediol can act as a buffer for the decreased pH, but then becomes a carbon source for both *P. aeruginosa* and *Rothia mucilaginosa*, and adds to the virulence of *P. aeruginosa* by allowing it to produce more reactive oxygen species and to provide additional electron acceptors to other anaerobic CF pathogens (Whiteson, Meinardi, et al. 2014).

Rothia mucilaginosa is an oral commensal species and perhaps the most closely tied to *P. aeruginosa* in CF lung infections as it forms similar biofilms (Yuan et al. 2013), may do so alongside *P. aeruginosa* (Y. W. Lim et al. 2013), and can act as a source of

metabolites necessary for the production of glutamate by *P. aeruginosa*, which is used as a component of its cell wall and may increase its virulence (Gao et al. 2018). *R. mucilaginosa* has been associated with a decline in lung function in CF (Paganin et al. 2015) and has been suggested to adapt very efficiently to the CF lung environment, to the point that individual CF patients may have unique strains of the species, as seen with *P. aeruginosa* and *Staphylococcus aureus* (Y. W. Lim et al. 2013). It is able to do so by the use of extracellular lactate, which is present at higher concentrations in CF individuals with poor lung function (Bensel et al. 2011), in order to undergo fermentation under anaerobic conditions, as well as adapting to resist antibiotics, protect against foreign nucleic acids, and to take advantage of the Fe^{2+} produced by the reduction of Fe^{3+} by both *P. aeruginosa* and *S. maltophilia* in the low-pH CF lung (Y. W. Lim et al. 2013).

Desulfobulbus is one of the only taxa found at higher abundance in the oral cavity of CF individuals in this study that, to our knowledge, has not been reported as a CF lung pathogen, though the evidence suggests that it may be well suited to a CF environment. *Desulfobulbus* is a sulfate-reducer (Widdel and Pfennig 1982), which is relevant because *P. aeruginosa* is able to use sulfate from mucins in the airways (Robinson et al. 2012), and the resulting volatile sulfur compounds may be utilized by the fungus *Aspergillus fumigatus*, a common CF pathogen (Scott et al. 2019) In the same vein, it has been shown that there is increased sulfation of mucus glycoproteins in CF (Mohapatra et al. 1995; Y. Zhang et al. 1995).

The frequent treatment of CF with antibiotics may result in the development of antibiotic resistance in a number of pathogens, with evidence that rates of resistance are increasing in some species (Hauser et al. 2011), including *Candida albicans* (Foweraker 2009), also found at greater prevalence in CF samples in our study. *C. albicans* has been implicated in CF lung infections (Bakare et al. 2003; Valenza et al. 2008; Lepesqueur et al. 2020), and in particular, it has been associated with a substantial decline in lung function (Paganin et al. 2015; Chotirmall et al. 2010), as well as

pancreatic insufficiency, exacerbation of lung infection, lower BMI and lower percentage of predicted forced expiratory volume in 1 second (FEV_1) (Chotirmall et al. 2010).

Although in this study we did not have specific information regarding the oral health status of the donors, as they were taken from a larger exploratory study that was not focused explicitly on oral diseases (SLL2 – from 'Saca La Lengua' in Spanish, see http://www.sacalalengua.org, **Chapter 2**, (Willis et al. 2018)), the results presented here broadly support the notion seen throughout the literature that, in CF patients, there is a lower incidence of periodontitis, though they develop an acidic oral environment prone to harm the enamel and dentin of the teeth. Factors that may indicate lower incidence of periodontitis in CF include low alpha diversity (LaMonte et al. 2018; Abusleme et al. 2013; Genco et al. 2019; Takeshita et al. 2016), as seen in our data, malfunction of the CFTR protein (Ajonuma et al. 2010), and generally higher use of anti-inflammatory and antibiotic medications (Pawlaczyk-Kamieńska et al. 2019). Indeed, nearly all of the genera found at lower abundance in CF samples here have been associated with periodontal disease. Many of these are members of the various bacterial complexes which form biofilms at different stages of the disease (Mohanty et al. 2019; A. P. V. Colombo et al. 2009; Koyanagi et al. 2013; Riggio and Lennon 2002; S. Bizzarro et al. 2016; Ramanan et al. 2014; Ruhl et al. 2014). Some of the organisms found at higher abundance in CF samples in this study have been implicated in dental caries and carious lesions on the teeth (J. Zhou et al. 2016; H. Chen et al. 2013; Q. Jiang et al. 2018; Mantzourani et al. 2009; Tanner et al. 2011; W. Jiang et al. 2014; Henne et al. 2015; Richards et al. 2017; Y. Wang et al. 2017; Naidu and Reginald 2016; Lozano Moraga et al. 2017; J. Xiao et al. 2018; Eidt et al. 2019) and, conversely, others found at lower abundance in CF samples have been linked to the absence of caries and tooth damage (W. Jiang et al. 2014; L. Xu et al. 2018; H. Xu et al. 2014; Lif Holgerson et al. 2015; Schoilew et al. 2019; Hernández et al. 2020).

Alloprevotella presents an interesting case that may merit further

study on its own regarding CF. Species in this genus reduce nitrate in the saliva which contributes to the anti-inflammatory response to periodontitis (Henskens et al. 1996; Qu et al. 2016; Espinoza et al. 2018), so its lower abundance in CF here could be linked in part to lower incidence of periodontitis. However, the nitrate reduction in CF patients seems to be a more complicated process. When it is metabolized to nitric oxide (NO), it has an anti-inflammatory effect in the airways of CF individuals. Some studies have shown that precursors to NO, like nitrate and nitrite, are higher in the saliva and exhaled breath condensate (EBC) of CF patients than in those of controls, but nonetheless, the amount of exhaled NO was lower in CF (Grasemann et al. 1998; Zetterquist et al. 2009). One proposed explanation is that NO may be produced normally in CF, but its diffusion is inhibited in the thick mucus produced in CF airways (de Winter-de Groot and van der Ent 2005), though these first two studies suggest that there is an impairment in the formation of NO in CF patients. Another study found that increasing the intake of dietary nitrate led to an increase in exhaled NO as compared to placebo treatments (Kerley et al. 2016). From this information, we cannot extrapolate to determine the exact mechanism in the impairment of the NO cycle in CF, but the low abundance of the nitrate-reducing genus *Alloprevotella* forms a link to this process that may warrant a deeper investigation.

Significant differences in many co-occurrences of taxa among the CF samples and the 100 matched control groups suggest underlying ecological differences in these two conditions. The Hamming distance calculations showed that the 100 matched control networks were more similar to each other than to the CF network (**Figure 4.2c**), and we also highlighted a number of particular co-occurrences that were specific to either the CF or control networks, and these may have implications for the connections between the mouth and lung in CF, as well as for oral health conditions. In the CF network (**Figure 4.2a**, **Table 4.3**), the negative association between the genera *Alloprevotella* and *Brevundimonas* follows with the results from our study and the literature, wherein *Alloprevotella* was lower and *Brevundimonas* was higher in CF (Fernández-Olmos et al. 2012; Menuet et al. 2008; Carter 2010). It is unclear whether

this connection is related to the NO cycle in CF, but at least some strains of *Brevundimonas* species are nitrate reducers (Nawaz, Hasan, and Shah 2015), so it may be that *Brevundimonas* is able to outcompete *Alloprevotella* in the CF environment. There may be a more concrete explanation for the negative associations between *Lactobacillus* and both *Fusobacterium* and *Gemella*, which are periodontal pathogens (Mohanty et al. 2019; Al-Jebouri 2016; A. P. V. Colombo et al. 2012). Many studies explore the use of *Lactobacillus* species as probiotic treatments and have shown that they can inhibit periodontal pathogens (Kõll Klais, Mändar, and Leibur 2005) by co-aggregating with them, promoting immune responses and interrupting biofilm formations (Chatterjee and Bhattacharya 2011). *Campylobacter* and *Leptotrichia* also have a meaningful connection in the CF oral microbiome, as they can be found in association in dental caries (Eribe and Olsen 2008), where they can metabolize sugars to lactic acid, aiding in the process of cariogenesis (Peterson et al. 2013; Thompson and Pikis 2012), and we have already mentioned that extracellular lactate levels are higher in CF sputum (Bensel et al. 2011). *Megasphaera* is sometimes associated with CF (Losada et al. 2016; Nielsen et al. 2016) and has been considered a member of the core microbiome of the lower lung, regardless of CF status (Zakharkina et al. 2013). This genus can work to neutralize acidic conditions that damage teeth (Nallabelli et al. 2016) and *Kingella* is acidogenic (F. E. Dewhirst et al. 1993), so their co-occurrence in the CF network may indicate that they complement each other in the acidic CF oral cavity. The association between *Corynebacterium* and *Selenomonas* in CF is reasonable, as both have been implicated in CF (Bittar, Cassagne, and Bosdure 2010; Pivot et al. 2019; Layeghifard et al. 2019).

Among the matched control samples, the co-occurrence networks primarily support our speculations on the underlying mechanisms that might lead to lower incidence of periodontitis and greater damage to enamel in CF patients as compared to non-CF controls. *Prevotella* and *Veillonella* are among the four most abundant genera in our dataset and were significantly associated in 92 of the matched control networks. *Prevotella*, *Veillonella*, and *Solobacterium* are all

periodontal pathogens (A. P. V. Colombo et al. 2009; Carrouel et al. 2016; Mashima et al. 2015), *and Prevotella melaninogenica* and *Solobacterium moorei*, whose genera were significantly associated in 61 matched control networks, utilize cysteine to produce hydrogen sulfide (H_2S), resulting in halitosis (Haraszthy et al. 2007; Stephen et al. 2014). Cysteine is a precursor to glutathione (Stipanuk et al. 2006), a peptide that helps to protect the lung from oxidants, and is found at lower levels in the lungs of CF patients (McKone et al. 2006), and acetylcysteine, a prodrug to cysteine, has been used to improve lung function in CF patients (Conrad et al. 2015). So perhaps there is a connection between greater overall availability of cysteine in non-CF individuals and the production of H_2S by organisms like *P. melaninogenica* and *S. moorei. Veillonella* species also produce volatile sulfur compounds that cause halitosis (Mashima et al. 2015). *Streptococcus* and *Gemella* are among the most abundant taxa present in the oral microbiome (**Figure 4.1a**) so it is not strange that they would have a significant association in 92 of the 100 control networks, though it is curious that they did not in the CF network. As mentioned above, *Streptococcus* was significantly more abundant in our CF samples and has been associated with CF lung infections (Zemanick et al. 2017; Paganin et al. 2015; Maeda et al. 2011; García-Castillo et al. 2007; Whiteson, Meinardi, et al. 2014), and *Gemella* may play a role in exacerbations of CF infections (Carmody et al. 2013). Nonetheless, they both have some involvement in periodontitis as well (Al-Jebouri 2016; A. P. V. Colombo et al. 2012; Dani et al. 2016; Contardo et al. 2011), and species of *Streptococcus* in particular make up the ‘yellow-complex’ of periodontal pathogens, which are early colonizers in that disease (Carrouel et al. 2016; Sigmund S. Socransky and Haffajee 2002). Somewhat unexpected associations also occurred exclusively in the matched control networks between *Gemella* and *Granulicatella* (occurs in 72 of the 100 matched control networks), and between *Rothia* and *Granulicatella* (55 of the 100 control networks), as *Granulicatella* has also been associated with CF (Sánchez-Bautista et al. 2019; Vandeplassche et al. 2019; Bevivino et al. 2019) and caries due to its acidogenic nature (Jagathrakshakan et al. 2015). The significant

associations seen exclusively in our matched control networks involving Ralstonia and Variovorax (Basavaraju et al. 2016; Utari, Vogel, and Quax 2017; Muras, Otero-Casal, et al. 2020; Muras, Mayer, et al. 2020), *Atopobium* (Camelo-Castillo et al. 2015; Ai et al. 2017), *Bradyrhizobium* (Shin et al. 2013), *Hyphomicrobium* (Anesti et al. 2005), *Actinomyces* (Vielkind et al. 2015), *Stomatobaculum* (W.-P. Chen et al. 2018), *Rothia* (Ramanan et al. 2014; Camelo-Castillo et al. 2015; Ai et al. 2017), Clostridiales Family_XIII (S. Bizzarro et al. 2016), *Parvimonas* (Mohanty et al. 2019; S. Bizzarro et al. 2016; Tindall and Euzéby 2006), *Treponema* (Mohanty et al. 2019; S. Bizzarro et al. 2016), *Filifactor* (S. Bizzarro et al. 2016; Deng et al. 2017; Naginyte et al. 2019), and *Dialister* (Ghayoumi, Chen, and Slots 2002; Oswal et al. 2020), all relate to organisms implicated in periodontitis.

4.6 Conclusions

In this study, we found significant differences in the oral microbiomes of CF and non-CF individuals from all around Spain, which have implications for the potential of the oral cavity to act as a reservoir of microorganisms for the lower airways, as well as for oral health in CF. Differential abundances of bacteria and fungi follow trends in the literature regarding CF lung infections, presenting similar microbial activity in the oral cavity. We highlighted underlying physiological differences that are apparent in the co-occurrences of taxa among CF samples and matched control samples, which add greater evidence to those trends discussed here. These results provide a snapshot of the unique composition of the CF oral microbiome and how it relates to oral health and the lower airways.

Supplementary information

Supplementary information referenced in this chapter accompanies the publication at https://doi.org/10.1080/20002297.2021.1897328.

Chapter 5: Citizen-science study of the oral microbiome reveals changes through age and the relative impacts of diverse biological, social and lifestyle factors.

Willis JR, Saus E, Iraola-Guzmán S, Ksiezopolska E, Cozzuto L, Bejarano LA, Andreu-Somavilla N, Alloza-Trabado M, Blanco A, Puig-Sola A, Broglio E, Carolis C, Ponomarenko J, Hecht J, & Gabaldón T. Citizen-science study of the oral microbiome reveals changes through age and the relative impacts of diverse biological, social and lifestyle factors. *(In preparation)*

Chapter 5: Citizen-science study of the oral microbiome reveals changes through age and the relative impacts of diverse biological, social and lifestyle factors.

5.1 Abstract

Background: The relevance of the human oral microbiome to our understanding of human health has grown in recent years as microbiome studies continue to develop. Given the links of the oral cavity with the digestive, respiratory and circulatory systems, the composition of the oral microbiome is relevant beyond just oral health, impacting systemic processes across the body. However, we still have a very limited understanding about intrinsic and extrinsic factors that shape the composition of the healthy oral microbiome. Here we followed a citizen-science approach to assess the relative impact on the oral microbiome of selected biological, social and lifestyle factors in 1,648 Spanish individuals. **Results:** We found that the oral microbiome changes across age, with middle ages showing a more homogeneous composition, and older ages showing more diverse microbiomes with increased representation of typically low abundance taxa. By measuring differences within and between groups of individuals sharing a given parameter, we were able to assess the relative impact of different factors in driving specific microbial compositions. Chronic health disorders present in the analyzed population were the most impactful factors, followed by smoking and the presence of yeasts in the oral cavity. Finally, we corroborate findings in the literature that relatives tend to have more similar oral microbiomes, and show for the first time a similar effect for classmates. **Conclusions:** Multiple intrinsic and extrinsic factors jointly shape the oral microbiome. Comparative analysis of metabarcoding data from a large sample set allows us to disentangle the individual effects.

Keywords: Oral microbiome, age, lifestyle, shared environment

5.2 Background

The oral cavity is inhabited by an abundant and diverse microbial community, the oral microbiome, which has been related to processes relevant for health and disease (Willis and Gabaldón 2020). The mouth is highly vascularized (Naumova et al. 2013), and is an entry point to the respiratory and digestive systems. As a result, changes in the composition of the oral microbiome can reflect and/or influence systemic changes across the human body, and as such it has an important diagnostic and therapeutic potential. A multitude of factors, both intrinsic (e.g. pH, immune system, chronic disorders) and extrinsic (e.g. lifestyle, diet), have the potential to shape the oral microbiome, but these are as yet only poorly understood. Increasing our knowledge on how these factors alter the oral microbiome is important for unveiling the specific roles that certain oral microbes play in disease processes, which in turn may pave the way for the development of innovative microbiome-based diagnostic and therapeutic approaches.

Most studies on the oral microbiome have focused on delineating its changes in the context of common oral diseases such as periodontitis, gingivitis, or dental caries (Costalonga and Herzberg 2014; Belibasakis et al. 2019). In recent years, however, the relationships of the oral microbiome with systemic diseases or chronic disorders have received growing attention. These include, among others, different cancer types (Fan, Alekseyenko, et al. 2018; Flemer et al. 2018), cardiovascular diseases (Chhibber-Goel et al. 2016; Teles and Wang 2011), diabetes (Casarin et al. 2013), celiac disease (Valitutti, Cucchiara, and Fasano 2019; Fernandez-Feo et al. 2013; Tian et al. 2017), Down Syndrome (Willis et al. 2020) or cystic fibrosis (Willis et al. 2021). Thanks to these studies, we are beginning to understand how oral or systemic disorders relate to changes in the composition of the oral microbiome. However, given

the strong focus on disease, we still lack a sufficient understanding of non-disease parameters that shape the healthy oral microbiome. These intrinsic (host biology) or extrinsic (environment, lifestyle) factors are pervasive and likely influence not only the overall composition of the oral microbial ecosystem, but also how it will respond in the context of disease, perhaps predisposing one to either relative dysbiosis or resilience.

A relevant intrinsic factor that has been poorly studied in relation to the oral microbiome is age. To our knowledge, there are no studies using high throughput sequencing techniques which focus specifically on the effects of aging on the oral microbiome in a state of relative health and which include a representative spectrum of ages. Recent reviews that have explored aging largely highlight the tendency toward increased periodontitis and dental caries, but they rely primarily on studies using culture-based identification techniques in regards to alterations in particular taxa (Feres et al. 2016; Belibasakis 2018; G. Hajishengallis 2014). Some studies which have compared age groups have some limitations, such as narrow age ranges or a focus on age only in the context of particular diseases (Feres et al. 2016; Lenartova et al. 2021; LaMonte et al. 2019; Burcham et al. 2020). Nonetheless, there are conjectures throughout the literature in reference to the oral microbiome's role in, and impact from, the physiological changes that occur during the human aging process. Perhaps most notable is the chronic low-grade systemic inflammation sometimes called "inflammaging", which coincides with immunosenescence, wherein the adaptive immune system declines and the efficiency of innate immunity diminishes with age (Franceschi and Campisi 2014; G. Hajishengallis 2014). Thus, further investigation into the connections between age and the oral microbiome is warranted.

Lifestyle and hygiene are perhaps the most studied extrinsic factors

with respect to changes in the oral microbiome (M. Kilian et al. 2016; Mogens Kilian 2018). Smoking (J. Wu et al. 2016; Kato et al. 2016; Mason et al. 2015; Vallès et al. 2018), wearing braces (Willis et al. 2018; Lucchese et al. 2018; Sun et al. 2018; Jing et al. 2019), and the composition of drinking water (Willis et al. 2018; McDonough et al. 2020) are factors that have been shown to drive particular changes in the oral microbiota. Extrinsic variables like these impact the oral microbial composition, and in fact, multiple studies have demonstrated that lifestyle, social structures, and shared environments are generally more significant than intrinsic factors like the human hosts' genetics. Family members have been shown to display more similar microbiome compositions to each other than to non-family members, while there was not a greater similarity amongst monozygotic twins than amongst dizygotic twins (Stahringer et al. 2012; Shaw et al. 2017; Song et al. 2013; Burcham et al. 2020; Yatsunenko et al. 2012).

Bacteria have received most of the attention in microbiome studies, but other organisms like fungi are also important components. In the oral cavity, species like *Candida albicans* have been implicated in dental caries (Diaz and Dongari-Bagtzoglou 2021), wherein it can adhere to the biofilms of the bacterial species *Streptococcus mutans* and both can act to demineralize tooth enamel (J. Xiao et al. 2018; Eidt et al. 2019). One study showed two distinct mycotypes (clusters of samples based on the fungal composition), with one being dominated by *Candida* species, and the other with higher fungal diversity and *Malassezia* as the main genus (Hong et al. 2020). This and another study (Janus et al. 2017) distinguished associations with bacterial taxa in *Candida*-dominated versus other samples, though those results do not seem to coincide entirely. The interactions between bacteria and fungi are an interesting aspect of the oral microbiome that deserves greater attention.

Here, we have taken advantage of the second edition of a large-scale citizen science-based project called “Saca La Lengua” (SLL2 - “Stick Out Your Tongue” in English) (“Saca La Lengua - Stick out Your Tongue” n.d.; Willis et al. 2018) to explore the effects of some of these factors in the oral microbiome. Contrary to disease-focused studies, studies on the overall population enabled by citizen-science provide a unique opportunity to infer the effects of commonly present factors. The dataset comprises 1,648 oral rinse samples taken from locations across Spain, representing a broad and balanced range of ages. A subset of the samples were from individuals with chronic disorders that are relevant to the physiology of the oral cavity, and all participants filled out a comprehensive survey with questions about lifestyle, diet, and hygiene habits. We coupled this information with 16S rRNA metabarcoding, as well as culture and proteomics-based identification of fungi to study some of the influences on and of the oral microbiome.

5.3 Materials and Methods

5.3.1 Sample collection

All participants signed an informed consent form allowing the use of their saliva samples for microbiological research. For participants under the age of 18, the consent form was also signed by one of the parents or a legal guardian. This project was approved by the ethics committee of the Barcelona Biomedical Research Park (PRBB). Samples were collected from January to November 2017. Participants were asked not to ingest any food or beverage (except water) for 1 hour before collecting the sample. All donors received clear indications about the sample collection procedure in person, and the collection of the samples was carried out with the assistance of a researcher involved in the project, following a demonstration. All participants responded to a uniform questionnaire about

lifestyle, diet, hygiene, and health. Before collection of the oral rinse, the pH of the saliva was measured using pH test strips (MColorpHast, Merck, range 5.0–10.0; 0.5 accuracy units), the accuracy of which have been previously validated (Willis et al. 2018). Saliva samples were collected using a mouthwash as described earlier (Willis et al. 2018). In brief, the protocol is as follows: participants rinsed their mouth with 15 mL of sterile phosphate-buffered saline (PBS) solution, for 1 min. Then, they returned the liquid into a 50 mL tube. The samples were then centrifuged at 4,500 g for 12 min at room temperature (r.t.) in an Eppendorf 5430 centrifuge equipped with an Eppendorf F-35-6-30 rotor. The supernatant was discarded and the pellets were resuspended with the remaining PBS, transferred to 1.5 ml tubes and centrifuged at 4,500 g for an additional 5 min at r.t. using an Eppendorf FA-45-24-11-HS rotor. Supernatants were discarded, and pellets were frozen and stored at −80 °C until further analysis.

The methods used for DNA extraction and 16S amplicon sequencing, fungal composition analysis, the pre-processing of 16S rRNA sequence reads and taxonomy assignment, as well as the alpha and beta diversity measures that we employed, were described in previous publications which used the same dataset (Willis et al. 2020, 2021).

5.3.2 Subsampling for analyses

When running statistical tests for a given variable, we first randomly select representative matched controls 100 times to ensure consistency in the results. In the case of binary variables, such as smoking, where the values are either “yes” or “no”, we randomly selected an equal number of samples from each group, and checked if each group had similar distributions of age, geographic location (based on the autonomous community within Spain from which the sample was collected), and gender. If there were over 100 total samples in both the “yes” and “no” groups for a given variable, 100 of each were selected for each of the 100 subsamplings, otherwise

the total number of the smaller group were selected and a matched random selection of the same size from the other group. In the case of age groups, we first classified our samples into six bins of ages: 13-20 (964 samples), 20-30 (41 samples), 30-40 (28 samples), 40-50 (85 samples), 50-60 (46 samples), >60 (42 samples). We then ensured that the six age bins had balanced geographical distributions and genders. The 100 subsamples based on the age bins were also used for calculations with age as a continuous variable, in order to ensure an even distribution of ages, as well as to account for the geographical distributions and genders. In the case of the ionic composition of drinking water, each participant indicated their primary source of drinking water, the options of which were bottled water, filtered tap water, unfiltered tap water, and untreated water (from a fountain, well, or river). For these analyses, those samples from individuals which drank primarily bottled water were removed, and 100 subsamples were selected from the remaining samples, balancing the distributions of age, geographical location, and gender. Relevant p-values mentioned throughout the text are the average from the tests across the 100 subsamplings, corrected with the "fdr" method in the p.adjust function from the base "stats" package (version 3.6.3) (R Core Team 2020), unless otherwise stated.

5.3.3 Statistical analyses

5.3.3.1 Comparisons of compositions

In order to determine the effects of variables on the composition of the oral microbiome, we performed a permutational multivariate analysis of variance (PERMANOVA) based on Aitchison distance metric using the *adonis* function from the "vegan" R package (version 2.5-6) (Oksanen et al. 2017). The model included the following fixed effects: variable of interest (e.g. smoking or age), gender, age (when it is not the primary variable of interest), and population of the city/town from which the sample came (as a generalized proxy of both location and lifestyle).

We used the *betadisper* function from the “vegan” package to test the homogeneity of group variances within the groups for a given variable to compare, for example, smokers versus non-smokers, or among the six age bins. Using the *anova* function from the “stats” package on the betadisper object, we could obtain a p-value to determine whether there was a significant difference in the homogeneity of the compositions of samples between groups. The betadisper objects also hold the distance of each sample within a group from that group’s spatial median, a measure of the centroid composition for each group, and we use these values to display the differences between groups in boxplots.

We used the *anosim* function from the “vegan” package to perform an analysis of similarities based on the Aitchison distance metric among family members and classmates. The different relationships considered were siblings, twins, partners, parents-children, grandparents-grandchildren, all family members (includes any of the previously mentioned relationships), and classmates (samples from students in the same school). For an anosim test of a given relationship, only those samples which had at least one relationship of that type were included. For example, there were 70 samples which had a sibling that also provided a sample, so the sibling anosim test included those 70 samples, wherein 34 distinct groups of siblings occurred (for any relationship, groups were by necessity of two samples or more).

5.3.3.2 Differential abundance and diversity calculations

Then, to determine differential abundances of taxa and variation in other variables like alpha diversity, oral pH, or the measurements of ions in drinking water, we performed a generalized linear model using the function *glm* from the “stats” package, again using the same fixed effects as for the PERMANOVA test. The abundance values used for these tests were the centered log ratios of the

amplicon sequence variant (ASV) counts, as described elsewhere (Willis et al. 2020, 2021). For the alpha diversity measures in relation to age in particular, we further used the *bs* function from the base R package “splines” to treat age as a second order fixed effect, in order to detect a parabolic trend. The *Anova* function from the “car” R package (version 3.0-7) (Fox and Weisberg 2019) was used to calculate type-II ANOVA tables, from which p-values were taken for each fixed effect in the models. These p-values were corrected for multiple testing with the *p.adjust* function from the “stats” package, using the “fdr” method.

5.3.3.3 Inferred co-occurrence networks

To produce co-occurrence networks within the groups for a given variable, we first filtered out very rare taxa to avoid spurious associations in taxa that do not appear regularly, by using the *filterTaxonMatrix* function from the “seqtime” R package (version 0.1.1) (Faust et al. 2020). We retained those taxa that had at least 15 counts in at least 20 samples. Then we calculated the networks for each of the groups in a given variable (for instance, for smokers and for non-smokers) in each of the 100 subsamplings using the *spiec.easi* function from the “SpiecEasi” package (version 1.0.7) (Kurtz et al. 2015). The chord diagrams that we used to represent the uniqueness of networks for particular variables were produced using the chordDiagram function from the “circlize” package (version 0.4.8) (Gu et al. 2014). To calculate the relative uniqueness of networks, we developed a score that is relative to the eight variables that we considered, which were those found to be significant with the PERMANOVA test. The scores were calculated as follows: for each variable, the co-occurrence networks were calculated among each of the 100 subsamples, and we retained those associations which occurred only in the groups of interest (samples with the indicated disorder, smokers, antibiotic users, or those samples in which yeast was absent). Then for each variable,

we calculated, pairwise with each other variable, the number of only those associations which occurred in all 100 subsamples and in 0 subsamples of the other variable being compared, weighted by the strengths of those associations that were determined by the *spiec.easi* function.

5.4 Results

5.4.1 Oral microbiome changes through age.

To assess the impact of aging on the oral microbiome, we compared the microbial profiles of oral rinse samples across ages, using a subsampling strategy that ensures comparable sample sizes (see Materials and Methods). We first tested for changes in the overall microbiome composition across age, including gender and population as fixed effects in 100 such subsamples (see Materials and Methods). PERMANOVA tests based on an Aitchison distance matrix considering age as a continuous value were consistently significant (mean adjusted P = 0.001, mean R^2 = 0.023, mean F statistic = 4.37). To further explore the age ranges that were most distinct, we also used age as a categorical variable, subsampling from the following age bins: 13-20 years old (964 samples), 20-30 (41 samples), 30-40 (28 samples), 40-50 (85 samples), 50-60 (46 samples), >60 (42 samples), which also showed significant differences (mean adjusted P = 0.001, mean R^2 = 0.051, mean F statistic = 1.90). Interestingly, when comparing each age bin separately against the group of all others, we observed a sort of parabolic effect, where only those comparisons of the extreme bins (13-20 and >60) against all other bins had a significant result on average across the 100 subsamples, and the differences involving the intermediate bins (30-40 and 40-50) were not significant (**Figure 5.1A**). We further calculated the homogeneity in the microbial composition of samples within a given bin. This homogeneity test first calculates a spatial median for each age bin (a sort of hypothetical centroid composition of the samples within a given age bin, derived from an Aitchison distance matrix), then

calculates the distance of each sample in that bin to the spatial median. This test resulted in a similar parabolic effect to that of the PERMANOVA tests (mean adjusted ANOVA P = 0.0016, mean F = 4.85), wherein the 40-50 bin was the most homogeneous in terms of microbiome composition, and the >60 bin was the most variable (**Figure 5.1B**).

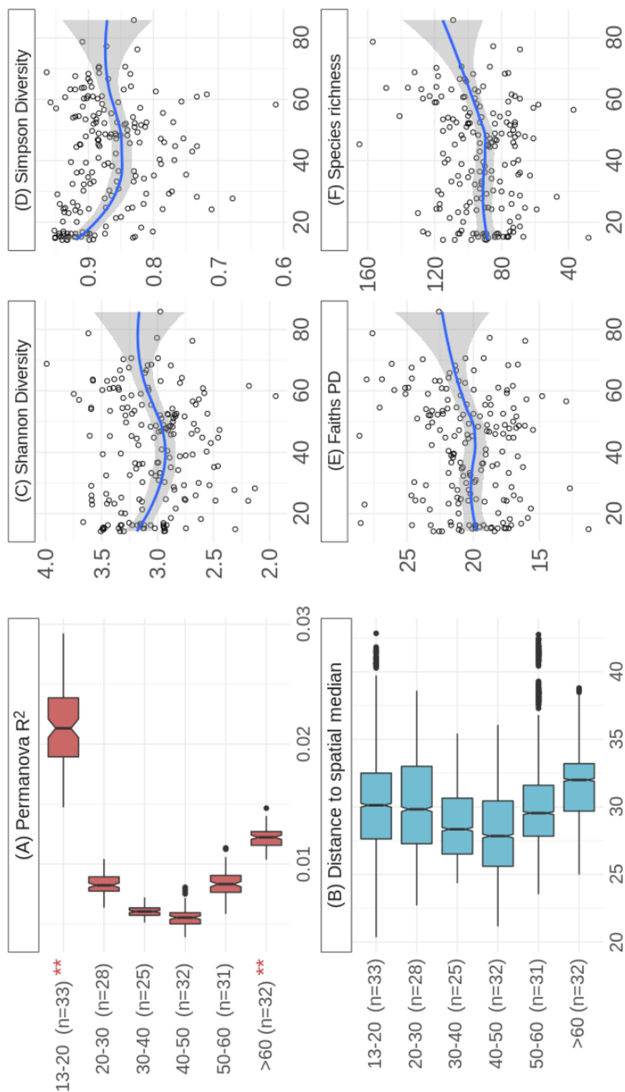

Figure 5.1: Homogeneity, distinction of composition, and alpha diversity across age. **(A)** Boxes of the R^2 values are from the PERMANOVA tests run separately for each of the 100 subsamples. The n in both plots indicates the number of samples in a given age bin in each subsample. Red stars indicate the magnitude of the mean adjusted p-values for the PERMANOVA tests. The representation of p-values are represented with symbols as indicated in the following value intervals: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' Not significant. **(B)** Boxes for the distances to the spatial median represent those distances of each sample from the spatial median of its particular age bin, as calculated by the *betadisper* function. The spatial medians for age bins and the associated ANOVAs were run separately for each of the 100 subsamples, but the boxes here display all such distances for each age bin. **(C-F)** Tests of the four alpha diversity measures (Shannon, Simpson, Faith's PD, species richness) were also run with those same 100 subsamples, using age as a continuous value, and the statistical values are summarized in **Table 5.1**. The four respective scatter plots here display only the values from one of those subsamples to give a representative depiction of the trend (the same subsample is used for all four), with age (in years) along the x-axis.

In addition, some alpha diversity measures showed parabolic relationships with age, wherein Shannon and Simpson diversity values were lower in the middle ages, consistent with the above result that these were the most homogeneous samples, while Faith's phylogenetic diversity (PD) and species richness each increased with age, especially in older individuals, though with less statistical significance than the Shannon and Simpson diversities (**Figure 5.1C-F, Table 5.1**). In **Table 5.1**, the p-values for both quadratic and linear models for these alpha diversity values are displayed, showing that indeed the quadratic model better explains the trends across age.

Table 5.1: Significance of differentially abundant taxa and alpha diversity measures as age increases. Columns indicate, in this order, the taxonomic level or type of variable, the organism name or variable name, the tendency of the change across age ("↗": increases with age, "↘": decreases with age, "↘ -

↗": parabolic effect seen in age, "→ - ↗": steady across most ages with an increase particularly in older samples), the mean adjusted p-value from the generalized linear or quadratic model, and the number of subsamples for which the test is significant. Rows are ordered first by the tendency with age, with organisms/variables that increase first, and then by mean adjusted p-value. In the last two columns for the alpha diversity measures, values are displayed for models based on both quadratic functions of age (Q) and linear functions of age (L).

Taxonomic level/Variable	**Organism/Value**	**Tendency across age**	**Mean adjusted P**	**Significant subsample tests**
Genus	*Anaeroglobus*	↗	0.0004	100
	Eikenella	↗	0.0033	100
	Fretibacterium	↗	0.0013	99
	Comamonas	↗	0.02	92
	Olsenella	↗	0.028	87
	Phocaeicola	↗	0.037	75
	Alloprevotella	↘	0.0003	100
	Streptobacillus	↘	0.0026	100
	Haemophilus	↘	0.0072	98
	Prevotella	↘	0.016	93
	Granulicatella	↘	0.02	93
	Bergeyella	↘	0.035	83
Phylum	Synergistetes	↗	0.0002	100
	Bacteroidetes	↘	< 0.0001	100
	Proteobacteria	↘	0.031	80

Physiology	BMI	↗	< 0.0001	100
	pH	↘	0.0026	100
Alpha Diversity	Simpson's diversity	↘ - ↗	Q = 0.0031 L = 0.26	Q = 100 L = 0
	Shannon's diversity	↘ - ↗	Q = 0.021 L = 0.99	Q = 90 L = 0
	Species Richness	→ - ↗	Q = 0.04 L = 0.071	Q = 83 L = 52
	Faith's PD	→ - ↗	Q = 0.076 L = 0.12	Q = 0 L = 9

We next investigated which organisms show significant differences across age. Our results (**Table 5.1**) show a number of taxa that increase with age, including the genera *Anaeroglobus*, *Eikenella*, *Fretibacterium*, *Comamonas*, *Olsenella*, and *Phocaeicola*, as well as the phylum Synergistetes, or decrease with age, including the genera *Alloprevotella*, *Streptobacillus*, *Haemophilus*, *Prevotella*, *Granulicatella*, and *Bergeyella*, as well as the phyla Bacteroidetes and Proteobacteria. Of note, genera that increase with age are typically found at low abundance among all samples, whereas those that decrease with age tend to display the opposite trend (**Supplementary Figure S5.1**). There was also a marked decrease in pH and increase in BMI as age increased (**Table 5.1**).

5.4.2 Chronic disorders, smoking and the presence of yeasts in the oral cavity, are important drivers of the oral microbiome composition.

We collected a comprehensive questionnaire regarding over 80 aspects of lifestyle, diet, hygiene, and health from all of the 1,648 participants in this study. To assess which of the considered variables had the largest effects on the overall composition of the oral microbiome, we used a PERMANOVA test for each variable with an Aitchison distance matrix, including age, gender, and population as fixed effects (see Materials and Methods). For each of the tested variables, 100 subsamples were taken to match the groups in that variable (Yes vs No) by geographic location, age, and gender. In these comparisons, we excluded samples from donors with any reported chronic disorders, except when the variable of interest was such a disorder. Our results (**Figure 5.2A**) show that chronic disorders like cystic fibrosis (mean adjusted P = 0.0011, mean R^2 = 0.054, mean F statistic = 3.39) and Down Syndrome (mean adjusted P = 0.0013, mean R^2 = 0.059, mean F statistic = 3.33) were the variables that were most distinct between groups. The detection of yeast species in a sample (as well as the detection of *Candida* specifically), smoking, celiac disease, hypertension, and the reported use of antibiotics also had significant (mean adjusted P < 0.05) PERMANOVA results.

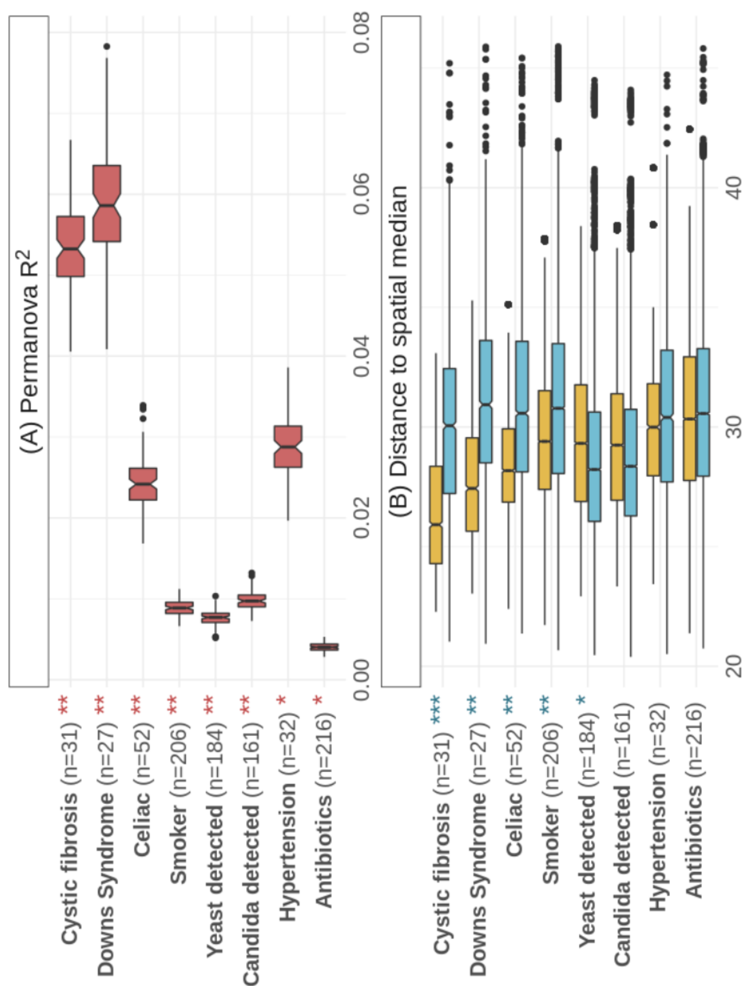

Figure 5.2: Homogeneity and distinction of composition across variables. (A) Boxes show the distribution of R^2 values (the proportion of sum of squares from the total) from the PERMANOVA tests comparing groups of a given variable for the 100 subsamples. **(B)** Boxes represent the distances of each sample from the spatial median of its group (Yes in yellow, No in blue), as calculated by the *betadisper* function. The spatial medians for groups and the associated ANOVAs were run separately for each of the 100 subsamples, but the boxes here display all such distances for each group. Pairs of boxes in both plots are ordered by the absolute value of the difference between the pairs. The n in both plots indicates the number of samples for which a given variable was indicated (the same number of matched controls were selected for each subsample test). Red stars in **(A)** and blue stars in **(B)** indicate the magnitude of the mean adjusted p-values for

the PERMANOVA tests and the ANOVAs of the betadisper tests, respectively. The representation of p-values are as follows: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' Not significant.

A test of homogeneity of each variable showed significant differences when compared to the respective matched controls for cystic fibrosis (CF), Down Syndrome (DS), the presence of yeast, smoking, and celiac disease (**Figure 5.2B**). As described above with age groups, the result of this test indicates that the samples of one group for a given variable (e.g. those with CF) were significantly more similar to their median composition than the samples of the other group for that variable (e.g. those without CF) were to their own median composition. Interestingly, those samples in which yeasts were not detected were more homogeneous than those in which yeasts were detected. Meanwhile the individuals with CF, DS, and celiac disease, as well as smokers were more homogeneous than those without these disorders and non-smokers, respectively. There was no difference in homogeneity based on hypertension, the use of antibiotics, or the presence of *Candida*, though as with the general detection of yeast, the absence of *Candida* did tend to present greater homogeneity.

Some of these variables displayed particular significant differences when compared to their matched controls (**Table 5.2**). CF (Willis et al. 2021) and DS (Willis et al. 2020) have been explored in detail elsewhere, and so are not included in this table. Celiac samples had higher abundances of the genera *Phocaeicola* and *Staphylococcus*, and also had lower Faith's PD values and species richness (the number of species detected in a sample). Smokers had higher abundances of *Megasphaera*, *Fretibacterium*, and *Streptococcus*, and lower abundances of *Fusobacterium*, *Capnocytophaga*, *Bergeyella*, *Porphyromonas*, *Leptotrichia*, *Haemophilus*, *Neisseria*, *Lautropia*, and an unclassified genus of the class Gracilibacteria,

and also had lower Simpson and Shannon alpha diversity values. Samples in which yeast were detected, in particular those with *Candida*, had higher abundance of *Lactobacillus*. There were no individual taxa that differed significantly for hypertension or antibiotics.

Table 5.2: Significance of differentially abundant taxa and alpha diversity measures between indicated variable and matched controls. Columns indicate, in this order, the variable considered, the organism name or the alpha diversity value, the tendency of the difference in the considered variable ("↗": higher in those samples where the variable is true, "↘": lower), the mean adjusted p-value of the statistical comparison between variable and matched controls, and the numbers of matched control subsamples for which the test is significant. Rows are ordered first by the tendency in the indicated variable, with organisms/diversities that were greater first, and then by mean adjusted p-value within each variable group.

Sample group	Organism/Diversity	Tendency in indicated variable	Mean adjusted P	Significant subsample tests
Celiac	*Phocaeicola*	↗	0.08	63
	Staphylococcus	↗	0.09	50
	Faith's phylogenetic diversity	↘	0.0009	100
	Species richness	↘	0.0004	100
Smokers	*Megasphaera*	↗	0.0017	100
	Fretibacterium	↗	0.037	77
	Streptococcus	↗	0.07	66
	Phylum: Synergistetes	↗	0.003	100
	Phylum: Firmicutes	↗	0.042	76

	Fusobacterium	↘	0.0003	100
	Capnocytophaga	↘	0.0004	100
	Bergeyella	↘	0.0028	100
	Porphyromonas	↘	0.018	89
	Leptotrichia	↘	0.022	88
	Haemophilus	↘	0.03	83
	Neisseria	↘	0.031	77
	Lautropia	↘	0.051	65
	C.Gracilibacteria.UCG	↘	0.056	70
	Phylum: Fusobacteria	↘	0.0016	100
	Phylum: Patescibacteria	↘	0.025	93
	Simpson diversity	↘	0.002	100
	Shannon diversity	↘	0.029	83
Yeast detected	*Lactobacillus*	↗	0.053	61
***Candida* detected**	*Lactobacillus*	↗	0.01	96

Co-occurrence networks represent patterns of taxa that present correlated abundances across different samples (Kurtz et al. 2015). We constructed such networks for groups of samples differing in the studied variables and compared them in the search of unique associations between taxa. From these network comparisons, we derived a score that indicates the relative network uniqueness (i.e. the fraction of significant co-occurrences that are unique to that

variable - see Materials and Methods) (**Figure 5.3**). The most unique co-occurrence networks were seen in samples with CF (the specifics of this network were discussed in a previous publication (Willis et al. 2021)) and hypertension, followed by the absence of yeast (and specifically the absence of *Candida*), then the other two chronic disorders, DS and celiac, and finally smoking, and the reported use of antibiotics. This largely follows the trend in the PERMANOVA results presented above, wherein the samples that are more distinct from their matched controls largely display the more unique sets of significant co-occurrences, though did not necessarily follow the same pattern as the explained variances from those PERMANOVA tests. DS had greater R^2 values than hypertension and the detection of yeast, and *Candida* specifically, yet generally had a less unique network than those three variables. Neither did the network uniquenesses show the same trend as the homogeneity results, as hypertension, for instance, showed no difference in homogeneity, yet had the second most unique network, while smoking showed one of the strongest differences in homogeneity, and was the second least unique network.

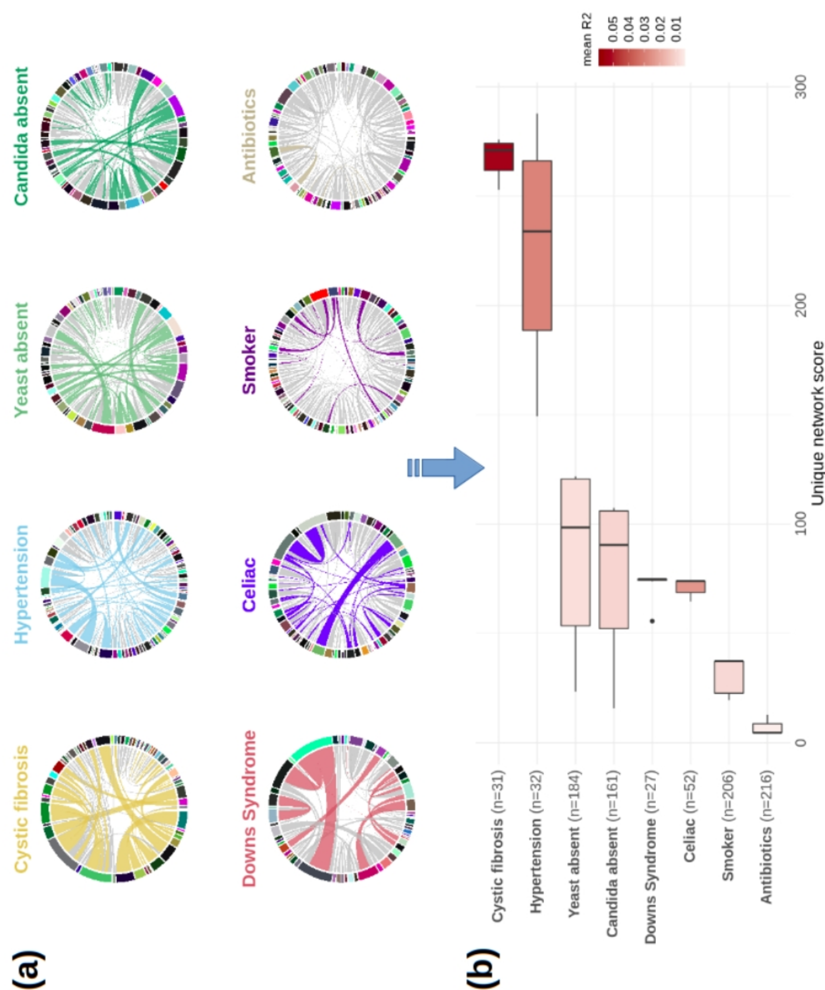

Figure 5.3: Uniqueness of core co-occurrence networks. (a) Chord diagram plots displaying significant co-occurrences and their relative strengths among the indicated samples for a given variable. Line widths are proportional to the strengths of the associations between genera. Colored lines represent associations that were unique to the samples indicated for a variable when compared to the network of at least one other variable. Grey lines are significant associations that were not unique to any of the displayed variables. **(b)** Boxes represent the distributions of scores derived from these unique associations. The values in a given box are the scores for that variable compared to each of the others. Thus the scores are relative only to these eight variables presented here. Scores were calculated as follows: for each variable, the co-occurrence networks were

calculated among each of the 100 subsamples, and we retained those associations which occurred only in the groups of interest (samples with the indicated disorder, smokers, antibiotic users, or those samples in which yeast was absent). Then for each variable, we calculated the number of only those associations which occurred in all 100 subsamples and in 0 subsamples of the other variable being compared, weighted by the strengths of those associations. Boxes are colored based on the mean R^2 value from the PERMANOVA tests comparing groups of a given variable. The number of samples for which a given variable was indicated in each subsample is indicated in the y axis.

5.4.3 Similarity of the oral microbiome composition among family members and classmates.

Our study included groups of samples that belong to members of the same family, and specified different degrees of relationship, such as parents and children, grandparents and grandchildren, partners, siblings and twins. In addition, given the active participation of schools in our project, we had several groups of samples from students attending the same school. Using an anosim test (analysis of similarities) on Aitchison distance matrices, we compared the similarity between the microbiome profiles of members of the same family or classroom, to determine whether the similarity was significantly higher than when compared to samples from different families or classes. With the exception of grandparents and grandchildren, all other relationships showed significantly greater similarity in oral microbiome compositions than was seen between samples from other families or classes (**Figure 5.4**). This similarity was highest for twins, followed by siblings, partners, family members (which included all of the non-classmate connections), parents-children, and classmates. Although the anosim statistic was higher for twins than for siblings, that merely indicates that the trend was stronger in twins. But twins were not statistically more significant to each other than siblings were to each other (P=0.33 for the Mann-Whitney test of Aitchison distances among twins vs those distances among non-twin siblings, the values represented by the blue boxes in **Figure 5.4**).

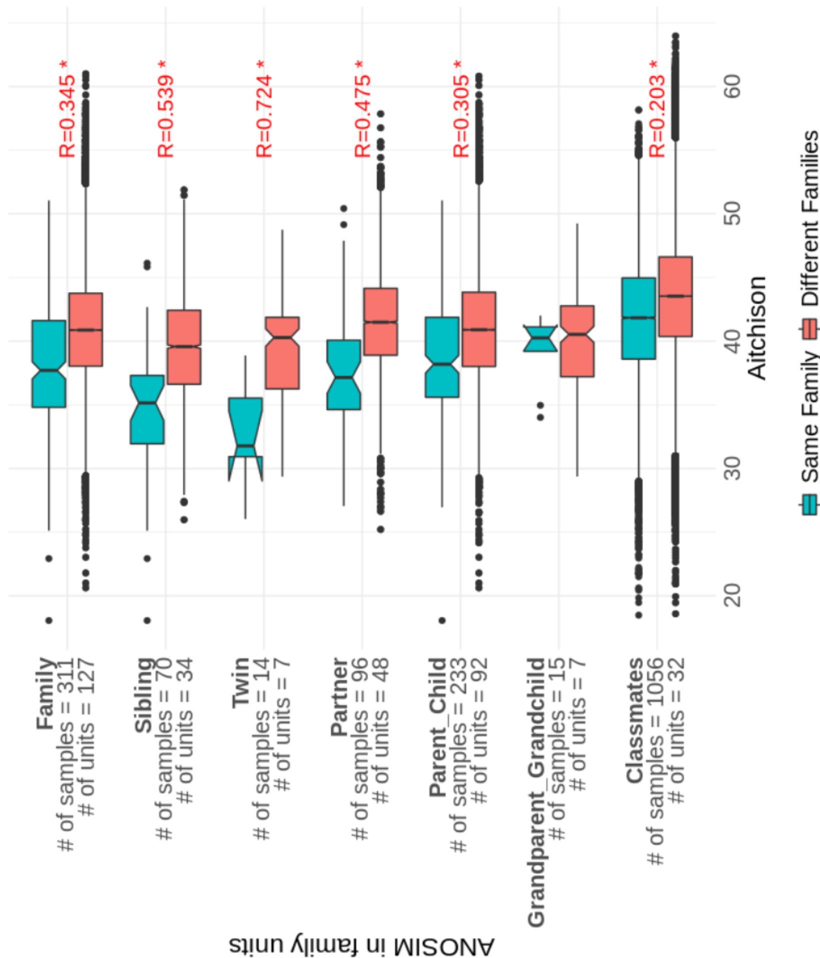

Figure 5.4: Anosim analyses of family units of various degrees of relationships, as well as classmates. Boxes show the distributions of Aitchison distance values between samples from the same unit (blue) or different units (red). The anosim R statistic is shown for those relationships that had significant results (anosim $P < 0.05$). The y axis labels indicate, for each relationship type, the number of samples for which that relationship occurred in at least one other sample, and the number of different units of two or more samples for which that relationship occurred.

5.5 Discussion

Our study builds on the first edition of the citizen-science project "Saca La Lengua" (Willis et al. 2018), which included 1,319 samples that were almost exclusively from 13-15 year old students in relative health. This first edition provided a comprehensive snapshot of the oral microbiome composition in adolescents and how it varied with different life-style parameters. In this second edition, we targeted a broad age range (7-85) as well as a few particular chronic disorders, namely cystic fibrosis (CF), Down Syndrome (DS), and celiac disease, in collaboration with relevant local and national patient associations. Participants also completed a comprehensive questionnaire about various daily habits, hygiene and diet. When collecting samples, we encouraged participants to bring along family members, and in the end 311 of the 1,648 total samples from this second edition of "Saca La Lengua" (SLL2) had some familial connection. To our knowledge, this is the first study to explore differences in the oral microbiome across a range of ages that is both balanced and encompasses most of the full range of the average human life expectancy (in Spain recent estimates were 86 years for women and 80 for men (Ho and Hendi 2018)). We have reported separately on the specific connections of the oral microbiome with DS (Willis et al. 2020) and CF (Willis et al. 2021), and here we present results based on the full SLL2 dataset.

5.5.1 Oral microbiome changes through age.

Studies exploring the trajectory of changes across the human lifespan have been limited, either comparing very disparate age groups (Nassar et al. 2014), a limited age range (LaMonte et al. 2019), or categorizing samples into very wide age ranges that do not effectively represent that entire range (Burcham et al. 2020). By spanning adolescence to late adulthood, our dataset provides some new insights into the topic. Our results show significant shifts in composition across time, wherein the younger and older samples were the most distinct, and the middle ages appear to represent an intermediate phase in which the oral microbiome is at its most

homogeneous. The parabolic trend in homogeneity was matched by the trend in both the Shannon and Simpson alpha diversity metrics, which were both lowest in approximately the 30-50 year old range. We can extrapolate similar results to ours from some of the studies mentioned above. In one, samples from women between the ages of 53 and 81 showed no significant change in alpha diversity (LaMonte et al. 2019), and at these ages, the diversity values in our samples have already risen to a relative plateau. In another study, a citizen-science project much like our own, youth samples (ages 8-16 with a mean age of 10) showed greater alpha diversity than adult samples (ages 20-75 with a mean age of 34) (Burcham et al. 2020). Though their "adult" group reaches up to age 75, the mean age of 34 suggests a similar result to our own.

Despite the parabolic trends in both alpha diversity and homogeneity across age, we did not find evidence of these patterns in the abundances of any particular organisms. Instead, we saw that with age there were statistically significant decreases in the genera *Alloprevotella*, *Streptobacillus*, *Haemophilus*, *Prevotella*, *Granulicatella*, and *Bergeyella*, and increases in *Anaeroglobus*, *Eikenella*, *Fretibacterium*, *Comamonas*, *Olsenella*, and *Phocaeicola*. As noted above, a typical trend in the aging oral cavity is an increase in the prevalence and severity of periodontal disease (G. Hajishengallis 2014; Feres et al. 2016; Belibasakis 2018). With the exception of *Olsenella*, each of the genera that were increased with age in our samples has been associated with periodontitis (Pérez-Chaparro et al. 2014; Camelo-Castillo et al. 2015; Bao et al. 2017; Xuyun Zhang et al. 2020; S. S. Socransky et al. 1998; Fujise et al. 2004; Deng et al. 2017; Lundmark et al. 2019; Nibali et al. 2020; Cao et al. 2018; Fernandez Y Mostajo et al. 2017). While we do not have data on salivary flow rate or nitrate levels from our samples, salivary flow rate has also been shown to decrease in the elderly (Percival, Challacombe, and Marsh 1994; Nassar et al. 2014), and is proportional to the systemic concentrations of anti-

inflammatory nitric oxide (NO) (Granli et al. 1989), the local concentrations of immunoglobulins and various molecules important for the mineralization of tooth enamel, and also maintains pH by removing substrates for the microbiota, as well as their acidic byproducts (Marcotte and Lavoie 1998). Indeed, we also found that pH decreased with age in our samples. Thus, future studies which track oral microbiome changes across age along with periodontal health, salivary nitrate levels and systemic NO levels, which result from an enzymatic process in oral commensal bacteria that humans cannot perform themselves (Hyde et al. 2014; Rammos et al. 2016), and how these combinations relate to inflammaging, would warrant further attention.

A noteworthy observation in the changes across age in our study is that those genera that decreased with age were typically among the most abundant oral taxa, while those that increased were found at relatively low or median abundances (**Supplementary Figure S5.1**). We speculate that the elderly oral microbiome may be more susceptible to colonization and establishment of rare opportunistic species whose growth is hindered by the more efficient immune responses in younger oral cavities. This would be in line with hypotheses proposed to explain the higher prevalence of periodontitis through aging (G. Hajishengallis 2014), which relate it to different factors, such as the accumulation of tissue damage, weaker immunity, increased adipose tissue (a source of cytokines), decreased anti-immflamatory sex hormones, diminished physical activity, and increased oxidative damage. Some of these factors may also explain the relatively high alpha diversity values in the elderly samples, though not necessarily in the teenaged samples. These instead may be a result of the continually developing microbiome composition, which appears to reach a more stable state in the 30s and 40s. It should also be noted that two of the alpha diversity measures we looked at (Faith’s phylogenetic distance and species

richness) were only higher in the older samples, and remained consistent up to the ages of approximately 50-55. Thus younger and older microbiomes present higher diversities of somewhat different natures, with the elderly being characterized by a higher number of species (Species richness) and more phylogenetically diverse compositions (Faith's PD), whereas both extreme age groups present similarly diverse microbiomes in terms of balanced representations of the different taxa (namely Shannon's and Simpson's diversity indexes). Other age groups, in comparison, are characterized by less diverse microbiomes, with more clear separations between dominant and minority taxa.

5.5.2 Chronic disorders, smoking and the presence of yeasts in the oral cavity, are important drivers of the oral microbiome composition.

The presence of chronic disorders such as CF and DS, the most impactful factors seen in this dataset, and their particular impacts were described elsewhere (Willis et al. 2021, 2020). Persons with CF, DS, or celiac disease, as well as smokers, had significantly more homogeneous compositions compared to the matched controls without these disorders and non-smokers, respectively. This finding suggests that those three disorders and smoking not only differentiate those samples significantly from their matched controls, but also that the bacterial compositions are shaped in consistently similar directions (i.e. towards a specific signature), while the controls are comparatively more variable. The reverse was the case for the detection of yeast, so that perhaps greater prevalence of these fungi promote a departure from typical bacterial ecosystems. This supports the existence of diverse synergistic and antagonistic ecological interactions between yeasts and bacterial species, and a role of fungi as keystone species in the oral ecosystem. Alternatively, the presence of yeasts might be a

consequence of already unbalanced microbiomes, suggesting they are opportunistic colonizers. In both cases, they could be considered as potential biomarkers for altered microbiomes. Finally, hypertension and antibiotics displayed significant differences to their matched controls, but there was no difference in homogeneity, so these factors did not direct the differences in any specific manner, perhaps depending on the specific antibiotic used or the severity of hypertension, for which we do not have specific information.

The particular differences seen in some of these variables here corroborate some findings in the literature. A study found that never-smokers and former smokers did not differ from each other in composition, but both differed significantly from current smokers, and that smokers had higher *Streptococcus* and *Atopobium*, and lower *Capnocytophaga*, *Leptotrichia*, and *Peptostreptococcus* (J. Wu et al. 2016). We found the same for *Streptococcus*, *Capnocytophaga*, and *Leptotrichia*. Three studies found smokers had increased *Megasphaera* and decreased *Neisseria (Kato et al. 2016; Mason et al. 2015; Vallès et al. 2018)*, though one of those (Kato et al. 2016) reported the family *Veillonellaceae*, of which *Megasphaera* is a member. There was also agreement with our finding of a decrease in *Haemophilus (Mason et al. 2015; Vallès et al. 2018)*, *Lautropia*, *Fusobacterium*, and *Leptotrichia (Vallès et al. 2018)*, though depending on the study, there were opposite findings for *Fusobacterium*, *Streptococcus*, and *Porphyromonas*. A study which described two distinct oral mycotypes (sample clusters defined by the fungal composition), found that one of these was dominated by *Candida*, and was enriched in *Lactobacillus* and *Propionibacterium (Hong et al. 2020)*, the former of which matches our own finding here. If their reported mycotypes are indeed ubiquitous structures of fungal composition, it may be that our samples also follow this dichotomy and the non-*Candida* samples would perhaps fall in the other mycotype, which was much more

diverse in fungi, though this would require further investigation.

Although the relative scores of uniqueness of the co-occurrence networks of the different variables mentioned here did not precisely match the patterns from either the PERMANOVA or homogeneity tests, the unique co-occurrences among particular sample groups suggest underlying ecological differences present under the various conditions. The networks of CF, for instance, were discussed at length elsewhere (Willis et al. 2021). Moreover, there was greater variation in the uniqueness scores for hypertension and absence of yeasts/*Candida* than in the other variables, as can be seen in **Figure 5.3B**, and thus a greater proportion of the associations in these networks were also seen in the networks of other variables. CF, as a contrasting example, had relatively little variation, and thus consistently displayed many of the same associations that did not appear in the networks of other variables, so its network is more universally unique. Similarly, although the uniqueness score for smokers was relatively low, it also had low variation, so the relatively few unique associations were also universally unique. The caveat to these findings is that here we only compare the networks of those eight variables which we found to significantly differentiate individuals from matched controls (as in **Figure 5.2**). To better understand the underlying ecologies, a more expansive comparative exploration of co-occurrence networks in particular cohorts should be performed.

5.5.3 Similarity of the oral microbiome composition among family members and classmates.

Our finding that the oral microbiomes among family members are more similar to each other than to those of non-family members

corroborates the trends seen in the literature (Stahringer et al. 2012; Song et al. 2013; Shaw et al. 2017; Burcham et al. 2020). One of these studies found that twins were not more similar to each other than non-twin siblings (Burcham et al. 2020), which we have corroborated in our results here, and another found that monozygotic twins were not more similar to each other than dizygotic twins (Stahringer et al. 2012), which was also seen in the gut microbiome (Yatsunenko et al. 2012). Moreover, a study using a genome-wide analysis of SNPs to compare genetic similarity with microbiome composition found no significant association (Shaw et al. 2017). All of this evidence points to the conclusion that the shared environment of the home strongly influences oral microbiome composition, more so than host genetics. In agreement with this, the only familial relationship that did not show a significant similarity in our data was that of the grandparent and grandchild, which is the connection least likely to share a living space. Indeed, while twins had the highest similarity score, they were not significantly more similar to each other than non-twin siblings, further supporting the findings in the literature. We even saw that, among the teenage samples obtained from different high schools, the oral microbiomes were more similar among classmates than non-classmates, though this was the comparison with the lowest magnitude of similarity among those that were significant (lowest anosim R statistic), as would be expected since it generally entails more distanced interactions than those among family members. The result about classmates may suggest that a regularly shared environment, even if only for a few hours a day, could impact the oral microbiome composition. Future studies could explore this notion further, for instance focusing on workplaces with close physical proximity like shared offices in contrast to more distanced outdoor working groups, as in construction sites.

5.6 Conclusions

This second edition of the citizen-science project Saca La Lengua (SLL2) extends the results of the first edition (Willis et al. 2018),

which provided a snapshot of the oral microbiome of teenagers in relative health across Spain. Here we have displayed the differences that occur across age, wherein a number of genera of bacteria either increase or decrease in abundance, and people in middle ages typically have more homogeneous compositions than teens or seniors, as well as lower alpha diversity, and seniors tend to harbor a greater number of low abundance organisms and a more acidic oral environment. In SLL2 we also compared the general influence of a number of different health and lifestyle factors on the oral microbiome composition. Cystic fibrosis and Down syndrome were the most impactful in terms of differentiating the composition, and the samples with these chronic disorders were significantly more homogeneous than matched controls, suggesting the disorders tend to direct the composition of the oral microbiome in specific and consistent ways. A similar effect was seen with celiac disease, smoking, and the absence of yeast species, while hypertension and recent use of antibiotics significantly differentiated samples, but did not show a difference in homogeneity. Nonetheless, hypertension, along with cystic fibrosis, displayed more unique associations between bacterial taxa in co-occurrences networks compared to these other variables, suggesting particular underlying ecologies. We also expanded upon findings in the literature that shared environments are important in shaping the oral microbiome. We saw that family members that typically live within the same household tend to have significantly more similar compositions compared to non-family members, and that twins are not significantly more similar than non-twin siblings, supporting the idea that the environment, more than host genetics, shape the microbiome. Furthermore, we saw that students in the same school were more similar to each other than those from different schools. This opens a door to further studies of shared spaces, like different working environments, as our finding suggests that regularly sharing the same environment for even a few hours impacts the microbiome. This study describes the manners in which an assortment of factors affect the oral microbiome in the Spanish population. The results lay some groundwork for future studies to expand upon in dedicated cohorts for particular factors, as well as in

other populations.

Abbreviations

SLL2	Saca La Lengua - 2nd edition (“Stick out your tongue”)
PRBB	Barcelona Biomedical Research Park
PBS	Phosphate-buffered saline
r.t.	Room temperature
PERMANOVA	Permutational multivariate analysis of variance
ASV	Amplicon sequence variant
DS	Down Syndrome
CF	Cystic fibrosis
Faith’s PD	Faith’s phylogenetic diversity
NO	Nitric oxide
CREAL	Center for Research into Environmental Epidemiology
ERDF	European Regional Development Fund
AGAUR	Catalan Research Agency
MEIC	Spanish Ministry for Economy, Industry and Competitiveness

Declarations

Ethics approval and consent to participate

All participants, and at least one of their parents or legal guardians for those under the age of 18, signed a consent form to use their saliva samples for microbiome research. This consent form and the purpose of this project received approval by the ethics committee of the Barcelona Biomedical Research Park (PRBB).

Consent for publication

All participants signed an informed consent form, and data is anonymized.

Availability of data and materials

The fastq files for the paired forward and reverse reads of the 16S rRNA sequencing of the 1,648 oral rinse samples used for the analyses in this study (57,221 Mb) were uploaded to the Sequence Read Archive (SRA) with the BioProject accession number **PRJNA667146** and can be found here: http://www.ncbi.nlm.nih.gov/bioproject/667146. We also provide a table with the results of MALDI-TOF analyses of fungal composition (44 kb), which can be found here: https://github.com/Gabaldonlab/ngs_public/tree/master/SLL2. The unique and anonymized identifiers for each sample can be found at the beginning of each fastq file, and these correspond to the row names in the fungal composition tables.

Competing interests

The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Funding

The project was financed by the CRG through Genomics and Bioinformatics Core Facilities funds, and by the EduCaixa programme through funds from the Fundación Bancaria “La Caixa”, with the participation of the Center for Research into Environmental Epidemiology (CREAL), and the “Centre d’Excel·lència Severo Ochoa 2013-2017” programme (SEV-2012-02-08) of the Ministry of Economy and Competitiveness. Eppendorf, Illumina, and ThermoFisher sponsored the research by donating some materials and reagents. TG group acknowledges support from the Spanish Ministry of Science and Innovation for grant PGC2018-099921-B-I00, cofounded by European Regional Development Fund (ERDF); from the CERCA Programme / Generalitat de Catalunya; from the Catalan Research Agency (AGAUR) SGR423; from the European Union’s Horizon 2020 research and innovation programme under

the grant agreement ERC-2016-724173; from Instituto de Salud Carlos III (INB Grant PT17/0009/0023 - ISCIII-SGEFI/ERDF).

Author Contributions

Conceptualization, T.G. and J.R.W.; formal analysis, J.R.W.; validation, J.R.W.; visualization J.R.W., T.G.; methodology, E.S., S.I.G., E.K., L.C. L.A.B., N.A.S., M.A.T., A.P.S., A.B., C.C., J.P., T.G., M.A.; data curation, L.C., J.P.; writing original draft, J.R.W., S.I., E.S., T.G.; supervision, T.G.; project administration, T.G.; funding acquisition, T.G.; resources T.G., J.P., J.H. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We are thankful to all citizens that participated in the second edition of the "Saca La Lengua" project by contributing samples and sharing ideas (see more details here www.sacalalengua.org). In particular, for the work described here we are extremely thankful to all the students and teachers of the schools that we have visited, for their enthusiasm and questions, which helped to expand our hypotheses, to those responsible for civic centers, libraries, museums, bars, for giving us their spaces and giving us the possibility of organizing all events, and to the following national associations: Spanish Federation of Cystic Fibrosis (www.fibrosisquistica.org), Down España (www.sindromedown.net), and Federación de Asociaciones de Celíacos de España (www.celiacos.org); and local associations: Asociación Madrileña de Fibrosis Quística, Associació Catalana de Fibrosi Quística, Cocemfe Cantabria, Down Lleida, Down Bilbao, Down Vigo, Down Málaga, Associació Celíacs de Catalunya, Celíacos Euskadi, Federación de Asociaciones Celíacos Andalucía, and Asojum Murcia. Only with their effort are studies like this possible. The authors acknowledge the CRG Genomics Core Facility, CRG Bioinformatics Core Facility, CRG Biomolecular Screening and Protein Technologies Unit, CRG Communication and Public Relationships Department, and UCT ICTS High Performance Computing unit for providing access to the computing

facilities. CRG authors acknowledge the Spanish Ministry for Economy, Industry and Competitiveness (MEIC) for the EMBL partnership, and Centro de Excelencia Severo Ochoa. The following reagents were obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: (1) Genomic DNA from Microbial Mock Community B (Even, Low Concentration), v5.1L, for 16S rRNA Gene Sequencing, HM-782D, and (2) Genomic DNA from Microbial Mock Community B (Staggered, Low Concentration), v5.2L, for 16S rRNA Gene Sequencing, HM-783D.

Discussion

Chapter 6: Summarizing Discussion

I have developed a sort of personal cliché at parties or dinners when someone asks what I do for a living, and I am sure I am not the only microbiomist to do so. I try to incite a bit of lighthearted existential angst by pointing out that a person is not only not alone in their own body, but that they are in fact outnumbered ten to one by microorganisms. It also makes for a snappy hook to open a paper (unsubtly repurposed here), although that ratio is based on an erroneous and evocatively described "back-of-the-envelope calculation" that has been cited hundreds of times since the 1970s (Savage 1977), and has recently been more rigorously estimated to be closer to a one to one ratio (Sender, Fuchs, and Milo 2016). Nonetheless, the implication is clear - the microbiome is an inextricable part of human existence. They may not perfectly fit the traditional definitions of tissues and organs, but the microbiomes of the different habitats throughout the human body constitute something like an organ system akin to the nervous or cardiovascular systems. Closely related microbes that produce biofilms and other proximal and strongly associated microbes might be likened to tissues, and the groupings of these in a particular niche to organs. The oral microbiome, for example, might itself be considered an organ, or a system of such organs found in the buccal mucosa, the tongue dorsum or the gingival plaques. Like any organ or system, the oral microbiome performs a set of regular and defined functions, such as the reduction of nitrate to produce anti-inflammatory nitric oxide (Hyde et al. 2014), or the maintenance of local pH levels (Bowen et al. 2018).

Thus, to understand the functions of human health, we must also understand the symbiosis between the human host and its microbial residents. As I discussed in **Chapter 1**, it is only in the last few decades that we have developed techniques which allow us to begin to approximate the totality of the microbiome, rather than relying on data regarding individual culturable organisms and observable

phenotypic traits gleaned from microscope slides. As a result, we have experienced a sort of evolutionary radiation in our conceptions of the structures and functions of the human microbiome, rapidly discovering and classifying previously unknown organisms from all across the tree of life (a process which has seemingly just begun (Bernard et al. 2018)), and subsequently developing new techniques for exploring their functions (Moran 2009; Heyer et al. 2017) and outputs (Fiehn 2002). These methods that are used today represent significant technological leaps, yet they each inherently reveal limited dimensions of the overall nature of the microbiome. If we were to call the era before high-throughput sequencing the field's infancy, today it remains in its adolescence. As the field matures, it should continue to move towards a more integrative approach using these and potentially as yet unanticipated techniques.

However, most of the sequencing-based studies of the microbiome to date have focused on classifying the organisms present within a particular body site. This cataloguing exercise is the most straightforward, but also the most fundamental, element of microbiome research, and one which is far from complete. The collective goal of these classification studies is to produce a sort of tapestry that weaves together the generalized compositions of microbiomes across different populations, giving us an abstracted picture of how these populations compare to each other. But the tapestry remains frayed along many of the edges, with a number of holes where understudied populations would lie. Like I mentioned in **Chapter 1**, there is a strong bias toward studying populations from the so-called “WEIRD” (Western, Educated, Industrialized, Rich, Democratic) nations, though sociopolitical and logistical factors limit accessibility to non-WEIRD samples that would provide the threads to fill in those gaps. Geography and culture are the broadest separating factors between these populations and their microbiomes, yet even within a given widely studied group, there remains room for expansion with ever finer threads, as more specific factors of lifestyle and health can be explored. This was the aim of “Saca La Lengua,” the large-scale oral microbiome project upon which this thesis is based, and which examined the oral microbiomes of particular subsets of the Spanish population. As I

move forward, I will discuss what the two editions of Saca La Lengua have added to the picture, how my techniques of analysis and those of the field at large have developed since the project began, and how this knowledge might be used in the future.

6.1 Lessons from Saca La Lengua

For this citizen science project, we relied on sequencing of the V3 and V4 hypervariable regions of the 16S rRNA gene. The classification capabilities of 16S sequencing suited the goals of Saca La Lengua (SLL), allowing for the collection and sequencing of approximately 1500 oral rinse samples. Although 16S sequencing inherently has lower taxonomic resolution than whole metagenome shotgun sequencing (WMS), it largely avoids the issue of predominating host genetic material, which can account for over 90% of DNA in a sample from the oral cavity (Marotz et al. 2018), and which would otherwise have to be sequenced, detected and discarded. We wanted to explore a representative sampling of a homogeneous segment of the population, adolescents in relative health, in order to obtain a clear and reliable picture of the composition of the oral microbiome in that group. This demographic was ideal for the other basis of SLL, its citizen science-based approach, and with the objective of increasing the interest in science in adolescents at the time when they are preparing to define their career paths toward academic specialization in high school. The design allowed us to travel to high schools all across Spain for the sample collection, where a member of the team could not only give a presentation about the ideas surrounding the project and the microbiome in general to raise scientific awareness and interest among youths, but also to receive input from them and their teachers about what might be the most interesting angles to approach the analyses. This type of collaboration between scientists and the general public is a primary aim of the citizen science movement (National Academies of Sciences, Engineering, and Medicine et al. 2019; Gura 2013). Some of their ideas were incorporated into the questionnaires that were completed by all of the participants, allowing them to take an active

role in the development of the project, with the hope of stimulating wider public interest in the processes and results of basic research.

I use the term microbiome throughout this discussion and the earlier chapters, though it is arguably not the most appropriate nomenclature in this context. The microbiome is often confounded with the bacteriome, perhaps because the abundance of bacterial cells in a human host is generally on the order of 10 to 100 times greater than that of eukaryotic and archaeal cells (Sender, Fuchs, and Milo 2016). Indeed, the first edition of SLL is based entirely on sequencing of bacteria, and the second edition includes only a limited perspective of the fungal component of the oral microbiome. However, I feel the generalized use of the term microbiome is justified for projects like these, not only because bacteria comprise the majority of the biomass in the human microbiome, but also because such projects are reductionist by their nature, a notion upon which I will expand later in this discussion in the **section 6.3.1**, and so aim to explore only particular components of the microbiome. It may be a pedantic argument, but the literature continues to reexamine the terminology within the field as it progresses (Ursell et al. 2012; Marchesi and Ravel 2015; Berg et al. 2020), so it is important to be explicit in our definitions.

The results from the SLL studies that this thesis comprises paint a vivid picture of the oral microbiome of the Spanish population. Based on the comparisons that have been made between westernized and non-westernized samples, which show significant separation in the compositions of gut and oral microbiomes from the different societies (Clemente et al. 2015; Segata 2015; Eisenstein 2020; Sánchez-Quinto et al. 2020), the connections that we have reported with drinking water, age, and particular chronic disorders could perhaps serve as generalized trends for other western European nations as well. If those studies do actually point to prevalent and consistent societal trends, a robust and representative dataset from oral microbiomes in Spain, like the 2967 total oral rinse samples from the two editions of SLL, should be indicative of trends from nations with generally similar diets and lifestyles. While it is too early in the development of the field of

microbiomics to definitively make such a claim, it is likely that the implications of our results are extendible beyond just the immediate context of our own sample set.

6.1.1 SLL1 - Drinking water, stomatotypes, and the “core” oral microbiome

The most impactful results from the first edition of SLL (**Chapter 2**) stemmed from the stomatotypes that we calculated. The concept of stomatotypes, a clustering of oral microbiome samples in a dataset based on a distance matrix, may already be obsolete in microbiome studies, (see the **sections 6.2** and **6.3.1** for further discussion), but I argue that they are still relevant for exploring trends, and they helped us to recognize some important features in our own data. This was one of the first kinds of analyses that I learned of for exploring microbiome data globally (Arumugam et al. 2011), and I approached it from many different angles, like plotting the distributions of each stomatotype on a map of Spain, based on the location from which a sample was obtained, to look for geographical trends (**Figure 6** in **Chapter 2**). When I showed the maps to my principal investigator and thesis director, Dr. Toni Gabaldón, he suddenly recalled a strikingly similar map of the ionic contents of water across Spain - it was his Archimedes moment, though without the shouting and public nudity. So we quickly contacted researchers that had analyzed the content of public drinking water from cities and towns all across Spain, and indeed we found significant correlations between the measured values and the abundances of taxa in our data, many of which were among the significant drivers of the two stomatotypes. We performed a similar analysis for the data from the second edition of SLL, though we did not include the results in the final version of the paper (**Chapter 5**), and again we saw the same trend, wherein many genera that strongly influenced the calculations of stomatotypes either positively or negatively associated with various ion concentrations (see **Appendix I** - “Drinking water impacts the oral microbiome composition”), despite the fact that by this time our statistical analyses had evolved (using centered log ratios instead of relative

abundances and generalized linear models including potential confounding variables instead of Pearson correlations - see **section 6.2 (Notes on methodologies)**). Subsequent studies by others have confirmed associations between the human microbiome and the composition of drinking water (Sinha et al. 2021; Bowyer et al. 2020).

We also compared our stomatotypes to others found in different populations, with some consensus and some deviation, which led us to suggest that our stomatotypes may represent ubiquitous equilibria of oral microbiome compositions, with the possibility that others may exist as well in different populations. In retrospect, this position may appear somewhat overstated because the comparisons between the stomatotypes among the different studies were based on trends in the abundances of only a handful of genera, including *Neisseria*, *Haemophilus*, *Prevotella*, and *Veillonella*. Though they are among the most abundant, the conclusion did not consider the trends in any rare genera. Nevertheless, this comparative analysis represents an important early foray into the search for a global image of the human oral microbiome and the different shapes it may take. In addition, our stomatotypes guided our perception of the gradients of abundances of many taxa, which can be useful in understanding the potential shift between alternative stable states (see **section 6.3.2**). This is a very similar idea to that of stomatotypes, though more abstract, as it does not necessarily impose strict boundaries and may account for transition states (Fukami and Nakajima 2011; Costello et al. 2012; Amor, Ratzke, and Gore 2020; Van de Guchte et al. 2020), in which abundances at a given time may be in flux and have not yet reached a relative stability. This is likely a more realistic conception of the state of the microbiome at a given time. The genera which most strongly influenced the separation of samples into particular stomatotypes generally had detectable gradients of abundances across the spectrum of compositions within our dataset. Thus we felt that these revealed the patterns in potential stable and transition states of the oral microbiome.

One of those studies with which we compared stomatotypes, based

on a large cohort of oral rinse samples from Japanese adults (Takeshita et al. 2016), listed what they referred to as “core” genera of the oral microbiome, those that were present in at least 75% of their samples. We performed the same calculation and found that 20 of the 32 “core” genera in our samples corresponded to theirs. This would seemingly hint at a persistently present set of organisms across human populations, especially considering the disparate demographics between the two studies. But it also raises the question of how this core should be defined, which others have also attempted (Zaura et al. 2009; Shade and Handelsman 2012; K. Li, Bihan, and Methé 2013; H. Chen and Jiang 2014), though the parameters are often arbitrarily designated as thresholds of presence vs abundance or of similarities in abundances for given taxa. Elsewhere it has been suggested that cores may exist only among subpopulations (Hamady and Knight 2009), which may be plausible based on the notable differences seen between WEIRD and non-WEIRD samples (Clemente et al. 2015; Pasolli et al. 2019), but which again would require standardized parameters of entry for core taxa.

Defining this core, or the series of cores present in particular populations, is central to discovering and understanding eubiosis in the oral microbiome, the optimal microbial balance, which ultimately is a primary goal of microbiome research. Hippocrates is believed to have said that “death is in the bowels” and “poor digestion is the root of all evil” (I have not found an original source of these quotes, but they are popular in today’s gut microbiome papers (Iebba et al. 2016)), and his notion was modernized by the Nobel laureate Elie Metchnikoff after the widespread acceptance of germ theory and his explorations of dysbiosis, or imbalances in the gut microbiome (Metchnikoff 1907). So microbiome researchers have a primal impulse to seek out causes of dysbiosis, so that we may continue to improve methods that ensure eubiosis. The literature is rife with attempts to link the oral microbiome to diseases, as we do in **Chapters 3** and **4**, based on states of dysbiosis relative to supposedly healthy controls, though these comparisons are limited by our current conception of relative eubiosis. Some studies have attempted to combine multiple publicly available

datasets from different populations (Leung, Wilkins, and Lee 2015; Pasolli et al. 2016; Aguirre de Cárcer 2018; Pasolli et al. 2019). As more data is produced and analyzed in this way, we will better be able to approach nuanced definitions of what may be called the pan-microbiome and accessory microbiome, analogous to the pangenome, which encompasses all genes that are present in all strains of a clade, and the accessory genome, those genes that are specific only to subsets of the clade or individual strains. While these efforts will still be subject to the inherent limitations of compositional datasets, which may complicate cross-study examinations (see the **sections 6.2** and **6.3.2**), they remain invaluable stepping stones in the progression of microbiome studies.

6.1.2 SLL2 - Chronic disorders and the oral microbiome

When conceiving the study design for the second edition of SLL, our group also felt the allure of dysbiosis explorations and targeted three chronic disorders whose connections to the oral microbiome were both understudied and likely meaningful. We chose Down Syndrome (DS) because of the unique immune and oral physiological characteristics (relatively late eruption of teeth, microdontia, lower salivary flow, greater number of missing teeth and greater dental spacing), cystic fibrosis (CF) because of the oral cavity's potential as a reservoir of microbes for the lungs, including the oral commensal *Pseudomonas* which is one of the primary pathogens in CF lung infections, and celiac disease because of the unique dietary restrictions that it imposes. As part of the questionnaire that all participants completed, we also asked if they experienced any other chronic disorders, and these, along with the three targeted disorders, accounted for 311 of the 1648 samples. Some of the other health issues included hypertension, diabetes, migraines, or hypothyroidism. We published dedicated analyses for both DS and CF (**Chapters 3** and **4**, respectively), but not for celiac, as we found relatively few significant differences compared to matched control samples, though it did have a notable global impact on the oral microbiome, as shown in **Chapter 5**. Both DS and CF

have been connected with oral diseases in the literature, though with opposite trends in regard to periodontitis (greater in DS and lower in CF) and some ambiguity in regard to dental caries. The differential abundances in our results of both studies largely supported these trends, further bolstered in the CF paper by the analysis of co-occurrence networks that pointed to the underlying ecological contexts.

In **Chapter 5**, we showed that samples from individuals with DS, CF, and celiac disease, as well as smokers and samples in which yeasts were not detected, not only separated significantly from matched controls in their overall compositions, but also that those compositions were significantly more homogeneous, suggesting that each of these variables has a consistent impact on the microbiome, driving its structure in particular directions. There were 32 individuals with hypertension, though it was not a disorder that we actively sought to study, and so did not have any other relevant information. Thus, we were left to speculate that, while the samples did differ significantly from controls, they may not have had more homogeneous compositions due to varying levels of severity. This is an instance in which metatranscriptomic data may have been illuminating. As we and our collaborators showed (Cutler et al. 2019), along with other studies (Hyde et al. 2014), oral bacteria are vital for the reduction of nitrate, which produces the vaso-dilating nitric oxide that circulates in our bloodstreams. We might hypothesize that our samples from individuals with hypertension may have had diminished nitrate-reduction potential, an effect that was not reflected in specific taxonomic differences, as there are many oral taxa with the required machinery (Rosier et al. 2020), but may have been revealed by insights into the functionality within samples.

We were careful to note that the specific conclusions about DS and CF that **Chapters 3** and **4** draw are speculative, as they are based on relative increases or decreases of taxa that have elsewhere been shown to follow similar trends in either periodontitis or dental caries. Essentially, our results strengthened the correlative links between oral diseases and DS and CF based on the correlative links

with taxa in our and other studies. But these results cannot by themselves reveal true causal links, an inherent limitation of modern microbiome studies, though researchers are often eager to wrest this kind of conclusion from association studies (Maruvada et al. 2017; Lynch, Parke, and O'Malley 2019; Walter et al. 2020). There are instances in which interventional studies can indeed point to causal relationships, such as our collaboration mentioned above in which an antibacterial mouthwash inhibited the nitrate reducing activity of oral bacteria, thus hindering exercise recovery (Cutler et al. 2019). However, in the case of exploratory classification-based studies, like our own from SLL, in order to determine causal connections between the microbiome and disease, they could be bolstered by a combination of omics data. Whole metagenome sequencing could be used so as to identify the functional potential within microbiome samples, but more appropriate would be metatranscriptome and metaproteome data to directly observe the functional impacts at the time of sampling. Metabolomics data would further elucidate the actions and effects of microbes on their human hosts. There is also the potential of epigenetic modification of microbes as a result of the action of the human host cells (Beaulaurier, Schadt, and Fang 2019; Morovic and Budinoff 2021), or of human host cells as a result of the action of resident microbes (Oelschlaeger 2010; Celluzzi and Masotti 2016), not to mention the potential inter-domain interactions which are rarely explored today (Rowan-Nash et al. 2019). Then there is the question of directionality. When is a shift in composition a cause of a disease state and when is it a result? In **Chapter 1** we use the example of *Neisseria* species as being potentially opportunistic in instances of dental caries, since it is acidogenic and may be taking advantage of the "hydrogen sink" created by methanogenic archaeal species and/or *Treponema*. But which of these groups took action first is difficult to discern, as is whether the alterations in the host environment promoted their activity or they exerted themselves upon the conditions of their environment. All of this is to say that, if we hope to determine causality in microbiome research, we would need a synthesis of omics techniques that would create improved simulations of the microbiome in its natural state (see **section 6.3.3**).

6.1.3 SLL2 - Aging in the oral microbiome

Part of the design of SLL2 was also to collect samples from a wide range of ages, as we had deliberately sought a homogeneous sample set for the first edition. We were fortunate to receive wide participation such that, with the subsampling approach we used for analyses to match age bins by gender and geographic location (see methods section in **Chapter 5**), we could have about 30 samples from each ten year age bin in each iteration. As I mentioned above, in **Chapter 2** we compared some basic characteristics of our samples from teenagers to those of Japanese adults in a separate study, which hinted at some pervasive features in the oral microbiome, but here we provided insights into the trajectory of the oral microbiome across age. As with our results surrounding chronic disorders, the specific taxonomic changes allowed merely for speculative conclusions that would require more in depth examination, but we also saw global effects that are likely more informative about fundamental attributes of the oral microbiome, in particular the parabolic trend in homogeneity. Samples from middle-aged individuals were both the least unique and the most homogeneous, and together with evidence from various alpha diversity measurements, we suggested that while youths and the elderly both have relatively variable compositions, the youth microbiomes are likely still developing alongside the rest of individuals' bodies, and the elderly are more susceptible to colonization by rare and opportunistic bacteria, possibly due to lower efficiency in immune responses.

Another interesting aspect of aging in the oral microbiome that we explored was its connection with Down Syndrome in **Chapter 3**. The literature has suggested that DS is a prematuring aging disorder, typically presenting with early immune senescence, increased plasma inflammatory markers at levels similar to that seen in the elderly, oxidative stress as a result of mitochondrial malfunction, and accelerated epigenetic modification rates (Franceschi et al. 2018; Horvath et al. 2015). So we wanted to determine if there were signs of premature aging in the oral microbiome in Down Syndrome as well. Our study design allowed

us to approximate a comparison of the trajectory of the oral microbiome across age in the DS and non-DS samples, where we expected that both or either the DS oral microbiome would be more similar to older non-DS microbiomes and the shifts in composition across age in DS would be more rapid than those in non-DS. Yet neither of those was the case. In fact the DS samples were most similar to the youngest non-DS samples, and not only was there no evidence of greater difference in composition with greater difference in age in DS compared to non-DS, there was no such correlation at all. We saw this as a sort of “anti-aging” effect in the DS oral microbiome that is likely related to the distinctive physiology in the DS oral cavity, like the diminished salivary flow, unique dentition, and hindered immune responses, leading to a relatively static environment.

In each of these cases we are basing our conclusions about aging on samples at an array of ages, but from different people, and so cannot account for interindividual inconsistencies. The ideal manner in which to explore aging in the microbiome would be extensive longitudinal studies. One study that I will cite with greater depth later during a discussion of microbiome simulations (see **section 6.3.2**) collected oral and stool microbiome samples from two individuals every day for most of a year (David et al. 2014). This was uniquely ambitious by today’s standards, as most longitudinal studies collect samples at far fewer timepoints from a larger group of individuals, but this design would be ideal for studying aging, so that we could see the true progression of microbial communities over time. Of course, speaking hypothetically, a study of this scale and resolution over the age range that we covered in SLL2 (ages 7-85) is not feasible for a variety of reasons. For one, we as a society, even just within the scientific community, probably would not have the patience to wait for the results and to continue to provide funding, even if those results were continually updated. While there are many examples of research which follows individuals across a significant portion of their lives, sample collections typically occur on the order of years, perhaps months, but not days, as I speculatively propose here. Thus, there would also be ethical considerations in expecting subjects to regularly provide samples

across their entire lifespan, and the participation would need to be wide to account for potential dropouts and early deaths. Moreover, there would be the danger not only of potential technical variation in processing the data over this time scale, but also of comparing samples processed with obsolete methods - I am likely writing this thesis in what will later be seen as primitive days of HTS and microbiome studies. In the few decades spanned by the modern incarnation of microbiome research, the field evolved far beyond what it was with rapid improvements in techniques and technologies, and there is no reason not to expect more of the same (see **section 6.3** for further speculation and extrapolation). This is largely hypothetical and ostensibly unrealistic at the moment, but the notion is undeniably appealing for microbiomists and only attainable if we continue to muse on these grander schemes.

6.1.4 SLL2 - The non-bacterial segment of the microbiome

At the beginning of this section I mentioned the dearth of microbiome studies that are not focused exclusively on bacteria. These studies are often very careful to account for potential confounding variables like medications taken by participants or smoking habits, yet they perpetuate an important oversight: the possibility for inter-domain interactions. We know that wide swaths of the tree of life are represented within the human microbiome, including many archaea (Koskinen et al. 2017), fungi and a variety of other unicellular eukaryotes, viruses, and some multicellular organisms, like helminths (Rowan-Nash et al. 2019). However, the exaggerated focus on bacteria in our field is not only because it dominates the biomass of the human microbiome, it is also the easiest to study with our current techniques. The internal transcribed spacer (ITS) region was accepted as the official marker gene for fungi (akin to the 16S rRNA gene in bacteria and archaea) by a consortium of mycologists (Schoch et al. 2012), but it is less universally applicable throughout that domain than 16S is among bacteria (Raja et al. 2017), and in general DNA extraction from fungi can be more complicated than that from bacteria because of

their thick cell walls (Ciardo et al. 2010). Sequencing of the 16S rRNA gene is viable for the archaeome, but primers have generally been designed for bacteria and often have been ineffectual for archaea. Nonetheless, protocols have been established that allow for proper study of archaea, and have found distinct compositions in different body sites (Koskinen et al. 2017). Given its inherent genetic diversity and lack of conserved genes, the virome does not have any universal marker genes like these other microbes, and viral genetic material is generally scant and difficult to detect, though there are improving techniques (Thurber et al. 2009; Allen et al. 2011). They are an important element of the microbiome; aside from the direct effects to human hosts and other microbes, viruses are likely involved in horizontal gene transfer events among prokaryotes, promoting virulence and antibiotic resistance within the bacteriome (Rowan-Nash et al. 2019).

Within Saca La Lengua, we only have fungal identification data for a segment of the dataset from the second edition. We aimed to include ITS sequencing data in both editions, but were unsuccessful in obtaining enough genetic material, for the reasons mentioned above, and because of the presence of amplicons of different sizes which complicated the process. In the end, we were able to culture fungal species from 1083 of the 1648 SLL2 samples, and used matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) analysis (Singhal et al. 2015) to identify the fungi. This provides a lower resolution of the fungal composition than the accompanying 16S sequencing for the bacteriome, as it does not give any information on fungal abundances, but nevertheless, we were able to glean some insights in each of the publications from the SLL2 dataset (**Chapters 3-5**). The relatively high prevalence of *Candida* species in both DS and CF samples supported the connections we drew with oral diseases and CF lung infections, respectively. Interestingly, in **Chapter 5** we saw that there were significant differences in the bacteriomes between samples in which fungal species were and were not detected, but it was the samples in which no fungi were detected that had the more homogeneous compositions. These samples also displayed some of the most unique co-occurrence networks among bacteria, compared to

variables like smoking and some chronic disorders. We posited that greater prevalence of fungi promotes a departure from the typical bacterial ecology in the oral cavity, though we could not be more specific based on our data.

6.2 Notes on Methodologies

Before most of the analyses were performed, we had our first experience with the technical biases to which high-throughput sequencing (HTS) projects can be susceptible. After sequencing the samples and processing the data, we noticed that a subset of the 1500 samples had considerably lower alpha diversity values than the rest. When backtracking, we discovered that these were all from a single sequencing run, for which the DNA extraction was performed in a different part of the lab under different conditions. In the end these precious samples were excluded from the dataset, so we learned the hard way that the procedures for a project like this must be rigid and standardized. There are various sources throughout the literature which suggest approaches to standardization at particular stages of microbiome analyses ("Raising Standards in Microbiome Research" 2016; Martin 2019; Amos et al. 2020), but there remains a lack of consensus. This is evident from the diversity of methodological decisions, like the use of different primers for targeting different variable regions of the 16S rRNA gene (Fouhy et al. 2016) or the choice of tool for sequence denoising and taxonomy assignment (Nearing et al. 2018; Prodan et al. 2020), not to mention unintentional or undetected technical variation that can occur within a given project, as we saw first-hand.

The current lack of consensus is due at least partially to continued efforts to improve and optimize the tools and techniques that microbiomists use (Gloor et al. 2017; Prodan et al. 2020), so it may simply be too early for the field to fully agree on the best practices. But it also may be exacerbating a number of inherent drawbacks to current analyses. Our example shows that even a small change in an early step of the sequencing procedure performed by the same

group can limit the comparability of samples within a given project, so the comparability between experiments performed by different groups attempting the same procedures could be unquantifiably dubious. There are platforms that aim to maximize reproducibility between projects or even within projects (Buza et al. 2019; Bai, Jhaney, and Wells 2019; Abdala Asbun et al. 2020; Murovec, Deutsch, and Stres 2021), but they cannot account for human error and technical variability in the sequencing process. There are statistical and experimental techniques that attempt to compensate for these issues, which I touched on briefly in **Chapter 2**. Most classification-based microbiome studies over the last decade have compared the relative abundances of taxa between samples, but this ignores the compositional nature of HTS experiments, which can have downstream statistical consequences. The most appropriate solution proposed thus far is to simply normalize the data using a transformation like the centered log-ratio, which is relatively robust to the depth of reads imposed by the sequencing instruments (Gloor et al. 2017). Others have also suggested methods to impute non-biological zero counts in the data (Kaul et al. 2017; R. Jiang, Li, and Li 2021). Essentially, they point to three different types of zeroes that might occur in HTS data: biological zeroes, which represent the true absence of an organism; technical zeroes, which occur due to experimental artefacts, and were the likely root of the issue we experienced in the first edition of SLL; and sampling zeroes, which occur due to the sequencing depth limitations. These methods attempt to distinguish the biological and non-biological zeroes, and to then correct the counts across the dataset. Even so, HTS data cannot reveal anything about the absolute abundances of microbes in a given sample because of the limited capacity for reads in the sequencing machines (hence the presence of sampling zeroes). Quantitative microbial profiling is an experimental technique that aims to account for the physiological relevance not only of differences in abundances of taxa, but of differences in total microbial load (Vandeputte et al. 2017). While this can account for the concerns of compositionality, it requires cell counting by flow cytometry (or other means) alongside sequencing, which is not commonly performed in microbiome studies.

These are issues that I came to understand over time and across multiple microbiome projects. The general ideas behind the analyses performed in the first and second editions of SLL were essentially the same: to classify the compositions of the oral microbiomes in particular contexts and search for correlative effects with various factors of health and lifestyle. But a closer comparison of the methods sections in **Chapter 2** versus those in **Chapters 3-5** will show a number of notable changes. In part this is because my path as a microbiomist has generally been autodidactic, as I expect it has for many others coming up in the early days of HTS-based microbiome studies. With standardizations of techniques and protocols in flux, there were certainly no formal education programs, so one had to learn from the trends present in the literature and practice with those technical tutorials that were available, building upon and honing analytical skills in parallel with the microbiome field itself. For instance, the first edition of SLL (**Chapter 2**), as well as a collaborative study of modulations of the skin microbiome (Paetzold et al. 2019), relied on the mothur platform (Schloss et al. 2009) to filter 16S rRNA gene sequences and assign them to operational taxonomic units (OTUs), which cluster the reads with a dissimilarity threshold of 97% by default. But after performing the analyses for those studies, we learned of tools like DADA2 (Benjamin J. Callahan et al. 2016), which use a more precise method to assign taxonomy based on amplicon sequence variants, accounting for sequencing errors to resolve single-nucleotide differences in reads (Benjamin J. Callahan, McMurdie, and Holmes 2017). It was not until another collaboration, one which examined the connections of oral bacteria to post-exercise hypotension and muscle oxygenation (Cutler et al. 2019), that I tested out and used the DADA2 platform, which I have used for all subsequent 16S-based microbiome studies.

The first edition of SLL also investigated the implications of stomatotypes among the samples, a notion conspicuously absent in **Chapters 3-5** (SLL2). The term stomatotype was a derivative of enterotype, coined in an early 16S-based study (Arumugam et al. 2011), merely substituting the relevant Greek roots (I also explored dermatotypes in our skin microbiome collaboration (Paetzold et al.

2019), and pneumotypes in another collaboration on the lung microbiome (Hérivaux et al. 2021)). It is meant to describe clusters of samples based on beta-diversity metrics which reveal general trends in the compositions of oral microbiome samples. As I described in **Chapter 1**, it is a concept that has received criticism for oversimplifying complex datasets, and which may better serve as guides for exploring the gradients of abundances of taxa across those datasets (Jeffery et al. 2012; Koren et al. 2013; Knights et al. 2014; Costea et al. 2018). I still believe that there is a practical use for these calculations as an early step in the analysis, especially if comparing the stomatotypes calculated with various distance metrics. Distinct clusters based on a Jaccard index, for instance, might reveal genuine patterns in rare or low abundance organisms, while a weighted UniFrac distance can show the effects of higher abundance organisms on the separation of samples while also accounting for phylogenetic distance between those organisms. This can be useful to the microbiomist as insights into the structure and tendencies inherent to a given dataset. But I have not continued to include this kind of analysis in the published results of studies, because while indeed it may be that there are a finite number of ubiquitous general conformations of the oral microbiome across human populations, the current state of analyses likely lacks the sophistication necessary to accurately simulate the processes and structures of the microbiome, which may have specious ramifications for the hopes of the microbiome as a tool of personalized medicine. I will further discuss the current limitations and potentialities of simulation in the microbiome in **section 6.3.3**. Instead of attempting to ascribe an artificial structure to the data as stomatotypes do, a converse approach may be more appropriate for describing trends; a calculation like the PERMANOVA (permutational multivariate analysis of variances) (M. J. Anderson 2001) can fit a multivariate model to a distance matrix to determine how the fixed effects in that model correspond to the separation of samples for the given distance calculation. This is a tool that I began to explore in two different collaborations, one exploring the connections of the oropharyngeal and stool microbiomes with obsessive-compulsive disorder (Domènech et al. 2020), and another

of the lung microbiome in pulmonary aspergillosis (Hérivaux et al. 2021). I carried this analysis over to the second edition of SLL (**Chapters 3-5**). An important element to this analysis was the ability to include multivariate models in order to simultaneously account for the effects of potential confounders like age and gender.

Arguably the most significant step in the progression of my analyses, and the impetus for this self-reflective section of the thesis, was the transition from simple calculations of the relative abundances of taxa to those of log-ratio transformations. Various methods of normalization have been proposed for the appropriate treatment of HTS-based microbiome data (Gloor et al. 2017; Morton et al. 2019; Lin and Peddada 2020). For the dataset from the second edition of SLL (**Chapters 3-5**), I began to use the centered log-ratio approach, which removes the bias that results from the unknown total microbial load in each sample by comparing the ratios of taxa between the samples in a dataset, and centering those values around zero to equally weight the relative differences (Gloor et al. 2017; Morton et al. 2019), and my more recent efforts with the SLL2 dataset have produced the most robust results of the studies in which I have been involved. The others, which have been referenced throughout this section of the thesis, not only compared relative abundances to determine differential abundances of taxa, but also compared global sample compositions with beta diversity metrics that are based on relative abundances, like the Bray-Curtis dissimilarity, Jensen-Shannon divergence and weighted UniFrac distance, or the presence and absence of taxa, like the Jaccard index and the unweighted UniFrac distance, all of which fall victim to the pitfalls of compositionally-unaware treatment of HTS data. As I mentioned in the previous paragraph, these calculations are useful for examining trends in the data from different angles, but I have relied primarily on the more compositionally sound Aitchison distance metric (Aitchison et al. 2000; Gloor et al. 2017) as a tool for reporting global trends in the dataset. Similarly, I constructed co-occurrence networks of taxa in **Chapter 2** based on Pearson correlations between genera, only to later understand that in a compositional dataset standard correlation measures, like Pearson or Spearman, have a bias towards negative correlations and are not

robust to subsetting of the data (Gloor et al. 2017). In **Chapters 4** and **5**, I instead use the SpiecEasi tool (Kurtz et al. 2015), which assumes a sparse data matrix (many zero counts, typical of rare and low abundance taxa in microbiome data) and uses a centered log-ratio transformation of the counts to infer robust co-occurrence networks. The datasets in each of the projects in which I have been involved have all been based on 16S sequencing, and generally report trends at the genus level. However, the 16S-based approach has been popular because it is a cheaper alternative to whole metagenome shotgun (WMS) sequencing, but as sequencing prices continue to decline (Wetterstrand 2020), researchers will more frequently opt for the greater taxonomic resolution offered by WMS (Brumfield et al. 2020; Durazzi et al. 2021). With this greater specificity will come wider sparsity throughout the data as particular strains will be detected in fewer individual samples, so proper treatment of this aspect of the data, as with SpiecEasi, will become increasingly important.

These evolutions in approaches to analysis may complicate the comparisons between many of my own studies, but I believe there is value in each of them beyond just their individual or collective results. They often have provided first glimpses of the microbiome in particular contexts, which should inspire further and deeper investigation, and the findings may be refined over time with improved data qualities and methods for analysis. As the field continues to develop, we can look back on what might be seen as antiquated methodologies to understand how they progressed and to better determine how to move forward.

6.3 The Being and the Becoming of Microbiomics

I started this discussion by describing the pre-HTS era as the infancy of microbiomics. Flipping through its history like the pages of a family album, we can see all its stages of growth. Early on, the field learned to express itself in crudely logical but naive ways, as Leeuwenhoek in 1676 described the bacteria, fungi and protozoa that he saw in his microscope as animalcules, meaning "tiny

animals”. The first vaccinations were introduced by Edward Jenner in 1796, despite some initial fears (Meynell 1995). We could see some early establishment of critical thinking skills, recognizing how processes in one’s environment work, like the discovery of the principles of fermentation in the late 1850s, and how to manipulate that environment to one’s benefit, like the development of pasteurization in 1865. Over time a youth recognizes dangers, develops rational fears and begins to find ways to cope with them, as with the espousal of the germ theory of disease in the 1880s and Robert Koch’s postulates on the causative relationships between microbes and diseases in 1884, even if those coping mechanisms need to be refined later on (Jindal 2018). We might relate the preadolescence of microbiomics with the advents of Sanger sequencing in 1977 (Sanger, Nicklen, and Coulson 1977) and the polymerase chain reaction in 1983 (Saiki et al. 1985), a period during which we could see that big changes were coming, that the field was about to hit a growth spurt and would never look the same again. And that is where we stand today, and precisely why I have called it the adolescence of the field of microbiomics. We are in a period of constant change and growth, probing at ways to best approach an array of challenges, uncertain and inconsistent at times (see section **6.2 Notes on methodologies**), but with a bright future and much still to learn.

6.3.1 Reductionism in Microbiome studies

Adolescents often have an as yet underdeveloped worldview because of limited experience with the complex systems at work about themselves. Such is the case even now for the field of microbiome studies. For instance, studies which attempt to represent the totality of the oral microbiome with one sample type are, by their nature (and by intentional design) reductionist, to which I alluded in **section 6.1**. This is a normal aspect of most scientific research, though it is important for these studies to

recognize the implications of the restrictions in their design. Many use only saliva to track differences between groups of samples, but this ignores minute niches, like the gingiva, or the surfaces of the teeth, gums, cheek, and tongue. One study, as an example, performed 16S sequencing of the saliva of children with and without severe early childhood caries (S-ECC), as well as a scraping of the deep dentinal plaque from the cavities in the teeth from children affected by S-ECC. Their results showed that, while the two saliva groups differed slightly from each other, both differed drastically from the caries-active cavity scraping samples (Hurley et al. 2019). In the methods section they cite a number of papers which justify the use of unstimulated saliva as "a representation of the whole oral ecosystem," yet their own results belie that assertion. I draw from this example not to suggest that their conclusions are invalid, but rather to point to the question of how best to obtain conclusions in a critical manner from each particular study design.

We necessarily derive a mereological debate from this issue, that of how to separate the whole of the microbiome from its parts. One of the earlier papers presented by the Human Microbiome Project (HMP) included samples from 18 different body sites, nine of which were in the oral cavity (Human Microbiome Project Consortium 2012). They found that, collectively, the oral cavity samples were readily distinguishable, based on Bray-Curtis dissimilarity values, from the stool, urogenital, skin, and nasal samples, and in fact were the group that most strongly separated from other body sites (Figure 1c of that paper). Each of the nine oral cavity sites could be identified by the combination of high abundances of *Streptococcus* and a few other genera, and similarly the groups of urogenital and skin samples were internally closely related, so we can rightly label the oral cavity a distinct part of the human microbiome as a whole. And yet, they describe subtle but detectable differences between the nine oral sites, showing that each of those is itself a distinct part of the whole microbiome. The term "oral ecosystem" is a useful one when designing a study, but here we can see that the tooth itself is an ecosystem, as is the tongue dorsum, and the keratinized gingiva, so that the oral cavity may be more akin to a biome with the human body as the biosphere. Of course, on the surface of a tooth with

dental caries, we might find a distinct ecosystem within the biofilm creating a cavity in the enamel which is distinct from the rest of the tooth, and on the tongue we might find distinct ecosystems among the different regions of gustatory cells and where amounts of air and moisture vary (analyses of the spatial organization of bacteria on the tongue already available (Wilbert, Mark Welch, and Borisy 2020)), so that if we continue in this manner it is ecosystems all the way down. This is why, to draw meaningful conclusions, reductionism in microbiomics is crucial, but it must be approached coherently. As described in **Chapter 1**, similarly to the paper about childhood caries that I referenced in the paragraph above, we designed the first edition of Saca La Lengua to include oral rinse samples because of previous works that also suggested that they offer a holistic depiction of the oral microbiome. But all of our samples were of the same type, obtained in the same manner, and so, while they may reduce the complexities of the oral ecosystem, they maintain an internal logic that provides comparable representations.

An exploration of an oral ecosystem is indeed the ultimate goal in a study like ours or the others that I mentioned above, but this exploration is limited, as by design they ignore some fundamental aspects of an ecosystem. An ecosystem is both biotic (represented by the composition of organisms) and abiotic (the nutrient and energy cycles), and is defined by how these two components interact (Odum 1971). An analysis of the taxonomic composition of a sample does not wholly address those systematic interactions, which is why I urged the adoption of multi-omics approaches in microbiome studies in **sections 6.1.2** and **6.2**, and do so again here. Practical budget concerns have of course restricted such an approach in most instances to date; given the options of 2000 samples with only 16S data or 20 samples with various omics data, most have leaned toward the former for the sake of greater statistical power, wider representation among populations, and more streamlined analyses. This is especially so since we have up to now been largely ensconced in an exploratory age that has been better served by broader views of the structure of the oral microbiome. Nonetheless, I expect that the field is becoming, and will continue to become, more interested in deeper examinations that will rely on

multiple omics datasets. Those studies that incorporate metabolomics, for example, begin to examine the system in which the microbes reside. The HMP paper that I reference above also included some functional analysis from metagenome sequencing data as well, which can point to the potential functionality of the identified organisms. The concern of relying solely on taxonomic classification is the potential for fallacious reification of results as the true totality of the microbiome. The notion of reification, also referred to as the fallacy of misplaced concreteness (Whitehead 1925), is often summarized with the aphorism “the map is not the territory” which rather evocatively illustrates the limitations in microbiomics. The field has generated a lot of data, and by amassing all of the details, we have begun to approximate an atlas of the human microbiome, replete with information and statistics on localized habits and preferences (via metatranscriptomic and metabolomic data, for instance), but even as the maps are continually refined, they themselves are not the microbiome. It seems obvious to say that there are always some elements missing from any HTS-based study, but a fundamental understanding of these holes is vital for interpreting results. There is aspiration for the manipulation of the microbiome in personalized medicine (Kashyap et al. 2017; Behrouzi, Nafari, and Siadat 2019; Pincelli et al. 2020; Cammarota et al. 2020), but the use of stomatotypes or enterotypes as diagnostic tools has been discouraged (Knights et al. 2014), for instance, precisely because they are reifications. In **Chapter 1**, I mention the functional niche shared by species of *Treponema* and methanogenic archaea in the oral cavity, each of which can act as “hydrogen sinks,” promoting the growth of secondary fermenters that exacerbate periodontitis. Stomatotypes are constructs of taxonomic classification, which might separate samples with differing abundances of *Treponema* and methanogens, though the functional impact may be equivalent. Thus, awareness of the limits of the map should inform the design of studies and their interpretations, and we should simultaneously strive to continually increase the complexity of our data so that the map better approximates the territory (see **section 6.3.3**).

6.3.2 Panta rhei

Panta rhei is a Greek phrase, meaning “everything flows,” attributed to the philosopher Heraclitus (Burnet 1930). It is the core tenet describing his doctrine of becoming, or the continual change that permeates and defines the world and everything in it, including the microbiome. And thus an ontological quandary stems from the reductionism in microbiomics. All microbiome studies to date have provided mere snapshots, whatever the omics discipline being applied. We intuitively understand the microbiome to be a dynamic system, yet our conclusions on its nature are derived from these static maps. The surfeit of information offered by HTS experiments over the last few decades, in comparison to what was offered in the preceding centuries, has vastly expanded our cartographic collections, though we should remain cognizant of their particular limitations. It is common for studies to propose oral microbiota as biomarkers for disease (Y. Lim et al. 2018; B. Chen et al. 2018; Chattopadhyay, Verma, and Panda 2019), and more advanced analyses, like those incorporating multiple omics techniques and analyzing the data with machine learning algorithms provide more robust predictions of biomarkers (Pasolli et al. 2016; B. Chen et al. 2018; Cammarota et al. 2020; Marcos-Zambrano et al. 2021). Yet they still represent only the “being” of the microbiome, and not its “becoming.” This, of course, is an issue common to most scientific research, but in the context of the microbiome, there are clear paths toward potential improvements.

Nonetheless, it is not a trivial distinction. Microbiome studies as they are usually structured today, whether they explore taxonomic composition, genomic or transcriptomic productivity, metabolic activity, or any combination of these or other techniques, do not discern the ephemeral from the perpetual. It is not to say that the field at large rejects the “becoming” of the microbiome (its perpetual change in the Heraclitean sense) in favor of a static “being” (an eternal and unified oneness in the Eleatic sense), but rather that the field today cannot readily represent its flow. A metabolomics experiment, for instance, may display an abundance of a particular metabolite in a sample that resulted from a brief

surge in activity by certain microbes after the human host tried a new topping on their pizza for the first time the previous night, and this source will likely go undetected by the microbiomists, as will the eventual return to average levels of that metabolite. Longitudinal studies that collect samples from the same individuals and sites over time are more informative in this sense, though they are still a limited series of snapshots.

One example that I briefly mentioned in **section 6.1.3** collected oral and stool samples from two individuals every day for most of a year (David et al. 2014). One traveled from the U.S. to southeast Asia for 51 days, during which time he experienced a two-fold increase in the Bacteroidetes to Firmicutes ratio. The other had *Salmonella* food poisoning for eight days, during which the most prominent taxa declined and were replaced by phylogenetically close competitors, and his microbiome remained in this transformed state for the final months of sample collections. In both individuals, there was apparent stability in the composition outside of the periods of disturbance, though in the former case the composition reverted to its original structure when the individual returned to the U.S. and in the latter it adopted a new structure. With just a single sample from an individual, we cannot determine whether the composition that we observe is in the process of a state transition or one of these stable states, and if so whether it had previously been in a different state. This similarly muddles the notion of eubiosis and how we might distinguish it from dysbiosis. In the example of the second individual, an apparent stable state is reached after infection, but it is largely constituted by taxa that thrived during the infection. Perhaps the composition during the time of the infection could be considered dysbiosis, but if it is similar afterward, when the individual has returned to relative health, should it still be seen as dysbiosis relative to the initial state? Is it a separate eubiosis? Where then is the border between eubiosis and dysbiosis? The literature continues to explore the ways in which we should understand multiple stable compositions (Fukami and Nakajima 2011; Costello et al. 2012; Amor, Ratzke, and Gore 2020; Van de Guchte et al. 2020), and it is still unclear what the implications may be in many contexts.

Another study collected samples from individuals every three hours (during waking hours) for three days and detected patterns of periodicity throughout the day in the abundances and diversity in the oral microbiome (Sarkar et al. 2021). This gives evidence of rapid turnover of microbes, further weakening the validity of the solitary snapshots that constitute most of the literature. Ideally, longitudinal study designs like these would be more widely adopted and amplified to the magnitude of sample sizes seen in most studies, which would indeed approach a representation of the perpetual conformation of the microbiome. However, it also raises the question of how we might quantify the impacts of this rapid turnover. There is an apt analogy that can help to conceptualize this issue and begin to decide how to confront it. Zeno, one of the Eleatic philosophers of pre-Socratic Greece, describes a supposed paradox about grains of millet. He noted that when a single grain falls, it apparently makes no sound, but 1000 grains do make a sound, and so the paradox is that 1000 nothings make something. Zeno and the Eleatics based their writings on the premise that human perception is fallible and therefore cannot be wholly trusted, but Aristotle later reasoned that imperceptible sounds can accumulate to surpass a threshold that is perceptible (Huggett 2019). In much the same way, both the human body and HTS experiments are subject to some threshold of detection of the turnover of microbes. While one's body may not react noticeably to the death of a single bacterium, there is undoubtedly some quantifiable impact as each cell is an inextricable piece of the microbiome and its host. To continue the reimagined quote of John Donne from the preface of this thesis: "any microbe's death diminishes me, because I am involved in the microbiome." Of course, the body and its reaction is one level of detection of turnover, but to study this empirically we rely on techniques like sequencing, and thus must consider their thresholds as well. As I discussed in **section 6.2**, the compositional nature of HTS experiments ensures that some amount of the genetic material will not be sequenced, and before being analyzed, more will be filtered out because of low quality sequencing reads. Thus, as we move forward the field will need to define, and continually refine,

standard thresholds for such impacts in a variety of contexts, like samples from different body sites and for different populations, demographics, or disease states. These will need to be especially accurate in fine-scale longitudinal studies, like the one referenced above which collected multiple samples per day, where the magnitude of changes will also be acute. It would also be useful to consider factors like the total microbial load using techniques like quantitative microbiome profiling (Vandeputte et al. 2017), which can further indicate changes in the microbial ecosystem. In short, what we will need are increasingly complex, multidimensional, dynamic representations that optimally simulate the microbiome and its systems.

6.3.3 The Evolving Microbiome Simulacrum

We understand microbiome studies today to be fairly straightforward images of the processes within a sampling site, but in reality they are many-layered abstractions of the true processes, so it is important to take an occasional moment to consider what this means for our analyses, and how we might continue to move toward more wholistic simulations of the microbiome. Taxonomic classification studies (16S and WMS sequencing) are intended to tell us exactly what is present in a sample, but there is always information loss at various steps. The collection is a sampling of the total microbial load, and from that collection not all genetic material will be extracted, and from that extraction some amount of the sequences will be of low quality and later discarded, and from that sequencing information, some amount is typically filtered, and from what is kept, there is often an amount of “microbial dark matter,” those organisms that are unclassified and unstudied (Rinke et al. 2013), so that knowing that they are present does not necessarily inform us much about what they might do. Functional analysis studies (metatranscriptomics, metaproteomics, metabolomics, and sometimes WMS sequencing which can reveal functional potential) present a different level of information than taxonomic studies since they tell us about what is happening or may happen, the actions taken by the microbes, or what has happened by revealing the byproducts in the case of metabolomics. Still, these show only a

moment in time, vulnerable to the same losses in information, metatranscriptomics again in relation to sequencing, while proteomics and metabolomics are subject to the estimates of nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry. Moreover, metabolomics studies can have difficulty in definitively identifying which organisms produce each molecule (Edlund et al. 2017); instead we are left to deduce these connections as best as we can.

These techniques, which in a short interval have allowed for immense growth in our awareness of the ecosystems present within the human body, are the best that we have at the moment, and their methods are continuously improving and building upon the findings that they provide. Single-cell genomics of the microbiome, for instance, avoids the uncertainties presented by samples with high levels of heterogeneity among strains of the same species, because it inspects one cell at a time, and has been used to sequence and identify some of the microbial dark matter that was as yet unclassified in bacteria and archaea (Rinke 2018), and in fungi (Ahrendt et al. 2018), as well as to explore bacteria, protists and viruses together (Yoon et al. 2011). Longitudinal studies, like those I mentioned in **sections 6.1.3** and **6.3.2**, are important steps toward lowering the abstractions of the true microbiome processes that are restricted by modern experiments. Yet these are still merely a series of time slices, the separations of which are like the framerate of a camera imposing constraints on the audience's perception of the totality of events. None of the techniques that we use today can perfectly perform the job we are ultimately trying to perform: the production of a simulacrum of the microbiome *in silico* that optimally resembles the microbiome *in vivo* so that we may examine its natural behaviors. As in any realm of investigation, this is merely a conjectural notion that can be useful for setting goals toward advancements, and in the field of microbiomics there are some concrete approaches that we can begin to imagine based on today's standards. Eventually it may be more appropriate to use the more specific term *in simulacra* to label such studies as our approximations improve. For this we would need a way to hypothetically focus a sort of high-definition camera on a site, such

as the surface of the tongue, that can zoom in to see what organisms are present, and in what numbers, and to track their turnover, examine their internal processes, like transcription of mRNA and protein synthesis, and also to quantify the ingestion and outputs of each cell, and weigh the strengths of adhesions to other cells and to the biofilm matrices that they produce. There is also the potential of epigenetic modification of microbes from the action of the human host cells (Beaulaurier, Schadt, and Fang 2019; Morovic and Budinoff 2021), or of human host cells from the action of resident microbes (Oelschlaeger 2010; Celluzzi and Masotti 2016).

The need for effective modelling of the microbiome has been discussed for at least the last decade (Borenstein 2012), and some complex tools have been developed which incorporate multiple omics data (M. Kumar et al. 2019; Popp and Centler 2020), and even *in vitro* simulacra of the human colon that model metabolic activities of microbes (Martínez-Cuesta, Peláez, and Requena 2019). However, many of the bioinformatics tools have so far been focused on modeling only the taxonomic structures (Fritz et al. 2019; Baldini et al. 2019; Rong et al. 2021). All of these tools can be insightful, but they have not yet been widely implemented throughout the literature to derive conclusive and consistent results, and I believe we should expect to see increasingly elaborate simulations as technologies and knowledge advance. The hope for manipulation of the microbiome in the field of personalized medicine may in many situations depend on the interpretations of such simulations, and we are considering a system with tens of trillions of moving parts (Sender, Fuchs, and Milo 2016), so it is vital that we maximize their accuracy. What follows is largely conjecture on how new simulacra may develop and some potential hazards that may arise from their interpretations. I will rely on science-adjacent analogies to the works of a few more of history’s great thinkers to clarify these points in the context of the oral microbiome.

In his dialogue *The Sophist*, Plato describes two kinds of simulacra: one that is a precise duplicate of the original, and another that is deliberately altered so that it appears like a duplicate to a viewer

whose perception may be hindered, which he explains like this: "in works either of sculpture or of painting, which are of any magnitude, there is a certain degree of deception; for artists were to give the true proportions of their fair works, the upper part, which is farther off, would appear to be out of proportion in comparison with the lower, which is nearer; and so they give up the truth in their images and make only the proportions which appear to be beautiful, disregarding the real ones" (Jowett 1871). We strive to make the first type, but technological constraints force that latter upon us when we try to correct for them. This is often done in the early bioinformatic stages of 16S sequencing studies by filtering low quality sequencing reads (Benjamin J. Callahan et al. 2016), or removing sequences whose abundances are low and variable and so are potentially erroneous (McMurdie and Holmes 2013), or adding pseudo-counts to account for non-biological zeroes in the data (Mandal et al. 2015). As Plato describes, microbiomists rely on technologies which are not infallible, and so must make informed adjustments to the data to approximate its true composition.

When striving to develop the former of Plato's simulacra, one which is an exact double of an individual's oral microbiome in our case, we should also remember the maxim quoted in **section 6.3.1** that "the map is not the territory." There are two main concerns in advancing microbiome simulations which should come with a fundamental awareness of this notion: (1) the difficulties in dealing with data overload, and (2) the unintentional confusion of the simulation with reality. Both of these issues can be exemplified by a short story (only a single paragraph) by Jorge Luis Borges called "Del rigor en la ciencia" ("On Exactitude in Science" in English) (Borges and Hurley 1999). It is about an empire whose cartographers perfect their science to the point that a map of the empire can only be deemed adequate if it is plotted at the same scale as the empire itself. Obviously this is excessive and would be cumbersome, so the future generations of the empire were put off by the giant map, and left it to wither in the elements. This is the first of the concerns with improving microbiome simulations. The map of the locations of bacterial species on the surface of the tongue that I referenced in **section 6.3.1** (Wilbert, Mark Welch, and

Borisy 2020) is a rudimentary parallel to Borges' map, which charts the bacterial taxonomy of the tongue based on 16S rRNA sequences. We will continue to produce larger and more complex datasets, so bioinformaticians will need to remain a step ahead, an issue that is discussed regularly in the literature (M. Baker 2010; Yang, Troup, and Ho 2017; Talia 2019). We need to consider not only processing power and scalability of software, but also the continuing development of statistical techniques that are appropriate for massive and complex datasets (D. R. Cox, Kartsonaki, and Keogh 2018). The high degree of abstraction in modern microbiome studies make their analyses manageable, but with greater complexity will come greater potential for confounded results, so we must also efficiently interpret them.

So our current microbiome studies actively create something like Plato's simulacrum in the form of the disproportionate statue, though we aim to move toward his other form of simulacrum, the precise depiction. With this attempt, we risk the second concern that I mentioned above, the misplaced equivalency between the simulacrum and the true microbiome. This may seem like a redundant argument, but it presents an important consideration for the field as a whole. As our technologies and techniques improve, we will asymptotically approach something more and more like Borges' map, perfectly and fully simulating the microbiome and its processes. But this may also increasingly limit our ability to recognize any imperfections in the simulation. The bioinformatic tool mbImpute (R. Jiang, Li, and Li 2021), for instance, posits that it can impute non-biological zeroes in microbiome data by comparing to a combination of similar taxa (close phylogenetic relatives with similar counts), similar samples in the dataset (in terms of total counts), and similar samples in the metadata (e.g. age and gender). Compared to a less advanced and comprehensive approach to imputation, like replacing all zeroes with the same pseudo-count value, this is less intuitive to the human observer and more reliant on a calculated simulation of the data. This may provide a more accurate estimation of the compositions than simple pseudo-counts, but the increased abstraction from the microbiomist's perspective also obscures sources of error to a greater degree, and thus obscures

the boundaries between real and simulated compositions, moving them more toward hyperreal representations. This is a simplified example of the notion, relevant for current sequencing-based studies (16S, WMS), but if we hope to harness the microbiome for the purposes of personalized medicine, as so many publications strive toward, we will rely on more accurate representations of the microbiome and its processes, which will necessarily incorporate multiple omics data and relevant metadata (information particular to the individual and the local environment, e.g. pH and salivary flow rate). Alongside improved simulacra must come an awareness of potential inaccuracies and efforts to vigilantly detect and limit them, which will largely fall on the shoulders of the bioinformaticians and statisticians (D. R. Cox, Kartsonaki, and Keogh 2018). The developers will generally be those with the greatest awareness of their tools' limitations and sources of error, but many, perhaps most, of the users may have weaker technical and computational backgrounds, or at least be less aware of the internal workings of those tools, and so will generally have no choice but the trust in the validity of the outputs. It is beyond my purview to describe the technologies that will build upon our current sequencing and chemical fingerprinting techniques, or the types of analyses that will come with them, but surely they will have their own limitations as well. Considering all of this, not only is there the potential for reification of the model as the true entirety of the microbiome, but unawareness of the possible errors can also threaten the validity of human interpretations.

With this section of the thesis, I merely hope to illustrate the potential advances and pitfalls in microbiomics that I foresee as fully as I am able, abstract though they may be. Indeed, I am very hopeful for the potential of this field and am proud to have played some small role in the early stages of what is already an explosive growth in its capabilities and knowledge base. In the analogy of Plato's statue as a simulacrum, my work from the two editions of SLL (**Chapters 2-5**) may be more akin to a bust, or really just a carving of the mouth, but it serves as an important jumping off point for further explorations and estimations of the oral microbiome in the Spanish population. While we have examined it only from the

perspective of its taxonomic classification, we have done so from a variety of angles that are foundational in structuring the oral microbiome's composition, particularly age, health, and lifestyle, and this is an equally important element of the construction of simulations.

6.4 The conceit of SLL

I opened the preface of this thesis with a quote from John Donne, an English metaphysical poet, part of a literary movement whose focus I believe should be, and often already is, internalized by microbiomists, at least as our field stands today. The metaphysical poets relied on the use of conceits to develop their social commentaries, a process which has been described like so: "a comparison becomes a conceit when we are made to concede likeness while being strongly conscious of unlikeness" (Gardner 1961). To help ourselves conceptualize the nature of the microbiome, we may concede that it is very much like human society, despite the obvious unlikenesses. Microbes, like humans, live in close proximity to others very much like themselves, though the complexity of the interindividual differences are magnitudes apart. Microbes, like humans, live in a variety of habitats with distinct environmental conditions and have come to colonize nearly every accessible niche of their host, though the ranges of adaptability are hardly comparable and operate on different scales. Microbes, like humans, can affect their neighbors both beneficially and adversely with their habits, as well as at various societal levels, individually or communally, locally or systemically, though the respective social structures arise through very different selective pressures. It is difficult to avoid that part of the human condition which compels us to hold a mirror up to anything we explore so that we might find our own reflections within it, and we may as well exploit that propensity to continue to build upon microbiome studies.

To that end, the work of this thesis, which primarily comprises the two editions of SLL, has added an important resource to the community of microbiome researchers. Our analyses provide the

first glimpses of the oral microbiome within the Spanish population at large. By comparisons with other findings throughout the literature, we have argued that our data may serve as a generalized proxy for trends in the oral microbiomes across most of the so-called "WEIRD" nations, as there are wide-scale patterns which distinguish oral microbiomes among populations with significantly different customs, diets, and socioeconomic parameters (Clemente et al. 2015; Pasolli et al. 2019; Handsley-Davis et al. 2020). But more directly, our work operated on a smaller scale (despite the nearly 3000 samples from the two editions of SLL together), focusing on individual trends within the oral microbiome throughout Spain, from which we were able to develop a map of its composition. Projects like ours allow microbiomists to finally start to envision the "societal" structures within the oral microbiome, and we have begun to annotate its map with findings on its internal interactions and its associations with various factors of health, aging, and lifestyle. As I have discussed at large in this chapter, modern technological limitations necessitate a distinction between the map and the territory it aims to chart, so we must move forward with a certain caution when drawing conclusions, and should continue to be critical of the quality of both the data being produced and its analyses. Our work from the two editions of SLL will be indispensable as we begin to synthesize the accumulating snapshots of microbiomes in specific contexts, which should allow us to move toward a systematic understanding of the entire human microbiome, particularly in western industrialized populations, but eventually in as many contexts as we can imagine. As techniques and technologies continue to develop, we will better be able to quantify the contributions by each microbe, and our approach to microbiome studies will better be able to apply a version of Donne's conceit stressing the need to recognize the worth of each individual to the whole, and to every other individual.

Conclusions

- The two editions of Saca La Lengua (SLL) used a total of nearly 3000 oral rinse samples, analyzed with 16S rRNA gene sequencing to produce an initial map of the composition of the oral microbiome throughout the Spanish population.
- The first edition of SLL showed that there are definitive correlative relationships between the ionic composition of drinking water and the structure of the oral microbiome.
- It also pointed to potentially ubiquitous generalized conformations of the oral microbiome, which have been corroborated to varying extents in other studies.
- Lifestyle factors, including diet and hygiene, also displayed associations with specific taxa within the oral microbiome.
- The second edition of SLL explored more varied segments of the Spanish population, including individuals with particular chronic disorders, including Down Syndrome (DS) and cystic fibrosis (CF). We provided early snapshots of the oral microbiome in the contexts of these disorders, and showed that they have strong implications in the balance between oral health and disease.
- DS appeared to present with a relatively static oral microbial environment across age, perhaps due to the particular physiological conditions inherent to the disorder, including low salivary flow, unique dentition, and poor immune responses.
- We saw evidence for the oral cavity as a potential reservoir of microorganisms associated with CF lung infections.
- We found evidence of the associations of fungal species,

particular *Candida*, in both DS and CF oral microbiomes.

- The second edition of SLL also examined the oral microbiome in a wide range of ages, which showed that there tends to be greater relative stability in the composition in the middle ages as compared to youths and seniors, potentially because of continued development and declining immunity, respectively.

- We found that chronic disorders had greater relative impacts on the overall composition of the oral microbiome as compared to other lifestyle factors, but also important were smoking and the presence or absence of yeasts.

- The second edition of SLL also expanded upon findings that shared environments are important in shaping the oral microbiome, in that closely related family members, and even students from the same class, tended to have more similar compositions compared to non-relatives and non-classmates, respectively.

Appendix I: Drinking water impacts the oral microbiome composition

As in the first edition of SLL (**Chapter 2**), we found significant associations between individual taxa and the concentrations of a number of ions in drinking water, and many of these associations were corroborated in our results from the second edition of SLL (**Chapter 5** - not included in the results submitted for publication, but presented here). For this analysis, we included only those samples without chronic disorders and who reported that they did not primarily drink bottled water (n = 841). The other sources were filtered tap water, unfiltered tap water, and untreated water (from a fountain, well, or river), and here, we selected subsamples of those 841 samples 100 times, balancing the distributions of those three sources of drinking water, age, geographical location, and gender. The strongest associations were negative trends between the genera *Acinetobacter*, *Variovorax*, *Mesorhizobium*, and *Pseudomonas* with nearly all of the ionic compounds and measurements, particularly magnesium (Mg), the amount of dry residue after boiling the water, alkalinity, bicarbonate (HCO_3), conductivity, water hardness, sulfate (SO_4), and calcium (Ca). A number of genera, including *Delftia*, an unclassified genus of the family Clostridiales_vadinBB60_group, and *Filifactor*, displayed positive associations with many of these same measurements, though the significance of the positive associations was typically lower than that of most of the negative associations. These and the other significant associations are displayed in **Figure A.1**.

Furthermore, using the 841 samples from this water analysis, we calculated stomatotypes, which are statistical clusters of the samples based on their oral microbiome compositions. We found in the first

edition of SLL (**Chapter 2**) that many of the most significantly differentiated organisms between the two stomatotypes were also significantly associated with many of these water values. In this study, we used the Aitchison distance metric to calculate the stomatotypes (see **Chapter 5** - Materials and Methods), and again we found two clusters. The genera which most significantly drive the composition of the first stomatotype (determined by between-class analysis) all displayed negative associations with many water values (**Table A.1, Figure A.1**). However, the driver genera for the second stomatotype did not display significant positive associations with the water values, with the exception of *Peptostreptococcus*, which was the strongest driver and was associated with the amount of dry residue after boiling, alkalinity, HCO_3, and nitrate (NO_3), though the significance of the positive associations was typically of a lower magnitude than that of the negative associations seen in the driver genera of the first stomatotype.

Table A.1: The 10 genera with the greatest individual impacts on the clustering of samples into each of the two stomatotypes. Values are the relative weights as determined by between-class analysis for each stomatotype, which are calculated using the Aitchison distance metric. ANPR = *Allorhizobium-Neorhizobium-Pararhizobium- Rhizobium*

Genus	Stomatotype 1 Weights	Genus	Stomatotype 2 Weights
Variovorax	0.504	*Peptostreptococcus*	0.419
Acinetobacter	0.504	*Solobacterium*	0.394
Pseudomonas	0.495	*Lachnoanaerobaculum*	0.393
Mesorhizobium	0.453	*Alloprevotella*	0.379
ANPR	0.441	*Prevotella*	0.369

Ralstonia	0.412	*Ruminococcaceae_ UCG-014*	0.330
Curvibacter	0.355	*Stomatobaculum*	0.313
Brevundimonas	0.340	*Johnsonella*	0.310
Bradyrhizobium	0.311	*Leptotrichia*	0.304
Hyphomicrobium	0.260	*Peptococcus*	0.303

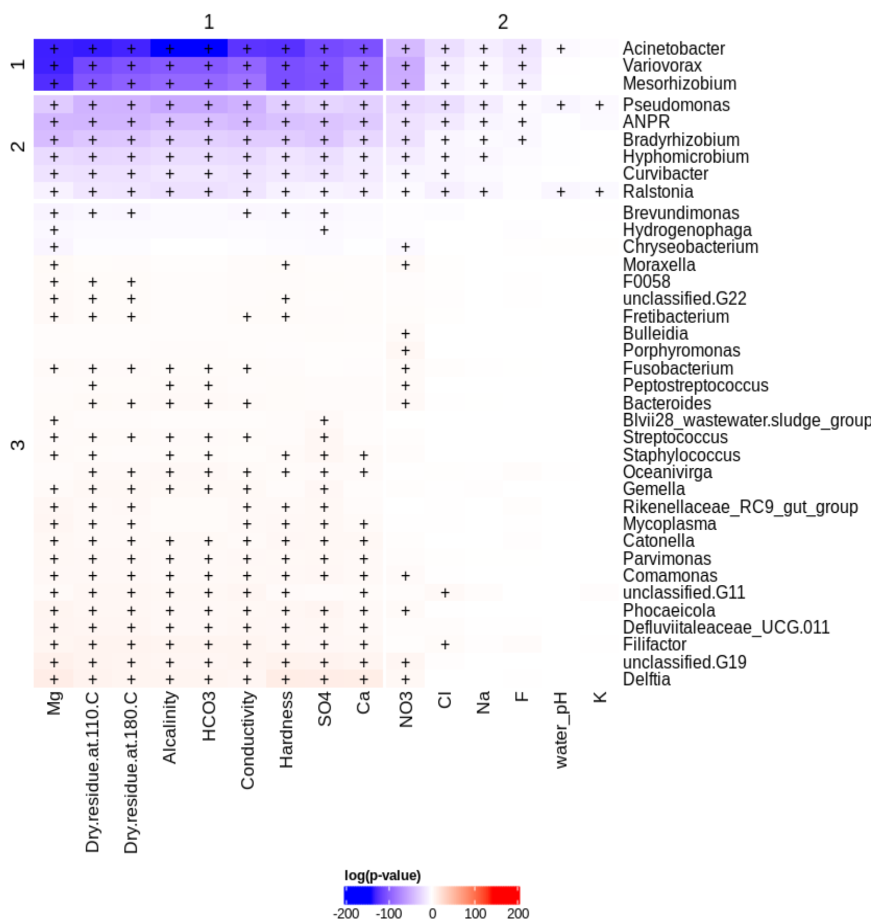

Figure A.1: Effects of drinking water on taxa abundances. Heatmap indicates

the log of the mean adjusted p-values for a given association across 100 subsamples. Negative values in blue indicate that there was a negative association between that genus and water value, while positive values in red indicate a positive association. Associations that were significant on average across the 100 subsamples (mean adjusted $P < 0.05$) are marked with a “+”. ANPR = *Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium*.

Bibliography

Aas, Jørn A., Ann L. Griffen, Sara R. Dardis, Alice M. Lee, Ingar Olsen, Floyd E. Dewhirst, Eugene J. Leys, and Bruce J. Paster. 2008. "Bacteria of Dental Caries in Primary and Permanent Teeth in Children and Young Adults." *Journal of Clinical Microbiology* 46 (4): 1407–17.

Abdala Asbun, Alejandro, Marc A. Besseling, Sergio Balzano, Judith D. L. van Bleijswijk, Harry J. Witte, Laura Villanueva, and Julia C. Engelmann. 2020. "Cascabel: A Scalable and Versatile Amplicon Sequence Data Analysis Pipeline Delivering Reproducible and Documented Results." *Frontiers in Genetics* 11 (November): 489357.

Abnet, Christian C., You-Lin Qiao, Sanford M. Dawsey, Zhi-Wei Dong, Philip R. Taylor, and Steven D. Mark. 2005. "Tooth Loss Is Associated with Increased Risk of Total Death and Death from Upper Gastrointestinal Cancer, Heart Disease, and Stroke in a Chinese Population-Based Cohort." *International Journal of Epidemiology* 34 (2): 467–74.

Abusleme, Loreto, Amanda K. Dupuy, Nicolás Dutzan, Nora Silva, Joseph A. Burleson, Linda D. Strausbaugh, Jorge Gamonal, and Patricia I. Diaz. 2013. "The Subgingival Microbiome in Health and Periodontitis and Its Relationship with Community Biomass and Inflammation." *The ISME Journal* 7 (5): 1016–25.

Agholme, M. B., G. Dahllöf, and T. Modéer. 1999. "Changes of Periodontal Status in Patients with Down Syndrome during a 7-Year Period." *European Journal of Oral Sciences* 107 (2): 82–88.

Aguayo, Sebastian, Christina Maria Anna Pia Schuh, Benjamin Vicente, and Luis Gerardo Aguayo. 2018. "Association between Alzheimer's Disease and Oral and Gut Microbiota: Are Pore Forming Proteins the Missing Link?" *Journal of Alzheimer's Disease: JAD* 65 (1): 29–46.

Aguiar-Pulido, Vanessa, Wenrui Huang, Victoria Suarez-Ulloa, Trevor Cickovski, Kalai Mathee, and Giri Narasimhan. 2016. "Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis." *Evolutionary Bioinformatics Online* 12 (Suppl 1): 5–16.

Aguirre de Cárcer, Daniel. 2018. “The Human Gut Pan-Microbiome Presents a Compositional Core Formed by Discrete Phylogenetic Units.” *Scientific Reports* 8 (1): 14069.
Ahrendt, Steven R., C. Alisha Quandt, Doina Ciobanu, Alicia Clum, Asaf Salamov, Bill Andreopoulos, Jan-Fang Cheng, et al. 2018. “Leveraging Single-Cell Genomics to Expand the Fungal Tree of Life.” *Nature Microbiology* 3 (12): 1417–28.
Ai, Dongmei, Ruocheng Huang, Jin Wen, Chao Li, Jiangping Zhu, and Li Charlie Xia. 2017. “Integrated Metagenomic Data Analysis Demonstrates That a Loss of Diversity in Oral Microbiota Is Associated with Periodontitis.” *BMC Genomics* 18 (Suppl 1): 1041.
Aitchison, J., C. Barceló-Vidal, J. A. Martín-Fernández, and V. Pawlowsky-Glahn. 2000. “Logratio Analysis and Compositional Distance.” *Mathematical Geology* 32 (3): 271–75.
Ajonuma, Louis, Qian Lu, Becky P. K. Cheung, W. Keung Leung, Lakshman P. Samaranayake, and Lijian Jin. 2010. “Expression and Localization of Cystic Fibrosis Transmembrane Conductance Regulator in Human Gingiva.” *Cell Biology International* 34 (2): 147–52.
Akiyama, K., N. Taniyasu, J. Hirota, Y. Iba, and K. Maisawa. 2001. “Recurrent Aortic Valve Endocarditis Caused by Gemella Morbillorum--Report of a Case and Review of the Literature.” *Japanese Circulation Journal* 65 (11): 997–1000.
Al-Ahmad, A., T. M. Auschill, R. Dakhel, A. Wittmer, K. Pelz, C. Heumann, E. Hellwig, and N. B. Arweiler. 2016. “Prevalence of Candida Albicans and Candida Dubliniensis in Caries-Free and Caries-Active Children in Relation to the Oral Microbiota-a Clinical Study.” *Clinical Oral Investigations* 20 (8): 1963–71.
Ali, Javed, K. Pramod, M. Abu Tahir, and S. H. Ansari. 2011. “Autoimmune Responses in Periodontal Diseases.” *Autoimmunity Reviews* 10 (7): 426–31.
Al-Jebouri, Mohemid. 2016. “The Relationship between Periodontal Disease and Predisposing Factors.” *Tikrit Journal for Dental Sciences* 4 (1): 68–80.
Allen, Lisa Zeigler, Thomas Ishoey, Mark A. Novotny, Jeffrey S. McLean, Roger S. Lasken, and Shannon J. Williamson. 2011. “Single Virus Genomics: A New Tool for Virus Discovery.” *PloS One* 6 (3): e17722.
Al-Mulla, Anas, Lena Karlsson, Saad Kharsa, Heidrun Kjellberg, and Dowen Birkhed. 2010. “Combination of High-Fluoride

Toothpaste and No Post-Brushing Water Rinsing on Enamel Demineralization Using an in-Situ Caries Model with Orthodontic Bands." *Acta Odontologica Scandinavica* 68 (6): 323–28.

Al Soub, Hussam, Sittana S. El-Shafie, Abdel-Latif M. Al-Khal, and Amar M. Salam. 2003. "Gemella Morbillorum Endocarditis." *Saudi Medical Journal* 24 (10): 1135–37.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.

Amano, A., T. Kishima, S. Akiyama, I. Nakagawa, S. Hamada, and I. Morisaki. 2001. "Relationship of Periodontopathic Bacteria with Early-Onset Periodontitis in Down's Syndrome." *Journal of Periodontology* 72 (3): 368–73.

Amano, Atsuo, Jumpei Murakami, Shigehisa Akiyama, and Ichijiro Morisaki. 2008. "Etiologic Factors of Early-Onset Periodontal Disease in Down Syndrome." *The Japanese Dental Science Review* 44 (2): 118–27.

Amir, Amnon, Daniel McDonald, Jose A. Navas-Molina, Evguenia Kopylova, James T. Morton, Zhenjiang Zech Xu, Eric P. Kightley, et al. 2017. "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns." *mSystems* 2 (2). https://doi.org/10.1128/mSystems.00191-16.

Amor, Daniel R., Christoph Ratzke, and Jeff Gore. 2020. "Transient Invaders Can Induce Shifts between Alternative Stable States of Microbial Communities." *Science Advances* 6 (8): eaay8676.

Amos, Gregory C. A., Alastair Logan, Saba Anwar, Martin Fritzsche, Ryan Mate, Thomas Bleazard, and Sjoerd Rijpkema. 2020. "Developing Standards for the Microbiome Field." *Microbiome* 8 (1): 98.

Anderson, Marti J. 2001. "A New Method for Non-Parametric Multivariate Analysis of Variance." *Austral Ecology* 26 (1): 32–46.

Anderson, S. 1981. "Shotgun DNA Sequencing Using Cloned DNase I-Generated Fragments." *Nucleic Acids Research* 9 (13): 3015–27.

Andrews, Simon. 2010. "FastQC A Quality Control Tool for High Throughput Sequence Data." Babraham Bioinformatics. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anesti, Vasiliki, Ian R. McDonald, Meghna Ramaswamy, William G. Wade, Donovan P. Kelly, and Ann P. Wood. 2005. "Isolation

and Molecular Detection of Methylotrophic Bacteria Occurring in the Human Mouth.” *Environmental Microbiology* 7 (8): 1227–38.

Areias, Cristina, Benedita Sampaio-Maia, Maria de Lurdes Pereira, Alvaro Azevedo, Paulo Melo, Casimiro Andrade, and Crispian Scully. 2012. “Reduced Salivary Flow and Colonization by Mutans Streptococci in Children with Down Syndrome.” *Clinics* 67 (9): 1007–11.

Arquitt, C. K., C. Boyd, and J. T. Wright. 2002. “Cystic Fibrosis Transmembrane Regulator Gene (CFTR) Is Associated with Abnormal Enamel Formation.” *Journal of Dental Research* 81 (7): 492–96.

Arumugam, Manimozhiyan, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R. Mende, Gabriel R. Fernandes, et al. 2011. “Enterotypes of the Human Gut Microbiome.” *Nature* 473 (7346): 174–80.

Asakawa, Mikari, Toru Takeshita, Michiko Furuta, Shinya Kageyama, Kenji Takeuchi, Jun Hata, Toshiharu Ninomiya, and Yoshihisa Yamashita. 2018. “Tongue Microbiota and Oral Health Status in Community-Dwelling Elderly Adults.” *mSphere* 3 (4). https://doi.org/10.1128/mSphere.00332-18.

Aßhauer, Kathrin P., Bernd Wemheuer, Rolf Daniel, and Peter Meinicke. 2015. “Tax4Fun: Predicting Functional Profiles from Metagenomic 16S rRNA Data.” *Bioinformatics* 31 (17): 2882–84.

Bai, Jinbing, Ileen Jhaney, and Jessica Wells. 2019. “Developing a Reproducible Microbiome Data Analysis Pipeline Using the Amazon Web Services Cloud for a Cancer Research Group: Proof-of-Concept Study.” *JMIR Medical Informatics* 7 (4): e14667.

Bakare, N., V. Rickerts, J. Bargon, and G. Just-Nübling. 2003. “Prevalence of Aspergillus Fumigatus and Other Fungal Species in the Sputum of Adult Patients with Cystic Fibrosis.” *Mycoses* 46 (1-2): 19–23.

Baker, J. L., J. T. Morton, M. Dinis, R. Alverez, N. C. Tran, R. Knight, and A. Edlund. 2019. “Deep Metagenomics Examines the Oral Microbiome during Dental Caries, Revealing Novel Taxa and Co-Occurrences with Host Molecules.” *bioRxiv*. https://doi.org/10.1101/804443.

Baker, Jonathon L., Batbileg Bor, Melissa Agnello, Wenyuan Shi, and Xuesong He. 2017. “Ecology of the Oral Microbiome:

Beyond Bacteria." *Trends in Microbiology* 25 (5): 362–74.
Baker, Monya. 2010. "Next-Generation Sequencing: Adjusting to Data Overload." *Nature Methods* 7 (7): 495–99.
Baldini, Federico, Almut Heinken, Laurent Heirendt, Stefania Magnusdottir, Ronan M. T. Fleming, and Ines Thiele. 2019. "The Microbiome Modeling Toolbox: From Microbial Interactions to Personalized Microbial Communities." *Bioinformatics* 35 (13): 2332–34.
Baliga, Sharmila, Sangeeta Muglikar, and Rahul Kale. 2013. "Salivary pH: A Diagnostic Biomarker." *Journal of Indian Society of Periodontology* 17 (4): 461–65.
Bandara, Herath Mudiyansalage Herath Nihal, Chamila P. Panduwawala, and Lakshman Perera Samaranayake. 2019. "Biodiversity of the Human Oral Mycobiome in Health and Disease." *Oral Diseases* 25 (2): 363–71.
Bao, Kai, Nagihan Bostanci, Thomas Thurnheer, and Georgios N. Belibasakis. 2017. "Proteomic Shifts in Multi-Species Oral Biofilms Caused by Anaeroglobus Geminatus." *Scientific Reports* 7 (1): 4409.
Barr-Agholme, M., G. Dahllöf, L. Linder, and T. Modéer. 1992. "Actinobacillus Actinomycetemcomitans, Capnocytophaga and Porphyromonas Gingivalis in Subgingival Plaque of Adolescents with Down's Syndrome." *Oral Microbiology and Immunology* 7 (4): 244–48.
Basavaraju, Mounika, Vidya Sagar Sisnity, Rajababu Palaparthy, and Pavan Kumar Addanki. 2016. "Quorum Quenching: Signal Jamming in Dental Plaque Biofilms." *Journal of Dental Sciences* 11 (4): 349–52.
Bassett, D. E., Jr, M. S. Boguski, and P. Hieter. 1996. "Yeast Genes and Human Disease." *Nature* 379 (6566): 589–90.
Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, October. https://doi.org/10.18637/jss.v067.i01.
Beaulaurier, John, Eric E. Schadt, and Gang Fang. 2019. "Deciphering Bacterial Epigenomes Using Modern Sequencing Technologies." *Nature Reviews. Genetics* 20 (3): 157–72.
Behrouzi, Ava, Amir Hossein Nafari, and Seyed Davar Siadat. 2019. "The Significance of Microbiome in Personalized Medicine." *Clinical and Translational Medicine* 8 (1): 16.
Belibasakis, Georgios N. 2018. "Microbiological Changes of the

Ageing Oral Cavity.” *Archives of Oral Biology* 96 (December): 230–32.

Belibasakis, Georgios N., Nagihan Bostanci, Philip D. Marsh, and Egija Zaura. 2019. “Applications of the Oral Microbiome in Personalized Dentistry.” *Archives of Oral Biology* 104 (August): 7–12.

Belstrøm, Daniel, Florentin Constancias, Yang Liu, Liang Yang, Daniela I. Drautz-Moses, Stephan C. Schuster, Gurjeet Singh Kohli, Tim Holm Jakobsen, Palle Holmstrup, and Michael Givskov. 2017. “Metagenomic and Metatranscriptomic Analysis of Saliva Reveals Disease-Associated Microbiota in Patients with Periodontitis and Dental Caries.” *NPJ Biofilms and Microbiomes* 3 (October): 23.

Belstrøm, Daniel, Palle Holmstrup, Nils-Erik Fiehn, Nikolai Kirkby, Alexis Kokaras, Bruce J. Paster, and Allan Bardow. 2017. “Salivary Microbiota in Individuals with Different Levels of Caries Experience.” *Journal of Oral Microbiology* 9 (1): 1270614.

Bensel, Tobias, Martin Stotz, Marianne Borneff-Lipp, Bettina Wollschläger, Andreas Wienke, Giovanni Taccetti, Silvia Campana, et al. 2011. “Lactate in Cystic Fibrosis Sputum.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 10 (1): 37–44.

Berg, Gabriele, Daria Rybakova, Doreen Fischer, Tomislav Cernava, Marie-Christine Champomier Vergès, Trevor Charles, Xiaoyulong Chen, et al. 2020. “Microbiome Definition Re-Visited: Old Concepts and New Challenges.” *Microbiome* 8 (1): 103.

Bernard, Guillaume, Jananan S. Pathmanathan, Romain Lannes, Philippe Lopez, and Eric Bapteste. 2018. “Microbial Dark Matter Investigations: How Microbial Studies Transform Biological Knowledge and Empirically Sketch a Logic of Scientific Discovery.” *Genome Biology and Evolution* 10 (3): 707–15.

Berni Canani, Roberto, Jack A. Gilbert, and Cathryn R. Nagler. 2015. “The Role of the Commensal Microbiota in the Regulation of Tolerance to Dietary Allergens.” *Current Opinion in Allergy and Clinical Immunology* 15 (3): 243–49.

Bernini, Patrizia, Ivano Bertini, Claudio Luchinat, Stefano Nepi, Edoardo Saccenti, Hartmut Schäfer, Birk Schütz, Manfred Spraul, and Leonardo Tenori. 2009. “Individual Human

Phenotypes in Metabolic Space and Time." *Journal of Proteome Research* 8 (9): 4264–71.
Bevivino, Annamaria, Giovanni Bacci, Pavel Drevinek, Maria T. Nelson, Lucas Hoffman, and Alessio Mengoni. 2019. "Deciphering the Ecology of Cystic Fibrosis Bacterial Communities: Towards Systems-Level Integration." *Trends in Molecular Medicine* 25 (12): 1110–22.
Biagi, Elena, Marco Candela, Manuela Centanni, Clarissa Consolandi, Simone Rampelli, Silvia Turroni, Marco Severgnini, et al. 2014. "Gut Microbiome in Down Syndrome." *PloS One* 9 (11): e112023.
Biagi, Elena, Claudio Franceschi, Simone Rampelli, Marco Severgnini, Rita Ostan, Silvia Turroni, Clarissa Consolandi, et al. 2016. "Gut Microbiota and Extreme Longevity." *Current Biology: CB* 26 (11): 1480–85.
Bittar, F., C. Cassagne, and E. Bosdure. 2010. "Outbreak of Corynebacterium Pseudodiphtheriticum Infection in Cystic Fibrosis Patients, France." *Emerging Infectious Diseases*. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3298292/.
Bizzarro, Sergio, Bruno G. Loos, Marja L. Laine, Wim Crielaard, and Egija Zaura. 2013. "Subgingival Microbiome in Smokers and Non-Smokers in Periodontitis: An Exploratory Study Using Traditional Targeted Techniques and a next-Generation Sequencing." *Journal of Clinical Periodontology* 40 (5): 483–92.
Bizzarro, S., M. L. Laine, M. J. Buijs, B. W. Brandt, W. Crielaard, B. G. Loos, and E. Zaura. 2016. "Microbial Profiles at Baseline and Not the Use of Antibiotics Determine the Clinical Outcome of the Treatment of Chronic Periodontitis." *Scientific Reports* 6 (February): 20205.
Blainey, Paul C., Carlos E. Milla, David N. Cornfield, and Stephen R. Quake. 2012. "Quantitative Analysis of the Human Airway Microbial Ecology Reveals a Pervasive Signature for Cystic Fibrosis." *Science Translational Medicine* 4 (153): 153ra130.
Borelli, Vincenzo, Valerie Vanhooren, Emanuela Lonardi, Karli R. Reiding, Miriam Capri, Claude Libert, Paolo Garagnani, Stefano Salvioli, Claudio Franceschi, and Manfred Wuhrer. 2015. "Plasma N-Glycome Signature of Down Syndrome." *Journal of Proteome Research* 14 (10): 4232–45.
Borenstein, Elhanan. 2012. "Computational Systems Biology and in Silico Modeling of the Human Microbiome." *Briefings in*

Bioinformatics 13 (6): 769–80.
Borges, Jorge Luis, and Andrew Hurley. 1999. *Collected Fictions*. London: The Penguin Press.
Boutin, Sébastien, and Alexander H. Dalpke. 2017. “Acquisition and Adaptation of the Airway Microbiota in the Early Life of Cystic Fibrosis Patients.” *Molecular and Cellular Pediatrics* 4 (1): 1.
Boutin, Sébastien, Simon Y. Graeber, Michael Weitnauer, Jessica Panitz, Mirjam Stahl, Diana Clausznitzer, Lars Kaderali, et al. 2015. “Comparison of Microbiomes from Different Niches of Upper and Lower Airways in Children and Adolescents with Cystic Fibrosis.” *PloS One* 10 (1): e0116029.
Bowen, William H., Robert A. Burne, Hui Wu, and Hyun Koo. 2018. “Oral Biofilms: Pathogens, Matrix, and Polymicrobial Interactions in Microenvironments.” *Trends in Microbiology* 26 (3): 229–42.
Bowyer, Ruth C. E., Daniel N. Schillereff, Matthew A. Jackson, Caroline Le Roy, Philippa M. Wells, Tim D. Spector, and Claire J. Steves. 2020. “Associations between UK Tap Water and Gut Microbiota Composition Suggest the Gut Microbiome as a Potential Mediator of Health Differences Linked to Water Quality.” *The Science of the Total Environment* 739 (October): 139697.
Broecker, Felix, Giancarlo Russo, Jochen Klumpp, and Karin Moelling. 2017. “Stable Core Virome despite Variable Microbiome after Fecal Transfer.” *Gut Microbes* 8 (3): 214–20.
Bronckers, Antonius, Lida Kalogeraki, Huub J. N. Jorna, Martina Wilke, Theodore J. Bervoets, Donacian M. Lyaruu, Behrouz Zandieh-Doulabi, Pamela Denbesten, and Hugo de Jonge. 2010. “The Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Is Expressed in Maturation Stage Ameloblasts, Odontoblasts and Bone Cells.” *Bone* 46 (4): 1188–96.
Brumfield, Kyle D., Anwar Huq, Rita R. Colwell, James L. Olds, and Menu B. Leddy. 2020. “Microbial Resolution of Whole Genome Shotgun and 16S Amplicon Metagenomic Sequencing Using Publicly Available NEON Data.” *PloS One* 15 (2): e0228899.
Brusca, María Isabel, Alcira Rosa, Olatz Albaina, María D. Moragues, Fernando Verdugo, and José Pontón. 2010. “The Impact of Oral Contraceptives on Women’s Periodontal Health and the Subgingival Occurrence of Aggressive

Periodontopathogens and Candida Species." *Journal of Periodontology* 81 (7): 1010–18.
Brusca, Samuel B., Steven B. Abramson, and Jose U. Scher. 2014. "Microbiome and Mucosal Inflammation as Extra-Articular Triggers for Rheumatoid Arthritis and Autoimmunity." *Current Opinion in Rheumatology* 26 (1): 101–7.
Burcham, Zachary M., Nicole L. Garneau, Sarah S. Comstock, Robin M. Tucker, Rob Knight, Jessica L. Metcalf, and Genetics of Taste Lab Citizen Scientists. 2020. "Patterns of Oral Microbiota Diversity in Adults and Children: A Crowdsourced Population Study." *Scientific Reports* 10 (1): 2133.
Burnet, John. 1930. *Early Greek Philosophy*. London: A. & C. Black.
Buza, Teresia M., Triza Tonui, Francesca Stomeo, Christian Tiambo, Robab Katani, Megan Schilling, Beatus Lyimo, et al. 2019. "iMAP: An Integrated Bioinformatics and Visualization Pipeline for Microbiome Data Analysis." *BMC Bioinformatics* 20 (1): 374.
Caliński, T., and J. Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics: Simulation and Computation* 3 (1): 1–27.
Callahan, Benjamin. 2018. *Silva Taxonomic Training Data Formatted for DADA2 (Silva Version 132)*. https://doi.org/10.5281/zenodo.1172783.
Callahan, Benjamin J., Paul J. McMurdie, and Susan P. Holmes. 2017. "Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis." *The ISME Journal* 11 (12): 2639–43.
Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." *Nature Methods* 13 (7): 581–83.
Callahan, Ben J., Kris Sankaran, Julia A. Fukuyama, Paul J. McMurdie, and Susan P. Holmes. 2016. "Bioconductor Workflow for Microbiome Data Analysis: From Raw Reads to Community Analyses." *F1000Research* 5 (June): 1492.
Camelo-Castillo, Anny J., Alex Mira, Alex Pico, Luigi Nibali, Brian Henderson, Nikolaos Donos, and Inmaculada Tomás. 2015. "Subgingival Microbiota in Health Compared to Periodontitis and the Influence of Smoking." *Frontiers in Microbiology* 6 (February): 119.

Cammarota, Giovanni, Gianluca Ianiro, Anna Ahern, Carmine Carbone, Andriy Temko, Marcus J. Claesson, Antonio Gasbarrini, and Giampaolo Tortora. 2020. "Gut Microbiome, Big Data and Machine Learning to Promote Precision Medicine for Cancer." *Nature Reviews. Gastroenterology & Hepatology* 17 (10): 635–48.

Cao, Yali, Min Qiao, Zhigang Tian, Yan Yu, Baohua Xu, Wansheng Lao, Xuguo Ma, and Wenge Li. 2018. "Comparative Analyses of Subgingival Microbiome in Chronic Periodontitis Patients with and Without IgA Nephropathy by High Throughput 16S rRNA Sequencing." *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology* 47 (2): 774–83.

Caporaso, J. Gregory, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, et al. 2010. "QIIME Allows Analysis of High-Throughput Community Sequencing Data." *Nature Methods* 7 (5): 335–36.

Carinci, F., M. Martinelli, M. Contaldo, R. Santoro, F. Pezzetti, D. Lauritano, V. Candotto, et al. 2018. "Focus on Periodontal Disease and Development of Endocarditis." *Journal of Biological Regulators and Homeostatic Agents* 32 (2 Suppl. 1): 143–47.

Carmody, Lisa A., Jiangchao Zhao, Patrick D. Schloss, Joseph F. Petrosino, Susan Murray, Vincent B. Young, Jun Z. Li, and John J. LiPuma. 2013. "Changes in Cystic Fibrosis Airway Microbiota at Pulmonary Exacerbation." *Annals of the American Thoracic Society* 10 (3): 179–87.

Caroline de Abreu Brandi, Thayse, Maristela Barbosa Portela, Paula Moraes Lima, Gloria Fernanda Barbosa de Araújo Castro, Lucianne Cople Maia, and Andréa Fonseca-Gonçalves. 2016. "Demineralizing Potential of Dental Biofilm Added with Candida Albicans and Candida Parapsilosis Isolated from Preschool Children with and without Caries." *Microbial Pathogenesis* 100 (November): 51–55.

Carrouel, Florence, Stéphane Viennot, Julie Santamaria, Philippe Veber, and Denis Bourgeois. 2016. "Quantitative Molecular Detection of 19 Major Pathogens in the Interdental Biofilm of Periodontally Healthy Young Adults." *Frontiers in Microbiology* 7 (June): 840.

Carter, Christopher. 2010. "Bioinformatics Analysis of Homologies between Pathogen Antigens, Autoantigens and the CFTR

Cystic Fibrosis Protein: A Role for Immunoadsorption Therapy?" *Nature Precedings*, December. https://doi.org/10.1038/npre.2010.5352.1.
Casarin, R. C. V., A. Barbagallo, T. Meulman, V. R. Santos, E. A. Sallum, F. H. Nociti, P. M. Duarte, M. Z. Casati, and R. B. Gonçalves. 2013. "Subgingival Biodiversity in Subjects with Uncontrolled Type-2 Diabetes and Chronic Periodontitis." *Journal of Periodontal Research* 48 (1): 30–36.
Castellarin, Mauro, René L. Warren, J. Douglas Freeman, Lisa Dreolini, Martin Krzywinski, Jaclyn Strauss, Rebecca Barnes, et al. 2012. "Fusobacterium Nucleatum Infection Is Prevalent in Human Colorectal Carcinoma." *Genome Research* 22 (2): 299–306.
Catalán, Marcelo A., Kathleen Scott-Anne, Marlise I. Klein, Hyun Koo, William H. Bowen, and James E. Melvin. 2011. "Elevated Incidence of Dental Caries in a Mouse Model of Cystic Fibrosis." *PloS One* 6 (1): e16549.
Celluzzi, Antonella, and Andrea Masotti. 2016. "How Our Other Genome Controls Our Epi-Genome." *Trends in Microbiology* 24 (10): 777–87.
Chalabi, Maryam, Sharareh Moghim, Ahamad Mogharehabed, Farid Najafi, and Faranak Rezaie. 2008. "EBV and CMV in Chronic Periodontitis: A Prevalence Study." *Archives of Virology* 153 (10): 1917–19.
Chambers, Stephen T., David Murdoch, Arthur Morris, David Holland, Paul Pappas, Manel Almela, Nuria Fernández-Hidalgo, et al. 2013. "HACEK Infective Endocarditis: Characteristics and Outcomes from a Large, Multi-National Cohort." *PloS One* 8 (5): e63181.
Chatterjee, A., and H. Bhattacharya. 2011. "Probiotics in Periodontal Health and Disease." *Society of Periodontology*. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3134041/.
Chattopadhyay, Indranil, Mukesh Verma, and Madhusmita Panda. 2019. "Role of Oral Microbiome Signatures in Diagnosis and Prognosis of Oral Cancer." *Technology in Cancer Research & Treatment* 18 (January): 1533033819867354.
Chaushu, Stella, Adrian Becker, Gabriel Chaushu, and Joseph Shapira. 2002. "Stimulated Parotid Salivary Flow Rate in Patients with Down Syndrome." *Special Care in Dentistry: Official Publication of the American Association of Hospital Dentists, the Academy of Dentistry for the Handicapped, and*

the American Society for Geriatric Dentistry 22 (1): 41–44.
Chen, Bin, Yan Zhao, Shufeng Li, Lanxiu Yang, Haiying Wang, Tao Wang, Bin Shi, et al. 2018. “Variations in Oral Microbiome Profiles in Rheumatoid Arthritis and Osteoarthritis with Potential Biomarkers for Arthritis Screening.” *Scientific Reports* 8 (1): 17126.
Chen, C. 1996. “Distribution of a Newly Described Species, Kingella Oralis, in the Human Oral Cavity.” *Oral Microbiology and Immunology* 11 (6): 425–27.
Chen, Casey, Chris Hemme, Joan Beleno, Zhou Jason Shi, Daliang Ning, Yujia Qin, Qichao Tu, et al. 2018. “Oral Microbiota of Periodontal Health and Disease and Their Changes after Nonsurgical Periodontal Therapy.” *The ISME Journal* 12 (5): 1210–24.
Cheng, Ronald H. W., Cynthia K. Y. Yiu, and W. Keung Leung. 2011. “Oral Health in Individuals with Down Syndrome.” *Prenatal Diagnosis and Screening for Down Syndrome. Rijeka, Croatia: In Tech*, 59–76.
Chen, Hui, and Wen Jiang. 2014. “Application of High-Throughput Sequencing in Understanding Human Oral Microbiome Related with Health and Disease.” *Frontiers in Microbiology* 5 (October): 508.
Chen, Hui, Zongxin Ling, Jenny Wen, and Jn Zhang. 2013. “Oral Microbiota in Chinese She Children with Dental Caries,” March. http://dx.doi.org/.
Chen, Wen-Pei, Shih-Hao Chang, Chuan-Yi Tang, Ming-Li Liou, Suh-Jen Jane Tsai, and Yaw-Ling Lin. 2018. “Composition Analysis and Feature Selection of the Oral Microbiota Associated with Periodontal Disease.” *BioMed Research International* 2018 (November): 3130607.
Chhibber-Goel, Jyoti, Varsha Singhal, Debaleena Bhowmik, Rahul Vivek, Neeraj Parakh, Balram Bhargava, and Amit Sharma. 2016. “Linkages between Oral Commensal Bacteria and Atherosclerotic Plaques in Coronary Artery Disease Patients.” *NPJ Biofilms and Microbiomes* 2 (December): 7.
Chomicz, L., J. Piekarczyk, B. Starosciak, P. Fiedor, B. Piekarczyk, D. Szubinska, P. J. Zawadzki, and M. Walski. 2002. “Comparative Studies on the Occurrence of Protozoans, Bacteria and Fungi in the Oral Cavity of Patients with Systemic Disorders.” *Acta Parasitologica / Witold Stefanski Institute of Parasitology, Warszawa, Poland* 2 (47). https://www.infona.pl//

resource/bwmeta1.element.agro-article-400b2855-6734-41ea-a377-fd91ab44111b.

Chotirmall, Sanjay H., Elaine O'Donoghue, Kathleen Bennett, Cedric Gunaratnam, Shane J. O'Neill, and Noel G. McElvaney. 2010. "Sputum Candida Albicans Presages FEV_1 Decline and Hospital-Treated Exacerbations in Cystic Fibrosis." *Chest* 138 (5): 1186–95.

Ciardo, Diana E., Katja Lucke, Alex Imhof, Guido V. Bloemberg, and Erik C. Böttger. 2010. "Systematic Internal Transcribed Spacer Sequence Analysis for Identification of Clinical Mold Isolates in Diagnostic Mycology: A 5-Year Study." *Journal of Clinical Microbiology* 48 (8): 2809–13.

Cichon, P., L. Crawford, and W. D. Grimm. 1998. "Early-Onset Periodontitis Associated with Down's Syndrome--Clinical Interventional Study." *Annals of Periodontology / the American Academy of Periodontology* 3 (1): 370–80.

Cielecka, D., L. Chomicz, J. Piekarczyk, M. Walski, P. J. Zawadzki, A. Bednarczyk, D. Szubińska, and Others. 2000. "Oral Cavity Condition and the Occurrence of Parasitic Protozoans in Patients with Genetic Diseases." *Acta Parasitologica / Witold Stefanski Institute of Parasitology, Warszawa, Poland* 45 (2): 107–12.

Clemente, Jose C., Erica C. Pehrsson, Martin J. Blaser, Kuldip Sandhu, Zhan Gao, Bin Wang, Magda Magris, et al. 2015. "The Microbiome of Uncontacted Amerindians." *Science Advances* 1 (3). https://doi.org/10.1126/sciadv.1500183.

Coburn, Bryan, Pauline W. Wang, Julio Diaz Caballero, Shawn T. Clark, Vijaya Brahma, Sylva Donaldson, Yu Zhang, et al. 2015. "Lung Microbiota across Age and Disease Stage in Cystic Fibrosis." *Scientific Reports* 5 (May): 10241.

Cocco, Fabio, Maria G. Cagetti, Peter Lingström, Nicole Camoni, and Guglielmo Campus. 2017. "The Strip Method and the Microelectrode Technique in Assessing Dental Plaque pH." *Minerva Stomatologica* 66 (6): 241–47.

Coelho, Edgar D., Joel P. Arrais, Sérgio Matos, Carlos Pereira, Nuno Rosa, Maria José Correia, Marlene Barros, and José Luís Oliveira. 2014. "Computational Prediction of the Human-Microbial Oral Interactome." *BMC Systems Biology* 8 (February): 24.

Coenye, Tom, Johan Goris, Theodore Spilker, Peter Vandamme, and John J. LiPuma. 2002. "Characterization of Unusual

Bacteria Isolated from Respiratory Secretions of Cystic Fibrosis Patients and Description of Inquilinus Limosus Gen. Nov., Sp. Nov.” *Journal of Clinical Microbiology* 40 (6): 2062–69.

Coffey, Michael J., Shaun Nielsen, Bernd Wemheuer, Nadeem O. Kaakoush, Millie Garg, Bronwen Needham, Russell Pickford, Adam Jaffe, Torsten Thomas, and Chee Y. Ooi. 2019. “Gut Microbiota in Children With Cystic Fibrosis: A Taxonomic and Functional Dysbiosis.” *Scientific Reports* 9 (1): 18593.

Colombo, Ana Paula V., Susan Bennet, Sean L. Cotton, J. Max Goodson, Ralph Kent, Anne D. Haffajee, Sigmund S. Socransky, et al. 2012. “Impact of Periodontal Therapy on the Subgingival Microbiota of Severe Periodontitis: Comparison between Good Responders and Individuals with Refractory Periodontitis Using the Human Oral Microbe Identification Microarray.” *Journal of Periodontology* 83 (10): 1279–87.

Colombo, Ana Paula V., Susan K. Boches, Sean L. Cotton, J. Max Goodson, Ralph Kent, Anne D. Haffajee, Sigmund S. Socransky, et al. 2009. “Comparisons of Subgingival Microbial Profiles of Refractory Periodontitis, Severe Periodontitis, and Periodontal Health Using the Human Oral Microbe Identification Microarray.” *Journal of Periodontology* 80 (9): 1421–32.

Colombo, Andrea V., Graziela M. Barbosa, Daniela Higashi, Giorgio di Micheli, Paulo H. Rodrigues, and Maria Regina L. Simionato. 2013. “Quantitative Detection of Staphylococcus Aureus, Enterococcus Faecalis and Pseudomonas Aeruginosa in Human Oral Epithelial Cells from Subjects with Periodontitis and Periodontal Health.” *Journal of Medical Microbiology* 62 (Pt 10): 1592–1600.

Conrad, C., J. Lymp, V. Thompson, C. Dunn, Z. Davies, B. Chatfield, D. Nichols, et al. 2015. “Long-Term Treatment with Oral N-Acetylcysteine: Affects Lung Function but Not Sputum Inflammation in Cystic Fibrosis Subjects. A Phase II Randomized Placebo-Controlled Trial.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 14 (2): 219–27.

Contardo, M. S., N. Díaz, O. Lobos, C. Padilla, and R. A. Giacaman. 2011. “Oral Colonization by Streptococcus Mutans and Its Association with the Severity of Periodontal Disease in Adults.” *Revista Clínica de Periodoncia, Implantología Y*

Rehabilitación Oral 4 (1): 9–12.
Costalonga, Massimo, and Mark C. Herzberg. 2014. “The Oral Microbiome and the Immunobiology of Periodontal Disease and Caries.” *Immunology Letters* 162 (2 Pt A): 22–38.
Costea, Paul I., Falk Hildebrand, Manimozhiyan Arumugam, Fredrik Bäckhed, Martin J. Blaser, Frederic D. Bushman, Willem M. de Vos, et al. 2018. “Enterotypes in the Landscape of Gut Microbial Community Composition.” *Nature Microbiology* 3 (1): 8–16.
Costello, Elizabeth K., Christian L. Lauber, Micah Hamady, Noah Fierer, Jeffrey I. Gordon, and Rob Knight. 2009. “Bacterial Community Variation in Human Body Habitats across Space and Time.” *Science* 326 (5960): 1694–97.
Costello, Elizabeth K., Keaton Stagaman, Les Dethlefsen, Brendan J. M. Bohannan, and David A. Relman. 2012. “The Application of Ecological Theory toward an Understanding of the Human Microbiome.” *Science* 336 (6086): 1255–62.
Cox, D. R., Christiana Kartsonaki, and Ruth H. Keogh. 2018. “Big Data: Some Statistical Issues.” *Statistics & Probability Letters* 136 (May): 111–15.
Cox, Michael J., Martin Allgaier, Byron Taylor, Marshall S. Baek, Yvonne J. Huang, Rebecca A. Daly, Ulas Karaoz, et al. 2010. “Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients.” *PloS One* 5 (6): e11044.
Cruz Peña, Maria José de la, Francisco Martinez-Hernandez, Inmaculada Garcia-Heredia, Mónica Lluesma Gomez, Òscar Fornas, and Manuel Martinez-Garcia. 2018. “Deciphering the Human Virome with Single-Virus Genomics and Metagenomics.” *Viruses* 10 (3). https://doi.org/10.3390/v10030113.
Cuervo, Adriana, Arancha Hevia, Patricia López, Ana Suárez, Carmen Diaz, Borja Sánchez, Abelardo Margolles, and Sonia González. 2016. “Phenolic Compounds from Red Wine and Coffee Are Associated with Specific Intestinal Microorganisms in Allergic Subjects.” *Food & Function* 7 (1): 104–9.
Cuthbertson, Leah, Alan W. Walker, Anna E. Oliver, Geraint B. Rogers, Damian W. Rivett, Thomas H. Hampton, Alix Ashare, et al. 2020. “Lung Function and Microbiota Diversity in Cystic Fibrosis.” *Microbiome* 8 (1): 45.
Cutler, C., M. Kiernan, J. R. Willis, L. Gallardo-Alfaro, P. Casas-Agustench, D. White, M. Hickson, T. Gabaldon, and R.

Bescos. 2019. “Post-Exercise Hypotension and Skeletal Muscle Oxygenation Is Regulated by Nitrate-Reducing Activity of Oral Bacteria.” *Free Radical Biology & Medicine* 143 (November): 252–59.

“DADA2 Pipeline Tutorial (1.16).” n.d. Accessed July 24, 2021. https://benjjneb.github.io/dada2/tutorial.html.

Dani, Sneha, Ashwin Prabhu, K. R. Chaitra, N. C. Desai, Sudhir R. Patil, and Ranjan Rajeev. 2016. “Assessment of Streptococcus Mutans in Healthy versus Gingivitis and Chronic Periodontitis: A Clinico-Microbiological Study.” *Contemporary Clinical Dentistry* 7 (4): 529–34.

Dassi, Erik, Pamela Ferretti, Giuseppina Covello, HTM-CMB-2015, Roberto Bertorelli, Michela A. Denti, Veronica De Sanctis, Adrian Tett, and Nicola Segata. 2018. “The Short-Term Impact of Probiotic Consumption on the Oral Cavity Microbiome.” *Scientific Reports* 8 (1): 10476.

David, Lawrence A., Arne C. Materna, Jonathan Friedman, Maria I. Campos-Baptista, Matthew C. Blackburn, Allison Perrotta, Susan E. Erdman, and Eric J. Alm. 2014. “Host Lifestyle Affects Human Microbiota on Daily Timescales.” *Genome Biology* 15 (7): R89.

Davidovich, Esti, Doron Jaacob Aframian, Josef Shapira, and Benjamin Peretz. 2010. “A Comparison of the Sialochemistry, Oral pH, and Oral Health Status of Down Syndrome Children to Healthy Children.” *International Journal of Paediatric Dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children* 20 (4): 235–41.

Debelius, Justine, Se Jin Song, Yoshiki Vazquez-Baeza, Zhenjiang Zech Xu, Antonio Gonzalez, and Rob Knight. 2016. “Tiny Microbes, Enormous Impacts: What Matters in Gut Microbiome Studies?” *Genome Biology* 17 (1): 217.

De Filippis, Francesca, Lucia Vannini, Antonietta La Storia, Luca Laghi, Paola Piombino, Giuseppina Stellato, Diana I. Serrazanetti, et al. 2014. “The Same Microbiota and a Potentially Discriminant Metabolome in the Saliva of Omnivore, Ovo-Lacto-Vegetarian and Vegan Individuals.” *PloS One* 9 (11): e112373.

De-La-Torre, Janire, Guillermo Quindós, Cristina Marcos-Arias, Xabier Marichalar-Mendia, María Luisa Gainza, Elena Eraso, Amelia Acha-Sagredo, and José Manuel Aguirre-Urizar. 2018.

“Oral Candida Colonization in Patients with Chronic Periodontitis. Is There Any Relationship?” *Revista Iberoamericana de Micologia* 35 (3): 134–39.

Deng, Zhi-Luo, Szymon P. Szafrański, Michael Jarek, Sabin Bhuju, and Irene Wagner-Döbler. 2017. “Dysbiosis in Chronic Periodontitis: Key Microbial Players and Interactions with the Human Host.” *Scientific Reports* 7 (1): 3703.

Deps, Tahyna Duda, Gabriela Lopes Angelo, Carolina Castro Martins, Saul Martins Paiva, Isabela Almeida Pordeus, and Ana Cristina Borges-Oliveira. 2015. “Association between Dental Caries and Down Syndrome: A Systematic Review and Meta-Analysis.” *PloS One* 10 (6): e0127484.

Devine, Deirdre A., Philip D. Marsh, and Josephine Meade. 2015. “Modulation of Host Responses by Oral Commensal Bacteria.” *Journal of Oral Microbiology* 7 (February): 26941.

Dewhirst, F. E., C-K C. Chen, B. J. Paster, and J. J. Zambon. 1993. “Phylogeny of Species in the Family Neisseriaceae Isolated from Human Dental Plaque and Description of Kingella Orale Sp. Nov.” *International Journal of Systematic Bacteriology* 43 (3): 490–99.

Dewhirst, Floyd E., Tuste Chen, Jacques Izard, Bruce J. Paster, Anne C. R. Tanner, Wen-Han Yu, Abirami Lakshmanan, and William G. Wade. 2010. “The Human Oral Microbiome.” *Journal of Bacteriology* 192 (19): 5002–17.

Dhotre, Shree V., Milind S. Davane, and Basavraj S. Nagoba. 2017. “Periodontitis, Bacteremia and Infective Endocarditis: A Review Study.” *Archives of Pediatric Infectious Diseases* 5 (3). https://doi.org/10.5812/pedinfect.41067.

Dias, Marcela F., Mariana P. Reis, Leonardo B. Acurcio, Anderson O. Carmo, Cristiane F. Diamantino, Amanda M. Motta, Evanguedes Kalapothakis, Jacques R. Nicoli, and Andréa M. A. Nascimento. 2018. “Changes in Mouse Gut Bacterial Community in Response to Different Types of Drinking Water.” *Water Research* 132 (April): 79–89.

Diaz, P. I., and A. Dongari-Bagtzoglou. 2021. “Critically Appraising the Significance of the Oral Mycobiome.” *Journal of Dental Research* 100 (2): 133–40.

Dickson, Robert P., John R. Erb-Downward, Christine M. Freeman, Lisa McCloskey, James M. Beck, Gary B. Huffnagle, and Jeffrey L. Curtis. 2015. “Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung

Biogeography." *Annals of the American Thoracic Society* 12 (6): 821–30.
Dige, I., M. K. Raarup, J. R. Nyengaard, M. Kilian, and B. Nyvad. 2009. "Actinomyces Naeslundii in Initial Dental Biofilm Formation." *Microbiology* 155 (Pt 7): 2116–26.
Ding, Tao, and Patrick D. Schloss. 2014. "Dynamics and Associations of Microbial Community Types across the Human Body." *Nature* 509 (7500): 357–60.
Dios Caballero, Juan de, Rosa Del Campo, Ana Royuela, Amparo Solé, Luis Máiz, Casilda Olveira, Esther Quintana-Gallego, et al. 2016. "Bronchopulmonary Infection-Colonization Patterns in Spanish Cystic Fibrosis Patients: Results from a National Multicenter Study." *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 15 (3): 357–65.
Dissick, Addie, Robert S. Redman, Miata Jones, Bavana V. Rangan, Andreas Reimold, Garth R. Griffiths, Ted R. Mikuls, Richard L. Amdur, John S. Richards, and Gail S. Kerr. 2010. "Association of Periodontitis with Rheumatoid Arthritis: A Pilot Study." *Journal of Periodontology* 81 (2): 223–30.
Domènech, Laura, Jesse Willis, Maria Alemany, Marta Morell, Eva Real, Geòrgia Escaramís, Sara Bertolín, et al. 2020. "Changes in the Stool and Oropharyngeal Microbiome in Obsessive-Compulsive Disorder." *bioRxiv*. medRxiv. https://doi.org/10.1101/2020.05.26.20113779.
Domingues, Natália Bertolo, Matheus Racy Mariusso, Marcia Hiromi Tanaka, Raquel Mantuaneli Scarel-Caminaga, Marcia Pinto Alves Mayer, Fernanda Lourenção Brighenti, Ângela Cristina Cilense Zuanon, Flávia Kazue Ibuki, Fernando Neves Nogueira, and Elisa Maria Aparecida Giro. 2017. "Reduced Salivary Flow Rate and High Levels of Oxidative Stress in Whole Saliva of Children with Down Syndrome." *Special Care in Dentistry: Official Publication of the American Association of Hospital Dentists, the Academy of Dentistry for the Handicapped, and the American Society for Geriatric Dentistry* 37 (6): 269–76.
Dominy, Stephen S., Casey Lynch, Florian Ermini, Malgorzata Benedyk, Agata Marczyk, Andrei Konradi, Mai Nguyen, et al. 2019. "Porphyromonas Gingivalis in Alzheimer's Disease Brains: Evidence for Disease Causation and Treatment with Small-Molecule Inhibitors." *Science Advances* 5 (1): eaau3333.
Donovan, Paul D., Gabriel Gonzalez, Desmond G. Higgins,

Geraldine Butler, and Kimihito Ito. 2018. “Identification of Fungi in Shotgun Metagenomics Datasets.” *PloS One* 13 (2): e0192898.

Down’s Heart Group. n.d. “Dental Care for Children and Adults with Heart Problems and Down’s Syndrome - Down's Heart Group.” Down’s Heart Group. Accessed July 7, 2020a. https://dhg.org.uk/information/dental-care/.

———. n.d. “Infective Endocarditis in Children and Adults with Heart Problems and Down’s Syndrome - Down's Heart Group.” Down’s Heart Group. Accessed July 7, 2020b. https://dhg.org.uk/information/infective-endocarditis/.

Dupuy, Amanda K., Marika S. David, Lu Li, Thomas N. Heider, Jason D. Peterson, Elizabeth A. Montano, Anna Dongari-Bagtzoglou, Patricia I. Diaz, and Linda D. Strausbaugh. 2014. “Redefining the Human Oral Mycobiome with Improved Practices in Amplicon-Based Taxonomy: Discovery of Malassezia as a Prominent Commensal.” *PloS One* 9 (3): e90899.

Durazzi, Francesco, Claudia Sala, Gastone Castellani, Gerardo Manfreda, Daniel Remondini, and Alessandra De Cesare. 2021. “Comparison between 16S rRNA and Shotgun Sequencing Data for the Taxonomic Characterization of the Gut Microbiota.” *Scientific Reports* 11 (1): 3030.

Duvallet, Claire, Sean M. Gibbons, Thomas Gurry, Rafael A. Irizarry, and Eric J. Alm. 2017. “Meta-Analysis of Gut Microbiome Studies Identifies Disease-Specific and Shared Responses.” *Nature Communications* 8 (1): 1784.

Easterly, Caleb W., Ray Sajulga, Subina Mehta, James Johnson, Praveen Kumar, Shane Hubler, Bart Mesuere, Joel Rudney, Timothy J. Griffin, and Pratik D. Jagtap. 2019. “metaQuantome: An Integrated, Quantitative Metaproteomics Approach Reveals Connections Between Taxonomy and Protein Function in Complex Microbiomes.” *Molecular & Cellular Proteomics: MCP* 18 (8 suppl 1): S82–91.

Edlund, Anna, Neha Garg, Hosein Mohimani, Alexey Gurevich, Xuesong He, Wenyuan Shi, Pieter C. Dorrestein, and Jeffrey S. McLean. 2017. “Metabolic Fingerprints from the Human Oral Microbiome Reveal a Vast Knowledge Gap of Secreted Small Peptidic Molecules.” *mSystems* 2 (4). https://doi.org/10.1128/mSystems.00058-17.

Eidt, Gustavo, Caroline Gomes de Andrade, Thais de Cássia

Negrini, and Rodrigo Alex Arthur. 2019. “Role of Candida Albicans on Enamel Demineralization and on Acidogenic Potential of Streptococcus Mutans in Vitro Biofilms.” *Journal of Applied Oral Science: Revista FOB* 27 (September): e20180593.

Eisenstein, Michael. 2020. “The Hunt for a Healthy Microbiome.” *Nature* 577 (7792): S6–8.

Eren, A. Murat, Gary G. Borisy, Susan M. Huse, and Jessica L. Mark Welch. 2014. “Oligotyping Analysis of the Human Oral Microbiome.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (28): E2875–84.

Eribe, Emenike Ribs K., and Ingar Olsen. 2008. “Leptotrichia Species in Human Infections.” *Anaerobe* 14 (3): 131–37.

Escapa, Isabel F., Tsute Chen, Yanmei Huang, Prasad Gajare, Floyd E. Dewhirst, and Katherine P. Lemon. 2018. “New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): A Resource for the Microbiome of the Human Aerodigestive Tract.” *mSystems* 3 (6). https://doi.org/10.1128/mSystems.00187-18.

Escudié, Frédéric, Lucas Auer, Maria Bernard, Mahendra Mariadassou, Laurent Cauquil, Katia Vidal, Sarah Maman, Guillermina Hernandez-Raquet, Sylvie Combes, and Géraldine Pascal. 2018. “FROGS: Find, Rapidly, OTUs with Galaxy Solution.” *Bioinformatics* 34 (8): 1287–94.

Espinoza, Josh L., Derek M. Harkins, Manolito Torralba, Andres Gomez, Sarah K. Highlander, Marcus B. Jones, Pamela Leong, et al. 2018. “Supragingival Plaque Microbiome Ecology and Functional Potential in the Context of Health and Disease.” *mBio* 9 (6). https://doi.org/10.1128/mBio.01631-18.

Fan, Xiaozhou, Alexander V. Alekseyenko, Jing Wu, Brandilyn A. Peters, Eric J. Jacobs, Susan M. Gapstur, Mark P. Purdue, et al. 2018. “Human Oral Microbiome and Prospective Risk for Pancreatic Cancer: A Population-Based Nested Case-Control Study.” *Gut* 67 (1): 120–27.

Fan, Xiaozhou, Brandilyn A. Peters, Deborah Min, Jiyoung Ahn, and Richard B. Hayes. 2018. “Comparison of the Oral Microbiome in Mouthwash and Whole Saliva Samples.” *PloS One* 13 (4): e0194729.

Faust, Karoline. 2017. “Microbial Association Network Construction Tutorial.” http://psbweb05.psb.ugent.be/conet/microbialnetworks/spiecea

si.php.
Faust, Karoline, Franziska Bauchinger, Sophie de Buyl, and Leo Lahti. 2020. *Seqtime: Time Series Analysis of Sequencing Data* (version 0.1.1). Github. https://github.com/hallucigenia-sparsa/seqtime.
Feki, A., B. Molet, R. Haag, and M. Kremer. 1981. “[Protozoa of the human oral cavity (epidemiological correlations and pathogenic possibilities].” *Journal de biologie buccale* 9 (2): 155–61.
Feres, Magda, Flavia Teles, Ricardo Teles, Luciene Cristina Figueiredo, and Marcelo Faveri. 2016. “The Subgingival Periodontal Microbiota of the Aging Mouth.” *Periodontology 2000* 72 (1): 30–53.
Fernandez-Feo, M., G. Wei, G. Blumenkranz, F. E. Dewhirst, D. Schuppan, F. G. Oppenheim, and E. J. Helmerhorst. 2013. “The Cultivable Human Oral Gluten-Degrading Microbiome and Its Potential Implications in Coeliac Disease and Gluten Sensitivity.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 19 (9): E386–94.
Fernández-Olmos, Ana, María García-Castillo, María-Isabel Morosini, Adelaida Lamas, Luis Máiz, and Rafael Cantón. 2012. “MALDI-TOF MS Improves Routine Identification of Non-Fermenting Gram Negative Isolates from Cystic Fibrosis Patients.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 11 (1): 59–62.
Fernandez Y Mostajo, M., R. A. M. Exterkate, M. J. Buijs, W. Beertsen, G. A. van der Weijden, E. Zaura, and W. Crielaard. 2017. “A Reproducible Microcosm Biofilm Model of Subgingival Microbial Communities.” *Journal of Periodontal Research* 52 (6): 1021–31.
Fiehn, Oliver. 2002. “Metabolomics--the Link between Genotypes and Phenotypes.” *Plant Molecular Biology* 48 (1-2): 155–71.
Filkins, L. M., T. H. Hampton, A. H. Gifford, M. J. Gross, D. A. Hogan, M. L. Sogin, H. G. Morrison, B. J. Paster, and G. A. O’Toole. 2012. “Prevalence of Streptococci and Increased Polymicrobial Diversity Associated with Cystic Fibrosis Patient Stability.” *Journal of Bacteriology* 194 (17): 4709–17.
Filosi, Michele, Roberto Visintainer, and Samantha Riccadonna. 2017. *Nettools: A Network Comparison Framework* (version 1.1.0).
Fine, Daniel H., Amey G. Patil, and Senthil K. Velusamy. 2019.

“Aggregatibacter Actinomycetemcomitans (Aa) Under the Radar: Myths and Misunderstandings of Aa and Its Role in Aggressive Periodontitis.” *Frontiers in Immunology* 10 (April): 728.

Fitzpatrick, Sarah G., and Joseph Katz. 2010. “The Association between Periodontal Disease and Cancer: A Review of the Literature.” *Journal of Dentistry* 38 (2): 83–95.

Flemer, Burkhardt, Ryan D. Warren, Maurice P. Barrett, Katryna Cisek, Anubhav Das, Ian B. Jeffery, Eimear Hurley, Micheal O’Riordain, Fergus Shanahan, and Paul W. O’Toole. 2018. “The Oral Microbiota in Colorectal Cancer Is Distinctive and Predictive.” *Gut* 67 (8): 1454–63.

Fouhy, Fiona, Adam G. Clooney, Catherine Stanton, Marcus J. Claesson, and Paul D. Cotter. 2016. “16S rRNA Gene Sequencing of Mock Microbial Populations- Impact of DNA Extraction Method, Primer Choice and Sequencing Platform.” *BMC Microbiology* 16 (1): 123.

Foweraker, Juliet. 2009. “Recent Advances in the Microbiology of Respiratory Tract Infection in Cystic Fibrosis.” *British Medical Bulletin* 89 (January): 93–110.

Fox, J., and S. Weisberg. 2019. *An R Companion to Applied Regression*. Third. Sage.

Franceschi, Claudio, and Judith Campisi. 2014. “Chronic Inflammation (inflammaging) and Its Potential Contribution to Age-Associated Diseases.” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 69 Suppl 1 (June): S4–9.

Franceschi, Claudio, Paolo Garagnani, Cristina Morsiani, Maria Conte, Aurelia Santoro, Andrea Grignolio, Daniela Monti, Miriam Capri, and Stefano Salvioli. 2018. “The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates.” *Frontiers of Medicine* 5 (March): 61.

Fridman, David, Ali Chaudhry, John Makaryus, Karen Black, and Amgad N. Makaryus. 2016. “Rothia Dentocariosa Endocarditis: An Especially Rare Case in a Previously Healthy Man.” *Texas Heart Institute Journal / from the Texas Heart Institute of St. Luke’s Episcopal Hospital, Texas Children's Hospital* 43 (3): 255–57.

Fritschi, B. Zinsli, A. Albert-Kiszely, and G. R. Persson. 2008. “Staphylococcus Aureus and Other Bacteria in Untreated Periodontitis.” *Journal of Dental Research* 87 (6): 589–93.

Fritz, Adrian, Peter Hofmann, Stephan Majda, Eik Dahms, Johannes Dröge, Jessika Fiedler, Till R. Lesker, et al. 2019. “CAMISIM: Simulating Metagenomes and Microbial Communities.” *Microbiome* 7 (1): 17.

Fujise, O., W. Chen, S. Rich, and C. Chen. 2004. “Clonal Diversity and Stability of Subgingival Eikenella Corrodens.” *Journal of Clinical Microbiology* 42 (5): 2036–42.

Fukami, Tadashi, and Mifuyu Nakajima. 2011. “Community Assembly: Alternative Stable States or Alternative Transient States?” *Ecology Letters* 14 (10): 973–84.

Fukui, Yuto, Kotaro Aoki, Yoshikazu Ishii, and Kazuhiro Tateda. 2018. “The Palatine Tonsil Bacteriome, but Not the Mycobiome, Is Altered in HIV Infection.” *BMC Microbiology* 18 (1): 127.

Galloway-Peña, Jessica R., Daniel P. Smith, Pranoti Sahasrabhojane, W. Duncan Wadsworth, Bryan M. Fellman, Nadim J. Ajami, Elizabeth J. Shpall, et al. 2017. “Characterization of Oral and Gut Microbiome Temporal Variability in Hospitalized Cancer Patients.” *Genome Medicine* 9 (1): 21.

Ganesan, Sukirth M., Vinayak Joshi, Megan Fellows, Shareef M. Dabdoub, Haikady N. Nagaraja, Benjamin O’Donnell, Neeta Rohit Deshpande, and Purnima S. Kumar. 2017. “A Tale of Two Risks: Smoking, Diabetes and the Subgingival Microbiome.” *The ISME Journal* 11 (9): 2075–89.

Gao, Bei, Tara Gallagher, Ying Zhang, Mona Elbadawi-Sidhu, Zijuan Lai, Oliver Fiehn, and Katrine L. Whiteson. 2018. “Tracking Polymicrobial Metabolism in Cystic Fibrosis Airways: Pseudomonas Aeruginosa Metabolism and Physiology Are Influenced by Rothia Mucilaginosa-Derived Metabolites.” *mSphere* 3 (2). https://doi.org/10.1128/mSphere.00151-18.

García-Castillo, María, María Isabel Morosini, Aránzazu Valverde, Felisa Almaraz, Fernando Baquero, Rafael Cantón, and Rosa del Campo. 2007. “Differences in Biofilm Development and Antibiotic Susceptibility among Streptococcus Pneumoniae Isolates from Cystic Fibrosis Samples and Blood Cultures.” *The Journal of Antimicrobial Chemotherapy* 59 (2): 301–4.

Gardner, Helen, ed. 1961. “Introduction.” In *The Metaphysical Poets*, xxiii. London: Oxford University Press.

Genco, R. J., M. J. LaMonte, D. I. McSkimming, M. J. Buck, L. Li, K. M. Hovey, C. A. Andrews, et al. 2019. “The Subgingival

Microbiome Relationship to Periodontal Disease in Older Women.” *Journal of Dental Research* 98 (9): 975–84.
Ghayoumi, Nooshin, Casey Chen, and Jørgen Slots. 2002. “Dialister Pneumosintes, a New Putative Periodontal Pathogen.” *Journal of Periodontal Research* 37 (1): 75–78.
Gloor, Gregory B., Jean M. Macklaim, Vera Pawlowsky-Glahn, and Juan J. Egozcue. 2017. “Microbiome Datasets Are Compositional: And This Is Not Optional.” *Frontiers in Microbiology* 8 (November): 2224.
Gloor, Gregory B., and Gregor Reid. 2016. “Compositional Analysis: A Valid Approach to Analyze Microbiome High-Throughput Sequencing Data.” *Canadian Journal of Microbiology* 62 (8): 692–703.
Gloor, Gregory B., Jia Rong Wu, Vera Pawlowsky-Glahn, and Juan José Egozcue. 2016. “It’s All Relative: Analyzing Microbiome Data as Compositions.” *Annals of Epidemiology* 26 (5): 322–29.
Gomes-Filho, Isaac S., Johelle S. Passos, and Simone Seixas da Cruz. 2010. “Respiratory Disease and the Role of Oral Bacteria.” *Journal of Oral Microbiology* 2 (December). https://doi.org/10.3402/jom.v2i0.5811.
Granli, T., R. Dahl, P. Brodin, and O. C. Bøckman. 1989. “Nitrate and Nitrite Concentrations in Human Saliva: Variations with Salivary Flow-Rate.” *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association* 27 (10): 675–80.
Grasemann, H., I. Ioannidis, R. P. Tomkiewicz, H. de Groot, B. K. Rubin, and F. Ratjen. 1998. “Nitric Oxide Metabolites in Cystic Fibrosis Lung Disease.” *Archives of Disease in Childhood* 78 (1): 49–53.
Griffen, Ann L., Clifford J. Beall, James H. Campbell, Noah D. Firestone, Purnima S. Kumar, Zamin K. Yang, Mircea Podar, and Eugene J. Leys. 2012. “Distinct and Complex Bacterial Profiles in Human Periodontitis and Health Revealed by 16S Pyrosequencing.” *The ISME Journal* 6 (6): 1176–85.
Gross, Erin L., Eugene J. Leys, Stephen R. Gasparovich, Noah D. Firestone, Judith A. Schwartzbaum, Daniel A. Janies, Kashmira Asnani, and Ann L. Griffen. 2010. “Bacterial 16S Sequence Analysis of Severe Caries in Young Permanent Teeth.” *Journal of Clinical Microbiology* 48 (11): 4121–28.
Gura, Trisha. 2013. “Citizen Science: Amateur Experts.” *Nature*

496 (7444): 259–61.
Gu, Zuguang, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. 2014. “Circlize Implements and Enhances Circular Visualization in R.” *Bioinformatics* 30 (19): 2811–12.
Hajishengallis, E., Y. Parsaei, M. I. Klein, and H. Koo. 2017. “Advances in the Microbial Etiology and Pathogenesis of Early Childhood Caries.” *Molecular Oral Microbiology* 32 (1): 24–34.
Hajishengallis, George. 2014. “Aging and Its Impact on Innate Immunity and Inflammation: Implications for Periodontitis.” *Journal of Oral Biosciences / JAOB, Japanese Association for Oral Biology* 56 (1): 30–37.
Hamady, Micah, and Rob Knight. 2009. “Microbial Community Profiling for Human Microbiome Projects: Tools, Techniques, and Challenges.” *Genome Research* 19 (7): 1141–52.
Handsley-Davis, Matilda, Lisa Jamieson, Kostas Kapellas, Joanne Hedges, and Laura S. Weyrich. 2020. “The Role of the Oral Microbiota in Chronic Non-Communicable Disease and Its Relevance to the Indigenous Health Gap in Australia.” *BMC Oral Health* 20 (1): 327.
Han, Xiang Y., and Enevold Falsen. 2005. “Characterization of Oral Strains of Cardiobacterium Valvarum and Emended Description of the Organism.” *Journal of Clinical Microbiology* 43 (5): 2370–74.
Han, Y. W., W. Shi, G. T. Huang, S. Kinder Haake, N. H. Park, H. Kuramitsu, and R. J. Genco. 2000. “Interactions between Periodontal Bacteria and Human Oral Epithelial Cells: Fusobacterium Nucleatum Adheres to and Invades Epithelial Cells.” *Infection and Immunity* 68 (6): 3140–46.
Haraguchi, A., M. Miura, O. Fujise, T. Hamachi, and F. Nishimura. 2014. “Porphyromonas Gingivalis Gingipain Is Involved in the Detachment and Aggregation of Aggregatibacter Actinomycetemcomitans Biofilm.” *Molecular Oral Microbiology* 29 (3): 131–43.
Haraszthy, Violet I., Joseph J. Zambon, Prem K. Sreenivasan, Margaret M. Zambon, Doralee Gerber, Rodrigo Rego, and Carol Parker. 2007. “Identification of Oral Bacterial Species Associated with Halitosis.” *Journal of the American Dental Association* 138 (8): 1113–20.
Haubek, Dorte. 2010. “The Highly Leukotoxic JP2 Clone of Aggregatibacter Actinomycetemcomitans: Evolutionary Aspects, Epidemiology and Etiological Role in Aggressive

Periodontitis.” *APMIS. Supplementum*, no. 130 (September): 1–53.
Hauser, Alan R., Manu Jain, Maskit Bar-Meir, and Susanna A. McColley. 2011. “Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis.” *Clinical Microbiology Reviews* 24 (1): 29–70.
He, Jinzhi, Yan Li, Yangpei Cao, Jin Xue, and Xuedong Zhou. 2015. “The Oral Microbiome Diversity and Its Relation to Human Diseases.” *Folia Microbiologica* 60 (1): 69–80.
Hendershot, Tabitha, Huaqin Pan, Jonathan Haines, William R. Harlan, Mary L. Marazita, Catherine A. McCarty, Erin M. Ramos, and Carol M. Hamilton. 2015. “Using the PhenX Toolkit to Add Standard Measures to a Study.” *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]* 86 (July): 1.21.1–1.21.17.
Henne, Karsten, Anke Rheinberg, Beate Melzer-Krick, and Georg Conrads. 2015. “Aciduric Microbial Taxa Including Scardovia Wiggsiae and Bifidobacterium Spp. in Caries and Caries Free Subjects.” *Anaerobe* 35 (Pt A): 60–65.
Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. “The Weirdest People in the World?” *The Behavioral and Brain Sciences* 33 (2-3): 61–83; discussion 83–135.
Henskens, Y. M., P. A. van den Keijbus, E. C. Veerman, G. A. Van der Weijden, M. F. Timmerman, C. M. Snoek, U. Van der Velden, and A. V. Nieuw Amerongen. 1996. “Protein Composition of Whole and Parotid Saliva in Healthy and Periodontitis Subjects. Determination of Cystatins, Albumin, Amylase and IgA.” *Journal of Periodontal Research* 31 (1): 57–65.
Hérivaux, Anaïs, Jesse R. Willis, Toine Mercier, Katrien Lagrou, Samuel M. Gonçalves, Relber A. Gonçales, Johan Maertens, Agostinho Carvalho, Toni Gabaldón, and Cristina Cunha. 2021. “Lung Microbiota Predict Invasive Pulmonary Aspergillosis and Its Outcome in Immunocompromised Patients.” *Thorax*, June. https://doi.org/10.1136/thoraxjnl-2020-216179.
Herman, Katarzyna, Małgorzata Kowalczyk-Zając, and Tomasz Pytrus. 2017. “Oral Cavity Health among Cystic Fibrosis Patients: Literature Overview.” *Advances in Clinical and Experimental Medicine: Official Organ Wroclaw Medical University* 26 (7): 1147–53.

Hernández, Miguel, Paloma Planells, Eva Martínez, Alex Mira, and Miguel Carda-Diéguez. 2020. "Microbiology of Molar-Incisor Hypomineralization Lesions. A Pilot Study." *Journal of Oral Microbiology* 12 (1): 1766166.

Heyer, Robert, Kay Schallert, Roman Zoun, Beatrice Becher, Gunter Saake, and Dirk Benndorf. 2017. "Challenges and Perspectives of Metaproteomic Data Analysis." *Journal of Biotechnology* 261 (November): 24–36.

Hibbett, David S., and John W. Taylor. 2013. "Fungal Systematics: Is a New Age of Enlightenment at Hand?" *Nature Reviews. Microbiology* 11 (2): 129–33.

Ho, Jessica Y., and Arun S. Hendi. 2018. "Recent Trends in Life Expectancy across High Income Countries: Retrospective Observational Study." *BMJ* 362 (August): k2562.

Hong, B. Y., A. Hoare, A. Cardenas, A. K. Dupuy, L. Choquette, A. L. Salner, P. K. Schauer, et al. 2020. "The Salivary Mycobiome Contains 2 Ecologically Distinct Mycotypes." *Journal of Dental Research* 99 (6): 730–38.

Horvath, Steve, Paolo Garagnani, Maria Giulia Bacalini, Chiara Pirazzini, Stefano Salvioli, Davide Gentilini, Anna Maria Di Blasio, et al. 2015. "Accelerated Epigenetic Aging in Down Syndrome." *Aging Cell* 14 (3): 491–95.

Horz, Hans-Peter. 2015. "Archaeal Lineages within the Human Microbiome: Absent, Rare or Elusive?" *Life* 5 (2): 1333–45.

Huang, Xuelian, Sara R. Palmer, Sang-Joon Ahn, Vincent P. Richards, Matthew L. Williams, Marcelle M. Nascimento, and Robert A. Burne. 2016. "A Highly Arginolytic Streptococcus Species That Potently Antagonizes Streptococcus Mutans." *Applied and Environmental Microbiology* 82 (7): 2187–2201.

Huggett, Nick. 2019. "Zeno's Paradoxes." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/paradox-zeno/.

Human Microbiome Project Consortium. 2012. "Structure, Function and Diversity of the Healthy Human Microbiome." *Nature* 486 (7402): 207–14.

Hurley, Eimear, Maurice P. J. Barrett, Martin Kinirons, Helen Whelton, C. Anthony Ryan, Catherine Stanton, Hugh M. B. Harris, and Paul W. O'Toole. 2019. "Comparison of the Salivary and Dentinal Microbiome of Children with Severe-Early Childhood Caries to the Salivary Microbiome of Caries-

Free Children." *BMC Oral Health* 19 (1): 13.
Huttenhower, Curtis, Rob Knight, C. Titus Brown, J. Gregory Caporaso, Jose C. Clemente, Dirk Gevers, Eric A. Franzosa, et al. 2014. "Advancing the Microbiome Research Community." *Cell* 159 (2): 227–30.
Hyde, Embriette R., Fernando Andrade, Zalman Vaksman, Kavitha Parthasarathy, Hong Jiang, Deepa K. Parthasarathy, Ashley C. Torregrossa, et al. 2014. "Metagenomic Analysis of Nitrate-Reducing Bacteria in the Oral Cavity: Implications for Nitric Oxide Homeostasis." *PloS One* 9 (3): e88645.
Iebba, Valerio, Valentina Totino, Antonella Gagliardi, Floriana Santangelo, Fatima Cacciotti, Maria Trancassini, Carlo Mancini, et al. 2016. "Eubiosis and Dysbiosis: The Two Sides of the Microbiota." *The New Microbiologica* 39 (1): 1–12.
Iheozor-Ejiofor, Zipporah, Helen V. Worthington, Tanya Walsh, Lucy O'Malley, Jan E. Clarkson, Richard Macey, Rahul Alam, Peter Tugwell, Vivian Welch, and Anne-Marie Glenny. 2015. "Water Fluoridation for the Prevention of Dental Caries." *Cochrane Database of Systematic Reviews* , no. 6 (June): CD010856.
Integrative HMP (iHMP) Research Network Consortium. 2014. "The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease." *Cell Host & Microbe* 16 (3): 276–89.
Jagathrakshakan, Sri Nisha, Raghavendra Jayesh Sethumadhava, Dhaval Tushar Mehta, and Arvind Ramanathan. 2015. "16S rRNA Gene-Based Metagenomic Analysis Identifies a Novel Bacterial Co-Prevalence Pattern in Dental Caries." *European Journal of Dentistry* 9 (1): 127.
Jakubovics, Nicholas S. 2015. "Intermicrobial Interactions as a Driver for Community Composition and Stratification of Oral Biofilms." *Journal of Molecular Biology* 427 (23): 3662–75.
Janus, M. M., W. Crielaard, C. M. C. Volgenant, M. H. van der Veen, B. W. Brandt, and B. P. Krom. 2017. "Candida Albicans Alters the Bacterial Microbiome of Early in Vitro Oral Biofilms." *Journal of Oral Microbiology* 9 (1): 1270613.
Jari, Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, H. M. H. Stevens, and Helene Wagner. 2016. "Vegan: Community Ecology Package." *R Package Version* 2: 3–5.

Jayaprakash, Kartheyaene, Hazem Khalaf, and Torbjörn Bengtsson. 2014. “Gingipains from Porphyromonas Gingivalis Play a Significant Role in Induction and Regulation of CXCL8 in THP-1 Cells.” *BMC Microbiology* 14 (July): 193.

Jeffery, Ian B., Marcus J. Claesson, Paul W. O’Toole, and Fergus Shanahan. 2012. “Categorization of the Gut Microbiota: Enterotypes or Gradients?” *Nature Reviews. Microbiology* 10 (9): 591–92.

Jewtuchowicz, V. M., M. T. Mujica, M. I. Brusca, N. Sordelli, M. C. Malzone, S. J. Pola, C. A. Iovannitti, and A. C. Rosa. 2008. “Phenotypic and Genotypic Identification of Candida Dubliniensis from Subgingival Sites in Immunocompetent Subjects in Argentina.” *Oral Microbiology and Immunology* 23 (6): 505–9.

Jiang, Qian, Jia Liu, Liang Chen, Ning Gan, and Deqin Yang. 2018. “The Oral Microbiome in the Elderly With Dental Caries and Health.” *Frontiers in Cellular and Infection Microbiology* 8: 442.

Jiang, Ruochen, Wei Vivian Li, and Jingyi Jessica Li. 2021. “mbImpute: An Accurate and Robust Imputation Method for Microbiome Data.” *Genome Biology* 22 (1): 192.

Jiang, Wen, Zongxin Ling, Xiaolong Lin, Yadong Chen, Jie Zhang, Jinjin Yu, Charlie Xiang, and Hui Chen. 2014. “Pyrosequencing Analysis of Oral Microbiota Shifting in Various Caries States in Childhood.” *Microbial Ecology* 67 (4): 962–69.

Jindal, Surinder K. 2018. “Koch’s Postulates - Pitfalls and Relevance in the 21st Century.” *The Indian Journal of Tuberculosis* 65 (1): 6–7.

Jing, Dian, Jin Hao, Yu Shen, Ge Tang, Lei Lei, and Zhihe Zhao. 2019. “Effect of Fixed Orthodontic Treatment on Oral Microbiota and Salivary Proteins.” *Experimental and Therapeutic Medicine* 17 (5): 4237–43.

Johansson, I., E. Witkowska, B. Kaveh, P. Lif Holgerson, and A. C. R. Tanner. 2016. “The Microbiome in Populations with a Low and High Prevalence of Caries.” *Journal of Dental Research* 95 (1): 80–86.

Jorth, Peter, Keith H. Turner, Pinar Gumus, Nejat Nizam, Nurcan Buduneli, and Marvin Whiteley. 2014. “Metatranscriptomics of the Human Oral Microbiome during Health and Disease.” *mBio* 5 (2): e01012–14.

Jose, Jubin Easo, Sridevi Padmanabhan, and Arun B. Chitharanjan. 2013. “Systemic Consumption of Probiotic Curd

and Use of Probiotic Toothpaste to Reduce Streptococcus Mutans in Plaque around Orthodontic Brackets." *American Journal of Orthodontics and Dentofacial Orthopedics: Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics* 144 (1): 67–72.

Jowett, Benjamin. 1871. "Sophist." In *The Dialogues of Plato*, edited by Benjamin Jowett, 1522–23.

Kageyama, Shinya, Toru Takeshita, Mikari Asakawa, Yukie Shibata, Kenji Takeuchi, Wataru Yamanaka, and Yoshihisa Yamashita. 2017. "Relative Abundance of Total Subgingival Plaque-Specific Bacteria in Salivary Microbiota Reflects the Overall Periodontal Condition in Patients with Periodontitis." *PloS One* 12 (4): e0174782.

Kashyap, Purna C., Nicholas Chia, Heidi Nelson, Eran Segal, and Eran Elinav. 2017. "Microbiome at the Frontier of Personalized Medicine." *Mayo Clinic Proceedings. Mayo Clinic* 92 (12): 1855–64.

Kataoka, Hideo, Makoto Taniguchi, Haruka Fukamachi, Takafumi Arimoto, Hirobumi Morisaki, and Hirotaka Kuwata. 2014. "Rothia Dentocariosa Induces TNF-Alpha Production in a TLR2-Dependent Manner." *Pathogens and Disease* 71 (1): 65–68.

Kato, Ikuko, Adrian A. Vasquez, Gregory Moyerbrailean, Susan Land, Jun Sun, Ho-Sheng Lin, and Jeffrey L. Ram. 2016. "Oral Microbiome and History of Smoking and Colorectal Cancer." *Journal of Epidemiological Research* 2 (2): 92–101.

Kaul, Abhishek, Siddhartha Mandal, Ori Davidov, and Shyamal D. Peddada. 2017. "Analysis of Microbiome Data in the Presence of Excess Zeros." *Frontiers in Microbiology* 8 (November): 2114.

Kembel, Steven W., Martin Wu, Jonathan A. Eisen, and Jessica L. Green. 2012. "Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance." *PLoS Computational Biology* 8 (10): e1002743.

Kerley, Conor P., Emma Kilbride, Peter Greally, and Basil Elnazir. 2016. "Dietary Nitrate Acutely and Markedly Increased Exhaled Nitric Oxide in a Cystic Fibrosis Case." *Clinical Medicine & Research* 14 (3-4): 151–55.

Khocht, Ahmed. 2011. "Down Syndrome and Periodontal Disease." In *Genetics and Etiology of Down Syndrome*, edited by

Subrata Dey. InTech.
Khocht, A., T. Yaskell, M. Janal, B. F. Turner, T. E. Rams, A. D. Haffajee, and S. S. Socransky. 2012. “Subgingival Microbiota in Adult Down Syndrome Periodontitis.” *Journal of Periodontal Research* 47 (4): 500–507.
Kianoush, Nima, Christina J. Adler, Ky-Anh T. Nguyen, Gina V. Browne, Mary Simonian, and Neil Hunter. 2014. “Bacterial Profile of Dentine Caries and the Impact of pH on Bacterial Population Diversity.” *PloS One* 9 (3): e92940.
Kilian, M., I. L. C. Chapple, M. Hannig, P. D. Marsh, V. Meuric, A. M. L. Pedersen, M. S. Tonetti, W. G. Wade, and E. Zaura. 2016. “The Oral Microbiome - an Update for Oral Healthcare Professionals.” *British Dental Journal* 221 (10): 657–66.
Kilian, Mogens. 2018. “The Oral Microbiome - Friend or Foe?” *European Journal of Oral Sciences* 126 Suppl 1 (October): 5–12.
Kim, Ga-Yeon, and Chong Heon Lee. 2015. “Antimicrobial Susceptibility and Pathogenic Genes of Staphylococcus Aureus Isolated from the Oral Cavity of Patients with Periodontitis.” *Journal of Periodontal & Implant Science* 45 (6): 223–28.
Kinane, Denis F., Marcello P. Riggio, Katie F. Walker, Duncan MacKenzie, and Barbara Shearer. 2005. “Bacteraemia Following Periodontal Procedures.” *Journal of Clinical Periodontology* 32 (7): 708–13.
Kistler, James O., Veronica Booth, David J. Bradshaw, and William G. Wade. 2013. “Bacterial Community Development in Experimental Gingivitis.” *PloS One* 8 (8): e71227.
Klaus, Katharina, Johanna Eichenauer, Rhea Sprenger, and Sabine Ruf. 2016. “Oral Microbiota Carriage in Patients with Multibracket Appliance in Relation to the Quality of Oral Hygiene.” *Head & Face Medicine* 12 (1): 28.
Knapp, J. S. 1988. “Historical Perspectives and Identification of Neisseria and Related Species.” *Clinical Microbiology Reviews* 1 (4): 415–31.
Knight, Rob, Alison Vrbanac, Bryn C. Taylor, Alexander Aksenov, Chris Callewaert, Justine Debelius, Antonio Gonzalez, et al. 2018. “Best Practices for Analysing Microbiomes.” *Nature Reviews. Microbiology* 16 (7): 410–22.
Knights, Dan, Tonya L. Ward, Christopher E. McKinlay, Hannah Miller, Antonio Gonzalez, Daniel McDonald, and Rob Knight.

2014. “Rethinking ‘Enterotypes.’” *Cell Host & Microbe* 16 (4): 433–37.

Kõll□Klais, P., R. Mändar, and E. Leibur. 2005. “Oral Lactobacilli in Chronic Periodontitis and Periodontal Health: Species Composition and Antimicrobial Activity.” *Oral Microbiology and Immunology*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1399-302X.2005.00239.x.

Konig, Maximilian F., Alizay S. Paracha, Malini Moni, Clifton O. Bingham 3rd, and Felipe Andrade. 2015. “Defining the Role of Porphyromonas Gingivalis Peptidylarginine Deiminase (PPAD) in Rheumatoid Arthritis through the Study of PPAD Biology.” *Annals of the Rheumatic Diseases* 74 (11): 2054–61.

Koo, Hyun, and William H. Bowen. 2014. “Candida Albicans and Streptococcus Mutans: A Potential Synergistic Alliance to Cause Virulent Tooth Decay in Children.” *Future Microbiology* 9 (12): 1295–97.

Koopman, Jessica E., Nicoline C. W. van der Kaaij, Mark J. Buijs, Yassaman Elyassi, Monique H. van der Veen, Wim Crielaard, Jacob M. Ten Cate, and Egija Zaura. 2015. “The Effect of Fixed Orthodontic Appliances and Fluoride Mouthwash on the Oral Microbiome of Adolescents - A Randomized Controlled Clinical Trial.” *PloS One* 10 (9): e0137318.

Koren, Omry, Dan Knights, Antonio Gonzalez, Levi Waldron, Nicola Segata, Rob Knight, Curtis Huttenhower, and Ruth E. Ley. 2013. “A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets.” *PLoS Computational Biology* 9 (1): e1002863.

Kornman, Kenneth S. 2008. “Mapping the Pathogenesis of Periodontitis: A New Look.” *Journal of Periodontology* 79 (8 Suppl): 1560–68.

Koskinen, Kaisa, Manuela R. Pausan, Alexandra K. Perras, Michael Beck, Corinna Bang, Maximilian Mora, Anke Schilhabel, Ruth Schmitz, and Christine Moissl-Eichinger. 2017. “First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin.” *mBio* 8 (6). https://doi.org/10.1128/mBio.00824-17.

Kostic, Aleksandar D., Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M. Earl, Akinyemi I.

Ojesina, et al. 2012. “Genomic Analysis Identifies Association of Fusobacterium with Colorectal Carcinoma.” *Genome Research* 22 (2): 292–98.
Koyanagi, Tatsuro, Mitsuo Sakamoto, Yasuo Takeuchi, Noriko Maruyama, Moriya Ohkuma, and Yuichi Izumi. 2013. “Comprehensive Microbiological Findings in Peri-Implantitis and Periodontitis.” *Journal of Clinical Periodontology* 40 (3): 218–26.
Krom, B. P., S. Kidwai, and J. M. Ten Cate. 2014. “Candida and Other Fungal Species: Forgotten Players of Healthy Oral Microbiota.” *Journal of Dental Research* 93 (5): 445–51.
Kumar, Manish, Boyang Ji, Karsten Zengler, and Jens Nielsen. 2019. “Modelling Approaches for Studying the Microbiome.” *Nature Microbiology* 4 (8): 1253–67.
Kumar, Purnima S., Ann L. Griffen, Melvin L. Moeschberger, and Eugene J. Leys. 2005. “Identification of Candidate Periodontal Pathogens and Beneficial Species by Quantitative 16S Clonal Analysis.” *Journal of Clinical Microbiology* 43 (8): 3944–55.
Kunzelmann, Karl, Rainer Schreiber, and Hans Beat Hadorn. 2017. “Bicarbonate in Cystic Fibrosis.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 16 (6): 653–62.
Kurkivuori, Johanna, Ville Salaspuro, Pertti Kaihovaara, Kirsti Kari, Riina Rautemaa, Lisa Grönroos, Jukka H. Meurman, and Mikko Salaspuro. 2007. “Acetaldehyde Production from Ethanol by Oral Streptococci.” *Oral Oncology* 43 (2): 181–86.
Kurtz, Zachary D., Christian L. Müller, Emily R. Miraldi, Dan R. Littman, Martin J. Blaser, and Richard A. Bonneau. 2015. “Sparse and Compositionally Robust Inference of Microbial Ecological Networks.” *PLoS Computational Biology* 11 (5): e1004226.
Lambiase, Antonietta, Mariassunta Del Pezzo, Valeria Raia, Angela Sepe, Pasqualina Ferri, and Fabio Rossano. 2007. “Chryseobacterium Respiratory Tract Infections in Patients with Cystic Fibrosis.” *The Journal of Infection* 55 (6): 518–23.
LaMonte, Michael J., Robert J. Genco, Michael J. Buck, Daniel I. McSkimming, Lu Li, Kathleen M. Hovey, Christopher A. Andrews, et al. 2019. “Composition and Diversity of the Subgingival Microbiome and Its Relationship with Age in Postmenopausal Women: An Epidemiologic Investigation.” *BMC Oral Health* 19 (1): 246.

LaMonte, Michael J., Robert J. Genco, Wei Zheng, Daniel I. McSkimming, Christopher A. Andrews, Kathleen M. Hovey, Lu Li, et al. 2018. “Substantial Differences in the Subgingival Microbiome Measured by 16S Metagenomics According to Periodontitis Status in Older Women.” *Dental Journal* 6 (4). https://doi.org/10.3390/dj6040058.

Langille, Morgan G. I., Jesse Zaneveld, J. Gregory Caporaso, Daniel McDonald, Dan Knights, Joshua A. Reyes, Jose C. Clemente, et al. 2013. “Predictive Functional Profiling of Microbial Communities Using 16S rRNA Marker Gene Sequences.” *Nature Biotechnology* 31 (9): 814–21.

Lassalle, Florent, Matteo Spagnoletti, Matteo Fumagalli, Liam Shaw, Mark Dyble, Catherine Walker, Mark G. Thomas, Andrea Bamberg Migliano, and Francois Balloux. 2018. “Oral Microbiomes from Hunter-Gatherers and Traditional Farmers Reveal Shifts in Commensal Balance and Pathogen Load Linked to Diet.” *Molecular Ecology* 27 (1): 182–95.

Laubichler, Manfred D., and Jürgen Renn. 2015. “Extended Evolution: A Conceptual Framework for Integrating Regulatory Networks and Niche Construction.” *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* 324 (7): 565–77.

Lax, Alistair J. 2005. “Opinion: Bacterial Toxins and Cancer--a Case to Answer?” *Nature Reviews. Microbiology* 3 (4): 343–49.

Layeghifard, Mehdi, Hannah Li, Pauline W. Wang, Sylva L. Donaldson, Bryan Coburn, Shawn T. Clark, Julio Diaz Caballero, et al. 2019. “Microbiome Networks and Change-Point Analysis Reveal Key Community Changes Associated with Cystic Fibrosis Pulmonary Exacerbations.” *Npj Biofilms and Microbiomes* 5 (1): 1–12.

Lenartova, Magdalena, Barbora Tesinska, Tatjana Janatova, Ondrej Hrebicek, Jaroslav Mysak, Jiri Janata, and Lucie Najmanova. 2021. “The Oral Microbiome in Periodontal Health.” *Frontiers in Cellular and Infection Microbiology* 11 (March): 629723.

Leo, L., and S. Shetty. 2017. *Microbiome: Microbiome R Package*. Github. https://github.com/microbiome/microbiome.

Lepesqueur, Laura Soares Souto, Marcia Hiromi Tanaka, Gabriela de Morais Gouvêa Lima, Sonia Mayumi Chiba, Adolfo José Mota, Samára Ferreira Santos, and Cristiane Yumi Koga-Ito.

2020. “Oral Prevalence and Antifungal Susceptibility of Candida Species in Cystic Fibrosis Patients.” *Archives of Oral Biology* 116 (August): 104772.

Lepp, Paul W., Mary M. Brinig, Cleber C. Ouverney, Katherine Palm, Gary C. Armitage, and David A. Relman. 2004. “Methanogenic Archaea and Human Periodontal Disease.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (16): 6176–81.

Leung, Marcus H. Y., David Wilkins, and Patrick K. H. Lee. 2015. “Insights into the Pan-Microbiome: Skin Microbial Communities of Chinese Individuals Differ from Other Racial Groups.” *Scientific Reports* 5 (July): 11845.

Lif Holgerson, Pernilla, Carina Öhman, Agneta Rönnlund, and Ingegerd Johansson. 2015. “Maturation of Oral Microbiota in Children with or without Dental Caries.” *PloS One* 10 (5): e0128534.

Li, Kelvin, Monika Bihan, and Barbara A. Methé. 2013. “Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome.” *PloS One* 8 (5): e63139.

Li, Kelvin, Monika Bihan, Shibu Yooseph, and Barbara A. Methé. 2012. “Analyses of the Microbial Diversity across the Human Microbiome.” *PloS One* 7 (6): e32118.

Lim, Yan Wei, Robert Schmieder, Matthew Haynes, Mike Furlan, T. David Matthews, Katrine Whiteson, Stephen J. Poole, et al. 2013. “Mechanistic Model of Rothia Mucilaginosa Adaptation toward Persistence in the CF Lung, Based on a Genome Reconstructed from Metagenomic Data.” *PloS One* 8 (5): e64285.

Lim, Yenkai, Naoki Fukuma, Makrina Totsika, Liz Kenny, Mark Morrison, and Chamindie Punyadeera. 2018. “The Performance of an Oral Microbiome Biomarker Panel in Predicting Oral Cavity and Oropharyngeal Cancers.” *Frontiers in Cellular and Infection Microbiology* 8 (August): 267.

Lim, Yenkai, Makrina Totsika, Mark Morrison, and Chamindie Punyadeera. 2017. “The Saliva Microbiome Profiles Are Minimally Affected by Collection Method or DNA Extraction Protocols.” *Scientific Reports* 7 (1): 8523.

Lin, Huang, and Shyamal Das Peddada. 2020. “Analysis of Microbial Compositions: A Review of Normalization and Differential Abundance Analysis.” *NPJ Biofilms and Microbiomes* 6 (1): 60.

Liu, Ailing, Hong Lv, Hongying Wang, Hong Yang, Yue Li, and Jiaming Qian. 2020. “Aging Increases the Severity of Colitis and the Related Changes to the Gut Barrier and Gut Microbiota in Humans and Mice.” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 75 (7): 1284–92.

Liu, Bo, Lina L. Faller, Niels Klitgord, Varun Mazumdar, Mohammad Ghodsi, Daniel D. Sommer, Theodore R. Gibbons, et al. 2012. “Deep Sequencing of the Oral Microbiome Reveals Signatures of Periodontal Disease.” *PloS One* 7 (6): e37919.

Li, Yanhui, Cheng-Gang Zou, Yu Fu, Yanhong Li, Qing Zhou, Bo Liu, Zhigang Zhang, and Juan Liu. 2016. “Oral Microbial Community Typing of Caries and Pigment in Primary Dentition.” *BMC Genomics* 17 (August): 558.

Li, Y., Y. Ge, D. Saxena, and P. W. Caufield. 2007. “Genetic Profiling of the Oral Microbiota Associated with Severe Early-Childhood Caries.” *Journal of Clinical Microbiology* 45 (1): 81–87.

Li, Yuchao, Xuexin Tan, Xida Zhao, Zhongfei Xu, Wei Dai, Weiyi Duan, Shaohui Huang, et al. 2020. “Composition and Function of Oral Microbiota between Gingival Squamous Cell Carcinoma and Periodontitis.” *Oral Oncology* 107 (May): 104710.

Loberto, Jussara Cia S., Clélia Ap de Paiva Martins, Silvana S. Ferreira dos Santos, José Roberto Cortelli, and Antonio Olavo Cardoso Jorge. 2004. “Staphylococcus Spp. in the Oral Cavity and Periodontal Pockets of Chronic Periodontitis Patients.” *Brazilian Journal of Microbiology: [publication of the Brazilian Society for Microbiology]* 35 (1-2): 64–68.

Lockhart, Peter B., Michael T. Brennan, Howell C. Sasser, Philip C. Fox, Bruce J. Paster, and Farah K. Bahrani-Mougeot. 2008. “Bacteremia Associated with Toothbrushing and Dental Extraction.” *Circulation* 117 (24): 3118–25.

Lockhart, Peter B., Michael T. Brennan, Martin Thornhill, Bryan S. Michalowicz, Jenene Noll, Farah K. Bahrani-Mougeot, and Howell C. Sasser. 2009. “Poor Oral Hygiene as a Risk Factor for Infective Endocarditis-Related Bacteremia.” *Journal of the American Dental Association* 140 (10): 1238–44.

Loozen, Gitte, Nico Boon, Martine Pauwels, Vera Slomka, Esteban Rodrigues Herrero, Marc Quirynen, and Wim Teughels. 2015. “Effect of Bdellovibrio Bacteriovorus HD100 on Multispecies

Oral Communities.” *Anaerobe* 35 (Pt A): 45–53.
López-López, Arantxa, Anny Camelo-Castillo, María D. Ferrer, Áurea Simon-Soro, and Alex Mira. 2017. “Health-Associated Niche Inhabitants as Oral Probiotics: The Case of Streptococcus Dentisani.” *Frontiers in Microbiology* 8 (March): 379.
Losada, Patricia Moran, Philippe Chouvarine, Marie Dorda, Silke Hedtfeld, Samira Mielke, Angela Schulz, Lutz Wiehlmann, and Burkhard Tümmler. 2016. “The Cystic Fibrosis Lower Airways Microbial Metagenome.” *ERJ Open Research* 2 (2). https://doi.org/10.1183/23120541.00096-2015.
Louca, Stilianos, Michael Doebeli, and Laura Wegener Parfrey. 2018. “Correcting for 16S rRNA Gene Copy Numbers in Microbiome Surveys Remains an Unsolved Problem.” *Microbiome* 6 (1): 41.
Lourenço, Talita Gomes Baêta, Débora Heller, Carina Maciel Silva-Boghossian, Sean L. Cotton, Bruce J. Paster, and Ana Paula Vieira Colombo. 2014. “Microbial Signature Profiles of Periodontally Healthy and Diseased Patients.” *Journal of Clinical Periodontology* 41 (11): 1027–36.
Lozano Moraga, Carla Paola, Gonzalo Andrés Rodríguez Martínez, Claudia Andrea Lefimil Puente, Irene Cecilia Morales Bozo, and Blanca Regina Urzúa Orellana. 2017. “Prevalence of Candida Albicans and Carriage of Candida Non-Albicans in the Saliva of Preschool Children, according to Their Caries Status.” *Acta Odontologica Scandinavica* 75 (1): 30–35.
Lucchese, Alessandra, Lars Bondemark, Marta Marcolina, and Maurizio Manuelli. 2018. “Changes in Oral Microbiota due to Orthodontic Appliances: A Systematic Review.” *Journal of Oral Microbiology* 10 (1): 1476645.
Lu, Haifeng, Zhigang Ren, Ang Li, Jinyou Li, Shaoyan Xu, Hua Zhang, Jianwen Jiang, et al. 2019. “Tongue Coating Microbiome Data Distinguish Patients with Pancreatic Head Cancer from Healthy Controls.” *Journal of Oral Microbiology* 11 (1): 1563409.
Lukowski, Angela F., Helen M. Milojevich, and Lauren Eales. 2019. “Cognitive Functioning in Children with Down Syndrome: Current Knowledge and Future Directions.” *Advances in Child Development and Behavior* 56 (February): 257–89.
Lundmark, Anna, Yue O. O. Hu, Mikael Huss, Gunnar Johannsen, Anders F. Andersson, and Tülay Yucel-Lindberg. 2019.

“Identification of Salivary Microbiota and Its Association With Host Inflammatory Mediators in Periodontitis.” *Frontiers in Cellular and Infection Microbiology* 9 (June): 216.

Lynch, Kate E., Emily C. Parke, and Maureen A. O’Malley. 2019. “How Causal Are Microbiomes? A Comparison with the Helicobacter Pylori Explanation of Ulcers.” *Biology & Philosophy* 34 (6): 62.

Maeda, Yasunori, J. Stuart Elborn, Michael D. Parkins, James Reihill, Colin E. Goldsmith, Wilson A. Coulter, Charlene Mason, et al. 2011. “Population Structure and Characterization of Viridans Group Streptococci (VGS) Including Streptococcus Pneumoniae Isolated from Adult Patients with Cystic Fibrosis (CF).” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 10 (2): 133–39.

Maffei, Vincent J., Sangkyu Kim, Eugene Blanchard 4th, Meng Luo, S. Michal Jazwinski, Christopher M. Taylor, and David A. Welsh. 2017. “Biological Aging and the Human Gut Microbiota.” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 72 (11): 1474–82.

Mager, D. L., A. D. Haffajee, P. M. Devlin, C. M. Norris, M. R. Posner, and J. M. Goodson. 2005. “The Salivary Microbiota as a Diagnostic Indicator of Oral Cancer: A Descriptive, Non-Randomized Study of Cancer-Free and Oral Squamous Cell Carcinoma Subjects.” *Journal of Translational Medicine* 3 (July): 27.

Mahajan, Ajay, Baljeet Singh, Divya Kashyap, Amit Kumar, and Poonam Mahajan. 2013. “Interspecies Communication and Periodontal Disease.” *TheScientificWorldJournal* 2013 (December): 765434.

Malla, Muneer Ahmad, Anamika Dubey, Ashwani Kumar, Shweta Yadav, Abeer Hashem, and Elsayed Fathi Abd Allah. 2018. “Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment.” *Frontiers in Immunology* 9: 2868.

Mallick, Himel, Siyuan Ma, Eric A. Franzosa, Tommi Vatanen, Xochitl C. Morgan, and Curtis Huttenhower. 2017. “Experimental Design and Quantitative Analysis of Microbial Community Multiomics.” *Genome Biology* 18 (1): 228.

Mandal, Siddhartha, Will Van Treuren, Richard A. White, Merete Eggesbø, Rob Knight, and Shyamal D. Peddada. 2015. “Analysis of Composition of Microbiomes: A Novel Method for

Studying Microbial Composition." *Microbial Ecology in Health and Disease* 26 (May): 27663.
Mantzourani, M., S. C. Gilbert, H. N. H. Sulong, E. C. Sheehy, S. Tank, M. Fenlon, and D. Beighton. 2009. "The Isolation of Bifidobacteria from Occlusal Carious Lesions in Children and Adults." *Caries Research* 43 (4): 308–13.
Maranhão, Fernanda Cristina de Albuquerque, Nayara Maciel Mendonça, Thamires Costa Teixeira, Gilvonete Alves da Costa Sobrinha Lages, Jacqueline Araújo de Melo, Carlos Guilherme Gaelzer Porciuncula, Eurípedes Alves da Silva Filho, and Denise Maria Wanderlei Silva. 2020. "Molecular Identification of Candida Species in the Oral Microbiota of Individuals with Down Syndrome: A Case-Control Study." *Mycopathologia* 185 (3): 537–43.
Maraver, Francisco, Isidro Vitoria, José Manuel Almerich-Silla, and Francisco Armijo. 2015. "[Fluoride content of bottled natural mineral waters in Spain and prevention of dental caries]." *Atencion primaria / Sociedad Espanola de Medicina de Familia y Comunitaria* 47 (1): 15–24.
Maraver, Francisco, Isidro Vitoria, Cíntia Ferreira-Pêgo, Francisco Armijo, and Jordi Salas-Salvadó. 2015. "Magnesium in Tap and Bottled Mineral Water in Spain and Its Contribution to Nutritional Recommendations." *Nutricion Hospitalaria: Organo Oficial de La Sociedad Espanola de Nutricion Parenteral Y Enteral* 31 (5): 2297–2312.
Marchesi, Julian R., and Jacques Ravel. 2015. "The Vocabulary of Microbiome Research: A Proposal." *Microbiome* 3 (July): 31.
Marcos-Zambrano, Laura Judith, Kanita Karaduzovic-Hadziabdic, Tatjana Loncar Turukalo, Piotr Przymus, Vladimir Trajkovik, Oliver Aasmets, Magali Berland, et al. 2021. "Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment." *Frontiers in Microbiology* 12 (February): 634511.
Marcotte, H., and M. C. Lavoie. 1998. "Oral Microbial Ecology and the Role of Salivary Immunoglobulin A." *Microbiology and Molecular Biology Reviews: MMBR* 62 (1): 71–109.
Mark Welch, Jessica L., Blair J. Rossetti, Christopher W. Rieken, Floyd E. Dewhirst, and Gary G. Borisy. 2016. "Biogeography of a Human Oral Microbiome at the Micron Scale." *Proceedings of the National Academy of Sciences of the United States of*

America 113 (6): E791–800.
Marotz, Clarisse A., Jon G. Sanders, Cristal Zuniga, Livia S. Zaramela, Rob Knight, and Karsten Zengler. 2018. "Improving Saliva Shotgun Metagenomics by Chemical Host DNA Depletion." *Microbiome* 6 (1): 42.
Marshall, B. C., A. Elbert, K. Petren, S. Rizvi, A. Fink, J. Ostrenga, and Sewall A. 2015. "Cystic Fibrosis Foundation Patient Registry 2014 Annual Data Report." Bethesda, Maryland: Cystic Fibrosis Foundation.
Marsh, P. D. 1994. "Microbial Ecology of Dental Plaque and Its Significance in Health and Disease." *Advances in Dental Research* 8 (2): 263–71.
Marsh, Philip D., Thuy Do, David Beighton, and Deirdre A. Devine. 2016. "Influence of Saliva on the Oral Microbiota." *Periodontology 2000* 70 (1): 80–92.
Martens, Miet, Peter Dawyndt, Renata Coopman, Monique Gillis, Paul De Vos, and Anne Willems. 2008. "Advantages of Multilocus Sequence Analysis for Taxonomic Studies: A Case Study Using 10 Housekeeping Genes in the Genus Ensifer (including Former Sinorhizobium)." *International Journal of Systematic and Evolutionary Microbiology* 58 (Pt 1): 200–214.
Martínez-Cuesta, M. Carmen, Carmen Peláez, and Teresa Requena. 2019. "Chapter 6 - Laboratory Simulators of the Colon Microbiome." In *Microbiome and Metabolome in Diagnosis, Therapy, and Other Strategic Applications*, edited by Joel Faintuch and Salomao Faintuch, 61–67. Elsevier Science.
Martinez-Martinez, R. E., J. P. Loyola-Rodriguez, S. E. Bonilla-Garro, N. Patiño-Marin, D. Haubek, A. Amano, and K. Poulsen. 2013. "Characterization of Periodontal Biofilm in Down Syndrome Patients: A Comparative Study." *The Journal of Clinical Pediatric Dentistry* 37 (3): 289–95.
Martin, Joseph. 2019. "Reproducibility: The Search for Microbiome Standards." *BioTechniques* 67 (3): 86–88.
Maruvada, Padma, Vanessa Leone, Lee M. Kaplan, and Eugene B. Chang. 2017. "The Human Microbiome and Obesity: Moving beyond Associations." *Cell Host & Microbe* 22 (5): 589–99.
Mashima, Izumi, M. Fujita, Y. Nakatsuka, T. Kado, Y. Furuichi, Sulistyani Herastuti, and F. Nakazawa. 2015. "The Distribution and Frequency of Oral Veillonella Spp. Associated with Chronic Periodontitis." https://www.semanticscholar.org/paper/

e8d55ce20a438733917336d7671f9228506628ae.
Mason, Matthew R., Philip M. Preshaw, Haikady N. Nagaraja, Shareef M. Dabdoub, Anis Rahman, and Purnima S. Kumar. 2015. "The Subgingival Microbiome of Clinically Healthy Current and Never Smokers." *The ISME Journal* 9 (1): 268–72.
Matarazzo, Flavia, Adriana Costa Ribeiro, Magda Feres, Marcelo Faveri, and Marcia Pinto Alves Mayer. 2011. "Diversity and Quantitative Analysis of Archaea in Aggressive Periodontitis and Periodontally Healthy Subjects." *Journal of Clinical Periodontology* 38 (7): 621–27.
McCoy, Amber N., Félix Araújo-Pérez, Andrea Azcárate-Peril, Jen Jen Yeh, Robert S. Sandler, and Temitope O. Keku. 2013. "Fusobacterium Is Associated with Colorectal Adenomas." *PloS One* 8 (1): e53653.
McDonough, Liza K., Karina T. Meredith, Chandima Nikagolla, Ryan J. Middleton, Jian K. Tan, Asanga V. Ranasinghe, Frederic Sierro, and Richard B. Banati. 2020. "The Water Chemistry and Microbiome of Household Wells in Medawachchiya, Sri Lanka, an Area with High Prevalence of Chronic Kidney Disease of Unknown Origin (CKDu)." *Scientific Reports* 10 (1): 18295.
McKone, Edward F., Jing Shao, Daisy D. Frangolias, Cassie L. Keener, Cynthia A. Shephard, Federico M. Farin, Mark R. Tonelli, et al. 2006. "Variants in the Glutamate-Cysteine-Ligase Gene Are Associated with Cystic Fibrosis Lung Disease." *American Journal of Respiratory and Critical Care Medicine* 174 (4): 415–19.
McLean, Jeffrey S. 2014. "Advancements toward a Systems Level Understanding of the Human Oral Microbiome." *Frontiers in Cellular and Infection Microbiology* 4 (July): 98.
McManus, Brenda A., Rory Maguire, Phillipa J. Cashin, Noel Claffey, Stephen Flint, Mohammed H. Abdulrahim, and David C. Coleman. 2012. "Enrichment of Multilocus Sequence Typing Clade 1 with Oral Candida Albicans Isolates in Patients with Untreated Periodontitis." *Journal of Clinical Microbiology* 50 (10): 3335–44.
McMurdie, Paul J., and Susan Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PloS One* 8 (4): e61217.
———. 2014. "Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible." *PLoS Computational Biology* 10 (4):

e1003531.
Mealey, Brian L., Thomas W. Oates, and American Academy of Periodontology. 2006. “Diabetes Mellitus and Periodontal Diseases.” *Journal of Periodontology* 77 (8): 1289–1303.
Menuet, Magalie, Fadi Bittar, Nathalie Stremler, Jean-Christophe Dubus, Jacques Sarles, Didier Raoult, and Jean-Marc Rolain. 2008. “First Isolation of Two Colistin-Resistant Emerging Pathogens, Brevundimonas Diminuta and Ochrobactrum Anthropi, in a Woman with Cystic Fibrosis: A Case Report.” *Journal of Medical Case Reports* 2 (December): 373.
Meskin, L. H., E. M. Farsht, and D. L. Anderson. 1968. “Prevalence of Bacteroides Melaninogenicus in the Gingival Crevice Area of Institutionalized Trisomy 21 and Cerebral Palsy Patients and Normal Children.” *Journal of Periodontology* 39 (6): 326–28.
Metchnikoff, Elie. 1907. *The Prolongation of Life; Optimistic Studies,*. London: Heinemann.
Meulen, Taco A. van der, Hermie J. M. Harmsen, Hendrika Bootsma, Silvia C. Liefers, Arnau Vich Vila, Alexandra Zhernakova, Jingyuan Fu, et al. 2018. “Dysbiosis of the Buccal Mucosa Microbiome in Primary Sjögren’s Syndrome Patients.” *Rheumatology* 57 (12): 2225–34.
Meurman, Jukka H. 2010. “Oral Microbiota and Cancer.” *Journal of Oral Microbiology* 2 (August). https://doi.org/10.3402/jom.v2i0.5195.
Meynell, E. 1995. “French Reactions to Jenner’s Discovery of Smallpox Vaccination: The Primary Sources.” *Social History of Medicine: The Journal of the Society for the Social History of Medicine / SSHM* 8 (2): 285–303.
Michaud, Dominique S., Zhuxuan Fu, Jian Shi, and Mei Chung. 2017. “Periodontal Disease, Tooth Loss, and Cancer Risk.” *Epidemiologic Reviews* 39 (1): 49–58.
Michaud, Dominique S., Jacques Izard, Charlotte S. Wilhelm-Benartzi, Doo-Ho You, Verena A. Grote, Anne Tjønneland, Christina C. Dahm, et al. 2013. “Plasma Antibodies to Oral Bacteria and Risk of Pancreatic Cancer in a Large European Prospective Cohort Study.” *Gut* 62 (12): 1764–70.
Miklossy, Judith. 2016. “Bacterial Amyloid and DNA Are Important Constituents of Senile Plaques: Further Evidence of the Spirochetal and Biofilm Nature of Senile Plaques.” *Journal of Alzheimer’s Disease: JAD* 53 (4): 1459–73.
Mohanty, Rinkee, Swati Joshi Asopa, M. Derick Joseph,

Bhupender Singh, Jagadish Prasad Rajguru, K. Saidath, and Uma Sharma. 2019. “Red Complex: Polymicrobial Conglomerate in Oral Flora: A Review.” *Journal of Family Medicine and Primary Care* 8 (11): 3480–86.

Mohapatra, N. K., P. W. Cheng, J. C. Parker, A. M. Paradiso, J. R. Yankaskas, R. C. Boucher, and T. F. Boat. 1995. “Alteration of Sulfation of Glycoconjugates, but Not Sulfate Transport and Intracellular Inorganic Sulfate Content in Cystic Fibrosis Airway Epithelial Cells.” *Pediatric Research* 38 (1): 42–48.

Mombelli, A., F. A. Gusberti, M. A. van Oosten, and N. P. Lang. 1989. “Gingival Health and Gingivitis Development during Puberty. A 4-Year Longitudinal Study.” *Journal of Clinical Periodontology* 16 (7): 451–56.

Moon, J-H, J-H Lee, and J-Y Lee. 2015. “Subgingival Microbiome in Smokers and Non-Smokers in Korean Chronic Periodontitis Patients.” *Molecular Oral Microbiology* 30 (3): 227–41.

Moon, Ji-Hoi, and Jae-Hyung Lee. 2016. “Probing the Diversity of Healthy Oral Microbiome with Bioinformatics Approaches.” *BMB Reports* 49 (12): 662–70.

Moran, Mary Ann. 2009. “Metatranscriptomics: Eavesdropping on Complex Microbial Communities.” *Microbe* 4 (7): 329–35.

Moreira, Maurício José Santos, Carolina Schwertner, Juliana Jobim Jardim, and Lina Naomi Hashizume. 2016. “Dental Caries in Individuals with Down Syndrome: A Systematic Review.” *International Journal of Paediatric Dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children* 26 (1): 3–12.

Morovic, Wesley, and Charles R. Budinoff. 2021. “Epigenetics: A New Frontier in Probiotic Research.” *Trends in Microbiology* 29 (2): 117–26.

Morris, E. Kathryn, Tancredi Caruso, François Buscot, Markus Fischer, Christine Hancock, Tanja S. Maier, Torsten Meiners, et al. 2014. “Choosing and Using Diversity Indices: Insights for Ecological Applications from the German Biodiversity Exploratories.” *Ecology and Evolution* 4 (18): 3514–24.

Morton, James T., Clarisse Marotz, Alex Washburne, Justin Silverman, Livia S. Zaramela, Anna Edlund, Karsten Zengler, and Rob Knight. 2019. “Establishing Microbial Composition Measurement Standards with Reference Frames.” *Nature Communications* 10 (1): 2719.

Moutsopoulos, Niki M., and Joanne E. Konkel. 2018. “Tissue-

Specific Immunity at the Oral Mucosal Barrier." *Trends in Immunology* 39 (4): 276–87.

Muras, Andrea, Celia Mayer, Paz Otero-Casal, Rob A. M. Exterkate, Bernd W. Brandt, Wim Crielaard, Ana Otero, and Bastiaan P. Krom. 2020. "Short-Chain N-Acylhomoserine Lactone Quorum-Sensing Molecules Promote Periodontal Pathogens in In Vitro Oral Biofilms." *Applied and Environmental Microbiology* 86 (3). https://doi.org/10.1128/AEM.01941-19.

Muras, Andrea, Paz Otero-Casal, Vanessa Blanc, and Ana Otero. 2020. "Acyl Homoserine Lactone-Mediated Quorum Sensing in the Oral Cavity: A Paradigm Revisited." *Scientific Reports* 10 (1): 9800.

Murovec, Boštjan, Leon Deutsch, and Blaž Stres. 2021. "General Unified Microbiome Profiling Pipeline (GUMPP) for Large Scale, Streamlined and Reproducible Analysis of Bacterial 16S rRNA Data to Predicted Microbial Metagenomes, Enzymatic Reactions and Metabolic Pathways." *Metabolites* 11 (6). https://doi.org/10.3390/metabo11060336.

Nadig, Suchetha Devendrappa, Deepak Timmasandra Ashwathappa, Muniraju Manjunath, Sowmya Krishna, Araleri Gopalkrishna Annaji, and Praveen Kunigal Shivaprakash. 2017. "A Relationship between Salivary Flow Rates and Candida Counts in Patients with Xerostomia." *Journal of Oral and Maxillofacial Pathology: JOMFP* 21 (2): 316.

Naginyte, Monika, Thuy Do, Josephine Meade, Deirdre Ann Devine, and Philip David Marsh. 2019. "Enrichment of Periodontal Pathogens from the Biofilms of Healthy Adults." *Scientific Reports* 9 (1): 5491.

Naidu, B. V., and B. A. Reginald. 2016. "Quantification and Correlation of Oral Candida with Caries Index Among Different Age Groups of School Children: A Case-Control Study." *Annals of Medical and Health Sciences Research* 6 (2): 80–84.

Nallabelli, Nayudu, Prashant P. Patil, Vijay Kumar Pal, Namrata Singh, Ashish Jain, Prabhu B. Patil, Vishakha Grover, and Suresh Korpole. 2016. "Biochemical and Genome Sequence Analyses of Megasphaera Sp. Strain DISK18 from Dental Plaque of a Healthy Individual Reveals Commensal Lifestyle." *Scientific Reports* 6 (1): 33665.

Nassar, Mohannad, Noriko Hiraishi, Md Sofiqul Islam, Masayuki Otsuki, and Junji Tagami. 2014. "Age-Related Changes in

Salivary Biomarkers." *Journal of Dental Sciences* 9 (1): 85–90.
National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Board on Science Education, and Committee on Designing Citizen Science to Support Science Learning. 2019. *Learning Through Citizen Science: Enhancing Opportunities by Design*. Edited by Kenne Ann Dibner and Rajul Pandya. Washington (DC): National Academies Press (US).
Naumova, Ella A., Tobias Dierkes, Jürgen Sprang, and Wolfgang H. Arnold. 2013. "The Oral Mucosal Surface and Blood Vessels." *Head & Face Medicine* 9 (March): 8.
Nawaz, Ahmed, Fariha Hasan, and Aamer A. Shah. 2015. "Degradation of Poly(ε-Caprolactone) (PCL) by a Newly Isolated Brevundimonas Sp. Strain MRL-AN1 from Soil." *FEMS Microbiology Letters* 362 (1): 1–7.
Nearing, Jacob T., Gavin M. Douglas, André M. Comeau, and Morgan G. I. Langille. 2018. "Denoising the Denoisers: An Independent Evaluation of Microbiome Sequence Error-Correction Approaches." *PeerJ* 6 (August): e5364.
Nibali, Luigi, Vanessa Sousa, Mehmet Davrandi, David Spratt, Qumasha Alyahya, Jose Dopico, and Nikos Donos. 2020. "Differences in the Periodontal Microbiome of Successfully Treated and Persistent Aggressive Periodontitis." *Journal of Clinical Periodontology* 47 (8): 980–90.
Nielsen, Shaun, Bronwen Needham, Steven T. Leach, Andrew S. Day, Adam Jaffe, Torsten Thomas, and Chee Y. Ooi. 2016. "Disrupted Progression of the Intestinal Microbiota with Age in Children with Cystic Fibrosis." *Scientific Reports* 6 (1): 24857.
NIH HMP Working Group, Jane Peterson, Susan Garges, Maria Giovanni, Pamela McInnes, Lu Wang, Jeffery A. Schloss, et al. 2009. "The NIH Human Microbiome Project." *Genome Research* 19 (12): 2317–23.
Ninomiya, Masami, Mari Hashimoto, Kouji Yamanouchi, Yoshiaki Fukumura, Toshihiko Nagata, and Koji Naruishi. 2020. "Relationship of Oral Conditions to the Incidence of Infective Endocarditis in Periodontitis Patients with Valvular Heart Disease: A Cross-Sectional Study." *Clinical Oral Investigations* 24 (2): 833–40.
Nørskov-Lauritsen, Niels, and Mogens Kilian. 2006. "Reclassification of Actinobacillus Actinomycetemcomitans, Haemophilus Aphrophilus, Haemophilus Paraphrophilus and

Haemophilus Segnis as Aggregatibacter Actinomycetemcomitans Gen. Nov., Comb. Nov., Aggregatibacter Aphrophilus Comb. Nov. and Aggregatibacter Segnis Comb. Nov., and Emended Description of Aggregatibacter Aphrophilus to Include V Factor-Dependent and V Factor-Independent Isolates." *International Journal of Systematic and Evolutionary Microbiology* 56 (Pt 9): 2135–46.

Odum, Eugene P. 1971. "Fundamentals of Ecology." Philadelphia: Saunders. 1971. http://catalog.hathitrust.org/api/volumes/oclc/154846.html.

Oelschlaeger, Tobias A. 2010. "Mechanisms of Probiotic Actions - A Review." *International Journal of Medical Microbiology: IJMM* 300 (1): 57–62.

Ogier, Jean-Claude, Sylvie Pagès, Maxime Galan, Matthieu Barret, and Sophie Gaudriault. 2019. "rpoB, a Promising Marker for Analyzing the Diversity of Bacterial Communities by Amplicon Sequencing." *BMC Microbiology* 19 (1): 171.

Oh, Julia, Allyson L. Byrd, Morgan Park, NISC Comparative Sequencing Program, Heidi H. Kong, and Julia A. Segre. 2016. "Temporal Stability of the Human Skin Microbiome." *Cell* 165 (4): 854–66.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, et al. 2017. *Vegan: Community Ecology Package* (version 2.4.6). https://CRAN.R-project.org/package=vegan.

Oswal, Pratima, Sandeep Katti, Vinayak Joshi, and Hawaabi Shaikh. 2020. "Identification of *Dialister Pneumosintes* in Healthy and Chronic Periodontitis Patients with Type 2 Diabetes Mellitus and Its Correlation with the Red Complex Bacteria." *Journal of Interdisciplinary Dentistry* 10 (1): 17.

Paetzold, Bernhard, Jesse R. Willis, João Pereira de Lima, Nastassia Knödlseder, Holger Brüggemann, Sven R. Quist, Toni Gabaldón, and Marc Güell. 2019. "Skin Microbiome Modulation Induced by Probiotic Solutions." *Microbiome* 7 (1): 95.

Paganin, Patrizia, Ersilia Vita Fiscarelli, Vanessa Tuccio, Manuela Chiancianesi, Giovanni Bacci, Patrizia Morelli, Daniela Dolce, et al. 2015. "Changes in Cystic Fibrosis Airway Microbial Community Associated with a Severe Decline in Lung Function." *PloS One* 10 (4): e0124348.

Palarea-Albaladejo, Javier, and Josep Antoni Martín-Fernández.

2015. “zCompositions — R Package for Multivariate Imputation of Left-Censored Data under a Compositional Approach.” *Chemometrics and Intelligent Laboratory Systems* 143 (April): 85–96.

Pasolli, Edoardo, Francesco Asnicar, Serena Manara, Moreno Zolfo, Nicolai Karcher, Federica Armanini, Francesco Beghini, et al. 2019. “Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle.” *Cell* 176 (3): 649–62.e20.

Pasolli, Edoardo, Duy Tin Truong, Faizan Malik, Levi Waldron, and Nicola Segata. 2016. “Machine Learning Meta-Analysis of Large Metagenomic Datasets: Tools and Biological Insights.” *PLoS Computational Biology* 12 (7): e1004977.

Paster, Bruce J., Ingar Olsen, Jørn A. Aas, and Floyd E. Dewhirst. 2006. “The Breadth of Bacterial Diversity in the Human Periodontal Pocket and Other Oral Sites.” *Periodontology 2000* 42: 80–87.

Patil, Priti Basgauda, and Basgauda Ramesh Patil. 2011. “Saliva: A Diagnostic Biomarker of Periodontal Diseases.” *Journal of Indian Society of Periodontology* 15 (4): 310–17.

Pawlaczyk-Kamieńska, Tamara, Maria Borysewicz-Lewicka, and Halina Batura-Gabryel. 2019. “Salivary Biomarkers and Oral Microbial Load in Relation to the Dental Status of Adults with Cystic Fibrosis.” *Microorganisms* 7 (12). https://doi.org/10.3390/microorganisms7120692.

Pawlaczyk-Kamieńska, Tamara, Maria Borysewicz-Lewicka, Renata Śniatała, Halina Batura-Gabryel, and Szczepan Cofta. 2019. “Dental and Periodontal Manifestations in Patients with Cystic Fibrosis - A Systematic Review.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 18 (6): 762–71.

Peano, Clelia, Alessandro Pietrelli, Clarissa Consolandi, Elio Rossi, Luca Petiti, Letizia Tagliabue, Gianluca De Bellis, and Paolo Landini. 2013. “An Efficient rRNA Removal Method for RNA Sequencing in GC-Rich Bacteria.” *Microbial Informatics and Experimentation* 3 (1): 1.

Percival, R. S., S. J. Challacombe, and P. D. Marsh. 1994. “Flow Rates of Resting Whole and Stimulated Parotid Saliva in Relation to Age and Gender.” *Journal of Dental Research* 73 (8): 1416–20.

Pérez-Brocal, Vicente, and Andrés Moya. 2018. “The Analysis of the Oral DNA Virome Reveals Which Viruses Are Widespread and Rare among Healthy Young Adults in Valencia (Spain).” *PloS One* 13 (2): e0191867.

Pérez-Chaparro, P. J., C. Gonçalves, L. C. Figueiredo, M. Faveri, E. Lobão, N. Tamashiro, P. Duarte, and M. Feres. 2014. “Newly Identified Pathogens Associated with Periodontitis: A Systematic Review.” *Journal of Dental Research* 93 (9): 846–58.

Peters, Brandilyn A., Jing Wu, Richard B. Hayes, and Jiyoung Ahn. 2017. “The Oral Fungal Mycobiome: Characteristics and Relation to Periodontitis in a Pilot Study.” *BMC Microbiology* 17 (1): 157.

Peters, Brandilyn A., Jing Wu, Zhiheng Pei, Liying Yang, Mark P. Purdue, Neal D. Freedman, Eric J. Jacobs, Susan M. Gapstur, Richard B. Hayes, and Jiyoung Ahn. 2017. “Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers.” *Cancer Research* 77 (23): 6777–87.

Peterson, Scott N., Erik Snesrud, Jia Liu, Ana C. Ong, Mogens Kilian, Nicholas J. Schork, and Walter Bretz. 2013. “The Dental Plaque Microbiome in Health and Disease.” *PloS One* 8 (3): e58487.

Pilcher, Elizabeth. 1998. “Dental Care for the Patient with Down Syndrome.” *Down’s Syndrome, Research and Practice: The Journal of the Sarah Duffen Centre / University of Portsmouth* 5 (3): 111–16.

Pincelli, João Vitor, Luiz Otávio Vittorelli, Nickolas Stabellini, and João Renato Rebello Pinho. 2020. “Chapter 2 - The Role of the Microbiome in Precision Medicine.” In *Precision Medicine for Investigators, Practitioners and Providers*, edited by Joel Faintuch and Salomao Faintuch, 13–18. Academic Press.

Pivot, Diane, Annlyse Fanton, Edgar Badell-Ocando, Marion Benouachkou, Karine Astruc, Frederic Huet, Lucie Amoureux, et al. 2019. “Carriage of a Single Strain of Nontoxigenic Corynebacterium Diphtheriae Bv. Belfanti (Corynebacterium Belfantii) in Four Patients with Cystic Fibrosis.” *Journal of Clinical Microbiology* 57 (5). https://doi.org/10.1128/JCM.00042-19.

Popp, Denny, and Florian Centler. 2020. “μBialSim: Constraint-Based Dynamic Simulation of Complex Microbiomes.” *Frontiers in Bioengineering and Biotechnology* 8 (June): 574.

Pramanik, Rashida, Hayley Thompson, James O. Kistler, William G. Wade, John Galloway, Tim Peakman, and Gordon B. Proctor. 2012. "Effects of the UK Biobank Collection Protocol on Potential Biomarkers in Saliva." *International Journal of Epidemiology* 41 (6): 1786–97.

Prasad, Mayuri, Toshi, Rajat Sehgal, Manisha Mallik, Aaysha Tabinda Nabi, and Sneha Singh. 2019. "Periodontal Disease and Salivary pH: Case Control Study." *Prasad Mayuri, Toshi, Sehgal Rajat, Mallik Manisha, Nabi Aaysha Tabinda, Singh Sneha. Periodontal Disease and Salivary pH: Case Control Study. International Archives of Integrated Medicine* 6 (2): 1–6.

Prodan, Andrei, Valentina Tremaroli, Harald Brolin, Aeilko H. Zwinderman, Max Nieuwdorp, and Evgeni Levin. 2020. "Comparing Bioinformatic Pipelines for Microbial 16S rRNA Amplicon Sequencing." *PloS One* 15 (1): e0227434.

Pruesse, Elmar, Christian Quast, Katrin Knittel, Bernhard M. Fuchs, Wolfgang Ludwig, Jörg Peplies, and Frank Oliver Glöckner. 2007. "SILVA: A Comprehensive Online Resource for Quality Checked and Aligned Ribosomal RNA Sequence Data Compatible with ARB." *Nucleic Acids Research* 35 (21): 7188–96.

Pushalkar, Smruti, Xiaojie Ji, Yihong Li, Cherry Estilo, Ramanathan Yegnanarayana, Bhuvanesh Singh, Xin Li, and Deepak Saxena. 2012. "Comparison of Oral Microbiota in Tumor and Non-Tumor Tissues of Patients with Oral Squamous Cell Carcinoma." *BMC Microbiology* 12 (July): 144.

Qu, X. M., Z. F. Wu, B. X. Pang, L. Y. Jin, L. Z. Qin, and S. L. Wang. 2016. "From Nitrate to Nitric Oxide: The Role of Salivary Glands and Oral Bacteria." *Journal of Dental Research* 95 (13): 1452–56.

"Raising Standards in Microbiome Research." 2016. *Nature Microbiology* 1 (7): 16112.

Raja, Huzefa A., Andrew N. Miller, Cedric J. Pearce, and Nicholas H. Oberlies. 2017. "Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community." *Journal of Natural Products* 80 (3): 756–70.

Rajani, Ronak, and John L. Klein. 2020. "Infective Endocarditis: A Contemporary Update." *Clinical Medicine* 20 (1): 31–35.

Ramanan, Poornima, Jason N. Barreto, Douglas R. Osmon, and Pritish K. Tosh. 2014. "Rothia Bacteremia: A 10-Year Experience at Mayo Clinic, Rochester, Minnesota." *Journal of*

Clinical Microbiology 52 (9): 3184–89.
Rammos, Christos, Ulrike B. Hendgen-Cotta, Matthias Totzeck, Julia Pohl, Peter Lüdike, Ulrich Flögel, René Deenen, et al. 2016. “Impact of Dietary Nitrate on Age-Related Diastolic Dysfunction.” *European Journal of Heart Failure* 18 (6): 599–610.
Rampelli, Simone, Matteo Soverini, Federica D’Amico, Monica Barone, Teresa Tavella, Daniela Monti, Miriam Capri, et al. 2020. “Shotgun Metagenomics of Gut Microbiota in Humans with up to Extreme Longevity and the Increasing Role of Xenobiotic Degradation.” *mSystems* 5 (2). https://doi.org/10.1128/mSystems.00124-20.
Rangé, Hélène, Julien Labreuche, Liliane Louedec, Philippe Rondeau, Cynthia Planesse, Uriel Sebbag, Emmanuel Bourdon, Jean-Baptiste Michel, Philippe Bouchard, and Olivier Meilhac. 2014. “Periodontal Bacteria in Human Carotid Atherothrombosis as a Potential Trigger for Neutrophil Activation.” *Atherosclerosis* 236 (2): 448–55.
Ranjan, Ravi, Asha Rani, Ahmed Metwally, Halvor S. McGee, and David L. Perkins. 2016. “Analysis of the Microbiome: Advantages of Whole Genome Shotgun versus 16S Amplicon Sequencing.” *Biochemical and Biophysical Research Communications* 469 (4): 967–77.
Razvi, Samiya, Lynne Quittell, Ase Sewall, Hebe Quinton, Bruce Marshall, and Lisa Saiman. 2009. “Respiratory Microbiology of Patients with Cystic Fibrosis in the United States, 1995 to 2005.” *Chest* 136 (6): 1554–60.
R Core Team. 2020. “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing. https://www.R-project.org/.
Reuland-Bosma, W., and J. van Dijk. 1986. “Periodontal Disease in Down’s Syndrome: A Review.” *Journal of Clinical Periodontology* 13 (1): 64–73.
Ricaurte, J. C., O. Klein, V. LaBombardi, V. Martinez, A. Serpe, and M. Joy. 2001. “Rothia Dentocariosa Endocarditis Complicated by Multiple Intracranial Hemorrhages.” *Southern Medical Journal* 94 (4): 438–40.
Richards, Vincent P., Andres J. Alvarez, Amy R. Luce, Molly Bedenbaugh, Mary Lyn Mitchell, Robert A. Burne, and Marcelle M. Nascimento. 2017. “Microbiomes of Site-Specific Dental Plaques from Children with Different Caries Status.”

Infection and Immunity 85 (8). https://doi.org/10.1128/IAI.00106-17.
Riggio, M. P., and A. Lennon. 2002. “Development of a PCR Assay Specific for Peptostreptococcus Anaerobius.” *Journal of Medical Microbiology* 51 (12): 1097–1101.
Rinke, Christian. 2018. “Single-Cell Genomics of Microbial Dark Matter.” *Methods in Molecular Biology* 1849: 99–111.
Rinke, Christian, Patrick Schwientek, Alexander Sczyrba, Natalia N. Ivanova, Iain J. Anderson, Jan-Fang Cheng, Aaron Darling, et al. 2013. “Insights into the Phylogeny and Coding Potential of Microbial Dark Matter.” *Nature* 499 (7459): 431–37.
Rivas Caldas, Rocio, Florence Le Gall, Krista Revert, Gilles Rault, Michèle Virmaux, Stephanie Gouriou, Geneviève Héry-Arnaud, Georges Barbier, and Sylvie Boisramé. 2015. “Pseudomonas Aeruginosa and Periodontal Pathogens in the Oral Cavity and Lungs of Cystic Fibrosis Patients: A Case-Control Study.” *Journal of Clinical Microbiology* 53 (6): 1898–1907.
Robinson, Camilla V., Mark R. Elkins, Katarzyna M. Bialkowski, David J. Thornton, and Michael A. Kertesz. 2012. “Desulfurization of Mucin by Pseudomonas Aeruginosa: Influence of Sulfate in the Lungs of Cystic Fibrosis Patients.” *Journal of Medical Microbiology* 61 (Pt 12): 1644–53.
Rong, Ruichen, Shuang Jiang, Lin Xu, Guanghua Xiao, Yang Xie, Dajiang J. Liu, Qiwei Li, and Xiaowei Zhan. 2021. “MB-GAN: Microbiome Simulation via Generative Adversarial Network.” *GigaScience* 10 (2). https://doi.org/10.1093/gigascience/giab005.
Rosenfeld, Margaret, Julia Emerson, Sharon McNamara, Valeria Thompson, Bonnie W. Ramsey, Wayne Morgan, Ronald L. Gibson, and EPIC Study Group. 2012. “Risk Factors for Age at Initial Pseudomonas Acquisition in the Cystic Fibrosis Epic Observational Cohort.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 11 (5): 446–53.
Rosier, Bob T., Eva M. Moya-Gonzalvez, Paula Corell-Escuin, and Alex Mira. 2020. “Isolation and Characterization of Nitrate-Reducing Bacteria as Potential Probiotics for Oral and Systemic Health.” *Frontiers in Microbiology* 11 (September): 555465.
Roszyk, Ewa, and Mariusz Puszczewicz. 2017. “Role of Human Microbiome and Selected Bacterial Infections in the

Pathogenesis of Rheumatoid Arthritis.” *Reumatologia* 55 (5): 242–50.

Rowan-Nash, Aislinn D., Benjamin J. Korry, Eleftherios Mylonakis, and Peter Belenky. 2019. “Cross-Domain and Viral Interactions in the Microbiome.” *Microbiology and Molecular Biology Reviews: MMBR* 83 (1). https://doi.org/10.1128/MMBR.00044-18.

Ruhl, Stefan, Andreas Eidt, Holger Melzl, Udo Reischl, and John O. Cisar. 2014. “Probing of Microbial Biofilm Communities for Coadhesion Partners.” *Applied and Environmental Microbiology* 80 (21): 6583–90.

“Saca La Lengua - Stick out Your Tongue.” n.d. Accessed July 27, 2021. https://www.sacalalengua.org/stick-out-your-tongue/.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. “Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia.” *Science* 230 (4732): 1350–54.

Saiman, Lisa, Jane Siegel, and Cystic Fibrosis Foundation Consensus Conference on Infection Control Participants. 2003. “Infection Control Recommendations for Patients with Cystic Fibrosis: Microbiology, Important Pathogens, and Infection Control Practices to Prevent Patient-to-Patient Transmission.” *American Journal of Infection Control* 31 (3 Suppl): S1–62.

Saint-Criq, Vinciane, and Michael A. Gray. 2017. “Role of CFTR in Epithelial Physiology.” *Cellular and Molecular Life Sciences: CMLS* 74 (1): 93–115.

Sakellari, D., K. N. Arapostathis, and A. Konstantinidis. 2005. “Periodontal Conditions and Subgingival Microflora in Down Syndrome Patients. A Case-Control Study.” *Journal of Clinical Periodontology* 32 (6): 684–90.

Salter, Susannah J., Michael J. Cox, Elena M. Turek, Szymon T. Calus, William O. Cookson, Miriam F. Moffatt, Paul Turner, Julian Parkhill, Nicholas J. Loman, and Alan W. Walker. 2014. “Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses.” *BMC Biology* 12 (November): 87.

Sánchez-Bautista, Antonia, Juan Carlos Rodríguez-Díaz, Inmaculada Garcia-Heredia, Carmen Luna-Paredes, and Pedro J. Alcalá-Minagorre. 2019. “Airway Microbiota in

Patients with Paediatric Cystic Fibrosis: Relationship with Clinical Status." *Enfermedades Infecciosas Y Microbiologia Clinica* 37 (3): 167–71.

Sánchez-Quinto, Andrés, Daniel Cerqueda-García, Luisa I. Falcón, Osiris Gaona, Santiago Martínez-Correa, Javier Nieto, and Isaac G-Santoyo. 2020. "Gut Microbiome in Children from Indigenous and Urban Communities in México: Different Subsistence Models, Different Microbiomes." *Microorganisms* 8 (10). https://doi.org/10.3390/microorganisms8101592.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. "DNA Sequencing with Chain-Terminating Inhibitors." *Proceedings of the National Academy of Sciences of the United States of America* 74 (12): 5463–67.

Santos, Bruna Rafaela Martins dos, Clarissa Favero Demeda, Eutália Elizabeth Novaes Ferreira da Silva, Maria Helena Marques Fonsecae de Britto, Kenio Costa Lima, and Maria Celeste Nunes de Melo. 2014. "Prevalence of Subgingival Staphylococcus at Periodontally Healthy and Diseased Sites." *Brazilian Dental Journal* 25 (4): 271–76.

Sarkar, Anujit, Melanie N. Kuehl, Amy C. Alman, and Brant R. Burkhardt. 2021. "Linking the Oral Microbiome and Salivary Cytokine Abundance to Circadian Oscillations." *Scientific Reports* 11 (1): 2658.

Saunders, Charles W., Annika Scheynius, and Joseph Heitman. 2012. "Malassezia Fungi Are Specialized to Live on Skin and Associated with Dandruff, Eczema, and Other Skin Diseases." *PLoS Pathogens* 8 (6): e1002701.

Savage, D. C. 1977. "Microbial Ecology of the Gastrointestinal Tract." *Annual Review of Microbiology* 31: 107–33.

Saxelin, Maija. 2008. "Probiotic Formulations and Applications, the Current Probiotics Market, and Changes in the Marketplace: A European Perspective." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 46 Suppl 2 (February): S76–79; discussion S144–51.

Scannapieco, Frank A., Ananda P. Dasanayake, and Nok Chhun. 2010. "'Does Periodontal Therapy Reduce the Risk for Systemic Diseases?'" *Dental Clinics of North America* 54 (1): 163–81.

Scher, Jose U., Carles Ubeda, Michele Equinda, Raya Khanin, Yvonne Buischi, Agnes Viale, Lauren Lipuma, et al. 2012. "Periodontal Disease and the Oral Microbiota in New-Onset

Rheumatoid Arthritis.” *Arthritis and Rheumatism* 64 (10): 3083–94.
Schliep, Klaus Peter. 2011. “Phangorn: Phylogenetic Analysis in R.” *Bioinformatics* 27 (4): 592–93.
Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. “Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities.” *Applied and Environmental Microbiology* 75 (23): 7537–41.
Schoch, Conrad L., Keith A. Seifert, Sabine Huhndorf, Vincent Robert, John L. Spouge, C. André Levesque, Wen Chen, Fungal Barcoding Consortium, and Fungal Barcoding Consortium Author List. 2012. “Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (16): 6241–46.
Schoilew, Kyrill, Helena Ueffing, Alexander Dalpke, Björn Wolff, Cornelia Frese, Diana Wolff, and Sébastien Boutin. 2019. “Bacterial Biofilm Composition in Healthy Subjects with and without Caries Experience.” *Journal of Oral Microbiology* 11 (1): 1633194.
Scoffield, Jessica A., and Hui Wu. 2015. “Oral Streptococci and Nitrite-Mediated Interference of Pseudomonas Aeruginosa.” *Infection and Immunity* 83 (1): 101–7.
Scott, Jennifer, Monica Sueiro-Olivares, Waqar Ahmed, Christoph Heddergott, Can Zhao, Riba Thomas, Michael Bromley, et al. 2019. “Pseudomonas Aeruginosa-Derived Volatile Sulfur Compounds Promote Distal Aspergillus Fumigatus Growth and a Synergistic Pathogen-Pathogen Interaction That Increases Pathogenicity in Co-Infection.” *Frontiers in Microbiology* 10 (October): 2311.
Segata, Nicola. 2015. “Gut Microbiome: Westernization and the Disappearance of Intestinal Diversity.” *Current Biology: CB*.
Segata, Nicola, Susan Kinder Haake, Peter Mannon, Katherine P. Lemon, Levi Waldron, Dirk Gevers, Curtis Huttenhower, and Jacques Izard. 2012. “Composition of the Adult Digestive Tract Bacterial Microbiome Based on Seven Mouth Surfaces, Tonsils, Throat and Stool Samples.” *Genome Biology* 13 (6): R42.
Sender, Ron, Shai Fuchs, and Ron Milo. 2016. “Revised Estimates

for the Number of Human and Bacteria Cells in the Body." *PLoS Biology* 14 (8): e1002533.

Shade, Ashley, and Jo Handelsman. 2012. "Beyond the Venn Diagram: The Hunt for a Core Microbiome." *Environmental Microbiology* 14 (1): 4–12.

Sharma, Poonam, Seydina M. Diene, Gregory Gimenez, Catherine Robert, and Jean-Marc Rolain. 2012. "Genome Sequence of Microbacterium Yannicii, a Bacterium Isolated from a Cystic Fibrosis Patient." *Journal of Bacteriology* 194 (17): 4785.

Sharma, Poonam, Seydina M. Diene, Sandrine Thibeaut, Fadi Bittar, Véronique Roux, Carine Gomez, Martine Reynaud-Gaubert, and Jean-Marc Rolain. 2013. "Phenotypic and Genotypic Properties of Microbacterium Yannicii, a Recently Described Multidrug Resistant Bacterium Isolated from a Lung Transplanted Patient with Cystic Fibrosis in France." *BMC Microbiology* 13 (May): 97.

Shaw, Liam, Andre L. R. Ribeiro, Adam P. Levine, Nikolas Pontikos, Francois Balloux, Anthony W. Segal, Adam P. Roberts, and Andrew M. Smith. 2017. "The Human Salivary Microbiome Is Shaped by Shared Environment Rather than Genetics: Evidence from a Large Family of Closely Related Individuals." *mBio* 8 (5). https://doi.org/10.1128/mBio.01237-17.

Shen, S., L. P. Samaranayake, and H-K Yip. 2005. "Coaggregation Profiles of the Microflora from Root Surface Caries Lesions." *Archives of Oral Biology* 50 (1): 23–32.

Sherman, Stephanie L., Emily G. Allen, Lora H. Bean, and Sallie B. Freeman. 2007. "Epidemiology of Down Syndrome." *Mental Retardation and Developmental Disabilities Research Reviews* 13 (3): 221–27.

Shillitoe, Edward, Ruth Weinstock, Taewan Kim, Howard Simon, Jessica Planer, Susan Noonan, and Robert Cooney. 2012. "The Oral Microflora in Obesity and Type-2 Diabetes." *Journal of Oral Microbiology* 4 (October). https://doi.org/10.3402/jom.v4i0.19013.

Shin, Mijeong, D. Han, H. Yi, and Haran Kim. 2013. "Characterization of Subgingival Microorganisms according to Periodontitis and Cardiovascular Status." In . unknown. http://dx.doi.org/.

Singhal, Neelja, Manish Kumar, Pawan K. Kanaujia, and Jugsharan S. Virdi. 2015. "MALDI-TOF Mass Spectrometry: An Emerging

Technology for Microbial Identification and Diagnosis.” *Frontiers in Microbiology* 6 (August): 791.
Sinha, Rashmi, Ni Zhao, James J. Goedert, Doratha A. Byrd, Yunhu Wan, Xing Hua, Autumn G. Hullings, et al. 2021. “Effects of Processed Meat and Drinking Water Nitrate on Oral and Fecal Microbial Populations in a Controlled Feeding Study.” *Environmental Research* 197 (June): 111084.
Siqueira, Walter Luiz Júnior, and José Nicolau. 2002. “Stimulated Whole Saliva Components in Children with Down Syndrome.” *Special Care in Dentistry: Official Publication of the American Association of Hospital Dentists, the Academy of Dentistry for the Handicapped, and the American Society for Geriatric Dentistry* 22 (6): 226–30.
Slots, J., C. Sugar, and J. J. Kamma. 2002. “Cytomegalovirus Periodontal Presence Is Associated with Subgingival Dialister Pneumosintes and Alveolar Bone Loss.” *Oral Microbiology and Immunology* 17 (6): 369–74.
Socransky, Sigmund S., and Anne D. Haffajee. 2002. “Dental Biofilms: Difficult Therapeutic Targets.” *Periodontology 2000* 28: 12–55.
Socransky, S. S., A. D. Haffajee, M. A. Cugini, C. Smith, and R. L. Kent Jr. 1998. “Microbial Complexes in Subgingival Plaque.” *Journal of Clinical Periodontology* 25 (2): 134–44.
Söder, B., L. J. Jin, B. Klinge, and P-O Söder. 2007. “Periodontitis and Premature Death: A 16-Year Longitudinal Study in a Swedish Urban Population.” *Journal of Periodontal Research* 42 (4): 361–66.
Song, Se Jin, Christian Lauber, Elizabeth K. Costello, Catherine A. Lozupone, Gregory Humphrey, Donna Berg-Lyons, J. Gregory Caporaso, et al. 2013. “Cohabiting Family Members Share Microbiota with One Another and with Their Dogs.” *eLife* 2 (April): e00458.
Soret, Perrine, Louise-Eva Vandenborght, Florence Francis, Noémie Coron, Raphael Enaud, Marta Avalos, Thierry Schaeverbeke, et al. 2020. “Respiratory Mycobiome and Suggestion of Inter-Kingdom Network during Acute Pulmonary Exacerbation in Cystic Fibrosis.” *Scientific Reports* 10 (1): 3589.
Stahringer, Simone S., Jose C. Clemente, Robin P. Corley, John Hewitt, Dan Knights, William A. Walters, Rob Knight, and Kenneth S. Krauter. 2012. “Nurture Trumps Nature in a

Longitudinal Survey of Salivary Bacterial Communities in Twins from Early Adolescence to Early Adulthood." *Genome Research* 22 (11): 2146–52.
Stephen, Abish S., Declan P. Naughton, Robert L. Pizzey, David J. Bradshaw, and Gary R. Burnett. 2014. "In Vitro Growth Characteristics and Volatile Sulfur Compound Production of Solobacterium Moorei." *Anaerobe* 26 (April): 53–57.
Stipanuk, Martha H., John E. Dominy Jr, Jeong-In Lee, and Relicardo M. Coloso. 2006. "Mammalian Cysteine Metabolism: New Insights into Regulation of Cysteine Metabolism." *The Journal of Nutrition* 136 (6 Suppl): 1652S – 1659S.
Sun, Fubo, Ayaz Ahmed, Lina Wang, Ming Dong, and Weidong Niu. 2018. "Comparison of Oral Microbiota in Orthodontic Patients and Healthy Individuals." *Microbial Pathogenesis* 123 (October): 473–77.
Tabatabaei, Mohammad, Mahdi Dastbarsar, and Mohammad Ashkan Moslehi. 2019. "Isolation and Identification of Pandoraea Spp. From Bronchoalveolar Lavage of Cystic Fibrosis Patients in Iran." *Italian Journal of Pediatrics* 45 (1): 118.
Takahashi, N., and C. F. Schachtele. 1990. "Effect of pH on the Growth and Proteolytic Activity of Porphyromonas Gingivalis and Bacteroides Intermedius." *Journal of Dental Research* 69 (6): 1266–69.
Takeshita, Toru, Shinya Kageyama, Michiko Furuta, Hidenori Tsuboi, Kenji Takeuchi, Yukie Shibata, Yoshihiro Shimazaki, et al. 2016. "Bacterial Diversity in Saliva and Oral Health-Related Conditions: The Hisayama Study." *Scientific Reports* 6 (February): 22164.
Takeshita, Toru, Kazuki Matsuo, Michiko Furuta, Yukie Shibata, Kaoru Fukami, Yoshihiro Shimazaki, Sumio Akifusa, et al. 2014. "Distinct Composition of the Oral Indigenous Microbiota in South Korean and Japanese Adults." *Scientific Reports* 4 (November): 6990.
Takeshita, Toru, Masaki Yasui, Yukie Shibata, Michiko Furuta, Yoji Saeki, Nobuoki Eshima, and Yoshihisa Yamashita. 2015. "Dental Plaque Development on a Hydroxyapatite Disk in Young Adults Observed by Using a Barcoded Pyrosequencing Approach." *Scientific Reports* 5 (January): 8136.
Talia, Domenico. 2019. "A View of Programming Scalable Data Analysis: From Clouds to Exascale." *Journal of Cloud*

Computing 8 (1): 1–16.
Talmaciu, I., L. Varlotta, J. Mortensen, and D. V. Schidlow. 2000. “Risk Factors for Emergence of Stenotrophomonas Maltophilia in Cystic Fibrosis.” *Pediatric Pulmonology* 30 (1): 10–15.
Tanner, A. C. R., R. L. Kent Jr, P. Lif Holgerson, C. V. Hughes, C. Y. Loo, E. Kanasi, N. I. Chalmers, and I. Johansson. 2011. “Microbiota of Severe Early Childhood Caries before and after Therapy.” *Journal of Dental Research* 90 (11): 1298–1305.
Tateno, Taro, Shinichi Ueno, Kiyokazu Hiwatashi, Masataka Matsumoto, Hiroshi Okumura, Tetsuro Setoyama, Yasuto Uchikado, et al. 2009. “Expression of Receptor for Advanced Glycation End Products (RAGE) Is Related to Prognosis in Patients with Esophageal Squamous Cell Carcinoma.” *Annals of Surgical Oncology* 16 (2): 440–46.
Tate, S., G. MacGregor, M. Davis, J. A. Innes, and A. P. Greening. 2002. “Airways in Cystic Fibrosis Are Acidified: Detection by Exhaled Breath Condensate.” *Thorax* 57 (11): 926–29.
Taylor, C. J., J. McGaw, R. Howden, B. I. Duerden, and P. S. Baxter. 1990. “Bacterial Reservoirs in Cystic Fibrosis.” *Archives of Disease in Childhood* 65 (2): 175–77.
Taylor, G. W. 2001. “Bidirectional Interrelationships between Diabetes and Periodontal Diseases: An Epidemiologic Perspective.” *Annals of Periodontology / the American Academy of Periodontology* 6 (1): 99–112.
Teles, R., and C-Y Wang. 2011. “Mechanisms Involved in the Association between Periodontal Diseases and Cardiovascular Disease.” *Oral Diseases* 17 (5): 450–61.
Templ, M., K. Hron, and P. Filzmoser, eds. 2011. “robCompositions: An R-Package for Robust Statistical Analysis of Compositional Data.” In *Compositional Data Analysis: Theory and Applications*, 341–55. John Wiley and Sons.
Thevaranjan, Netusha, Alicja Puchta, Christian Schulz, Avee Naidoo, J. C. Szamosi, Chris P. Verschoor, Dessi Loukov, et al. 2018. “Age-Associated Microbial Dysbiosis Promotes Intestinal Permeability, Systemic Inflammation, and Macrophage Dysfunction.” *Cell Host & Microbe* 23 (4): 570.
Thompson, J., and A. Pikis. 2012. “Metabolism of Sugars by Genetically Diverse Species of Oral Leptotrichia.” *Molecular Oral Microbiology*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-

1014.2011.00627.x.
Thurber, Rebecca V., Matthew Haynes, Mya Breitbart, Linda Wegley, and Forest Rohwer. 2009. “Laboratory Procedures to Generate Viral Metagenomes.” *Nature Protocols* 4 (4): 470–83.
Tian, Na, Lina Faller, Daniel A. Leffler, Ciaran P. Kelly, Joshua Hansen, Jos A. Bosch, Guoxian Wei, Bruce J. Paster, Detlef Schuppan, and Eva J. Helmerhorst. 2017. “Salivary Gluten Degradation and Oral Microbial Profiles in Healthy Individuals and Celiac Disease Patients.” *Applied and Environmental Microbiology* 83 (6). https://doi.org/10.1128/AEM.03330-16.
Tilley, Ann E., Matthew S. Walters, Renat Shaykhiev, and Ronald G. Crystal. 2015. “Cilia Dysfunction in Lung Disease.” *Annual Review of Physiology* 77: 379–406.
Tindall, B. J., and J. P. Euzéby. 2006. “Proposal of Parvimonas Gen. Nov. and Quatrionicoccus Gen. Nov. as Replacements for the Illegitimate, Prokaryotic, Generic Names Micromonas Murdoch and Shah 2000 and Quadricoccus Maszenan et Al. 2002, Respectively.” *International Journal of Systematic and Evolutionary Microbiology* 56 (Pt 11): 2711–13.
Torlakovic, Lino, Vanja Klepac-Ceraj, Bjørn Ogaard, Sean L. Cotton, Bruce J. Paster, and Ingar Olsen. 2012. “Microbial Community Succession on Developing Lesions on Human Enamel.” *Journal of Oral Microbiology* 4 (March). https://doi.org/10.3402/jom.v4i0.16125.
Torres, Pedro J., Erin M. Fletcher, Sean M. Gibbons, Michael Bouvet, Kelly S. Doran, and Scott T. Kelley. 2015. “Characterization of the Salivary Microbiome in Patients with Pancreatic Cancer.” *PeerJ* 3 (November): e1373.
Torres, Sandra R., Camila Bernardo Peixoto, Daniele Manhães Caldas, Eline Barboza Silva, Tiyomi Akiti, Márcio Nucci, and Milton de Uzeda. 2002. “Relationship between Salivary Flow Rates and Candida Counts in Subjects with Xerostomia.” *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics* 93 (2): 149–54.
Uchida-Fukuhara, Yoko, Daisuke Ekuni, Md Monirul Islam, Kota Kataoka, Ayano Taniguchi-Tabata, Daiki Fukuhara, Naoki Toyama, et al. 2020. “Caries Increment and Salivary Microbiome during University Life: A Prospective Cohort Study.” *International Journal of Environmental Research and Public Health* 17 (10). https://doi.org/10.3390/ijerph17103713.
Uranga, Carla C., Pablo Arroyo Jr, Brendan M. Duggan, William H.

Gerwick, and Anna Edlund. 2020. “Commensal Oral Rothia Mucilaginosa Produces Enterobactin, a Metal-Chelating Siderophore.” *mSystems* 5 (2). https://doi.org/10.1128/mSystems.00161-20.

Ursell, Luke K., Jessica L. Metcalf, Laura Wegener Parfrey, and Rob Knight. 2012. “Defining the Human Microbiome.” *Nutrition Reviews* 70 Suppl 1 (August): S38–44.

Urzúa, B., G. Hermosilla, J. Gamonal, I. Morales-Bozo, M. Canals, S. Barahona, C. Cóccola, and V. Cifuentes. 2008. “Yeast Diversity in the Oral Microbiota of Subjects with Periodontitis: Candida Albicans and Candida Dubliniensis Colonize the Periodontal Pockets.” *Medical Mycology: Official Publication of the International Society for Human and Animal Mycology* 46 (8): 783–93.

Utari, Putri D., Jan Vogel, and Wim J. Quax. 2017. “Deciphering Physiological Functions of AHL Quorum Quenching Acylases.” *Frontiers in Microbiology* 8 (June): 1123.

Valenza, Giuseppe, Dennis Tappe, Doris Turnwald, Matthias Frosch, Corinne König, Helge Hebestreit, and Marianne Abele-Horn. 2008. “Prevalence and Antimicrobial Susceptibility of Microorganisms Isolated from Sputa of Patients with Cystic Fibrosis.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 7 (2): 123–27.

Valitutti, Francesco, Salvatore Cucchiara, and Alessio Fasano. 2019. “Celiac Disease and the Microbiome.” *Nutrients* 11 (10). https://doi.org/10.3390/nu11102403.

Vallès, Yvonne, Claire K. Inman, Brandilyn A. Peters, Raghib Ali, Laila Abdel Wareth, Abdishakur Abdulle, Habiba Alsafar, et al. 2018. “Types of Tobacco Consumption and the Oral Microbiome in the United Arab Emirates Healthy Future (UAEHFS) Pilot Study.” *Scientific Reports* 8 (1): 11327.

Van de Guchte, Maarten, Sebastian D. Burz, Julie Cadiou, Jiangbo Wu, Stanislas Mondot, Hervé M. Blottière, and Joël Doré. 2020. “Alternative Stable States in the Intestinal Ecosystem: Proof of Concept in a Rat Model and a Perspective of Therapeutic Implications.” *Microbiome* 8 (1): 153.

Vandeplassche, Eva, Sarah Tavernier, Tom Coenye, and Aurélie Crabbé. 2019. “Influence of the Lung Microbiome on Antibiotic Susceptibility of Cystic Fibrosis Pathogens.” *European Respiratory Review: An Official Journal of the European Respiratory Society* 28 (152).

https://doi.org/10.1183/16000617.0041-2019.
Vandeputte, Doris, Gunter Kathagen, Kevin D’hoe, Sara Vieira-Silva, Mireia Valles-Colomer, João Sabino, Jun Wang, et al. 2017. “Quantitative Microbiome Profiling Links Gut Community Variation to Microbial Load.” *Nature* 551 (7681): 507–11.
Vartoukian, S. R., R. M. Palmer, and W. G. Wade. 2009. “Diversity and Morphology of Members of the Phylum ‘Synergistetes’ in Periodontal Health and Disease.” *Applied and Environmental Microbiology* 75 (11): 3777–86.
Vázquez-Baeza, Yoshiki, Antonio Gonzalez, Larry Smarr, Daniel McDonald, James T. Morton, Jose A. Navas-Molina, and Rob Knight. 2017. “Bringing the Dynamic Microbiome to Life with Animations.” *Cell Host & Microbe* 21 (1): 7–10.
Vieira Colombo, Ana Paula, Clarissa Bichara Magalhães, Fátima Aparecida Rocha Resende Hartenbach, Renata Martins do Souto, and Carina Maciel da Silva-Boghossian. 2016. “Periodontal-Disease-Associated Biofilm: A Reservoir for Pathogens of Medical Importance.” *Microbial Pathogenesis* 94 (May): 27–34.
Vielkind, Paul, Holger Jentsch, Klaus Eschrich, Arne C. Rodloff, and Catalina-Suzana Stingu. 2015. “Prevalence of Actinomyces Spp. in Patients with Chronic Periodontitis.” *International Journal of Medical Microbiology: IJMM* 305 (7): 682–88.
Vigild, M. 1986. “Dental Caries Experience among Children with Down’s Syndrome.” *Journal of Mental Deficiency Research* 30 (Pt 3) (September): 271–76.
Vitoria, Isidro, Francisco Maraver, Félix Sánchez-Valverde, and Francisco Armijo. 2015. “[Nitrate concentrations in tap water in Spain].” *Gaceta sanitaria / S.E.S.P.A.S* 29 (3): 217–20.
Vos, Michiel, Christopher Quince, Agata S. Pijl, Mattias de Hollander, and George A. Kowalchuk. 2012. “A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity.” *PloS One* 7 (2): e30600.
Wade, William G. 2013. “The Oral Microbiome in Health and Disease.” *Pharmacological Research: The Official Journal of the Italian Pharmacological Society* 69 (1): 137–43.
Walter, Jens, Anissa M. Armet, B. Brett Finlay, and Fergus Shanahan. 2020. “Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents.” *Cell* 180 (2): 221–32.

Wang, Hannah, Jessica Altemus, Farshad Niazi, Holly Green, Benjamin C. Calhoun, Charles Sturgis, Stephen R. Grobmyer, and Charis Eng. 2017. “Breast Tissue, Oral and Urinary Microbiomes in Breast Cancer.” *Oncotarget* 8 (50): 88122–38.

Wang, Laura, and Ian Ganly. 2014. “The Oral Microbiome and Oral Cancer.” *Clinics in Laboratory Medicine* 34 (4): 711–19.

Wang, Yuan, Jie Zhang, Xi Chen, Wen Jiang, Sa Wang, Lei Xu, Yan Tu, et al. 2017. “Profiling of Oral Microbiota in Early Childhood Caries Using Single-Molecule Real-Time Sequencing.” *Frontiers in Microbiology* 8 (November): 2244.

Wantland, W. W., E. M. Wantland, J. W. Remo, and D. L. Winquist. 1958. “Studies on Human Mouth Protozoa.” *Journal of Dental Research* 37 (5): 949–50.

Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane. 1991. “16S Ribosomal DNA Amplification for Phylogenetic Study.” *Journal of Bacteriology* 173 (2): 697–703.

Wei, Yiping, Meng Shi, Min Zhen, Cui Wang, Wenjie Hu, Yong Nie, and Xiaolei Wu. 2019. “Comparison of Subgingival and Buccal Mucosa Microbiome in Chronic and Aggressive Periodontitis: A Pilot Study.” *Frontiers in Cellular and Infection Microbiology* 9 (March): 53.

Wetterstrand, Kris A. 2020. “DNA Sequencing Costs: Data.” NHGRI Genome Sequencing Program (GSP). 2020. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.

Whiley, Robert A., Emily V. Fleming, Ridhima Makhija, and Richard D. Waite. 2015. “Environment and Colonisation Sequence Are Key Parameters Driving Cooperation and Competition between Pseudomonas Aeruginosa Cystic Fibrosis Strains and Oral Commensal Streptococci.” *PloS One* 10 (2): e0115513.

Whitehead, Alfred North. 1925. *Science and the Modern World*. New York: Free Press.

Whiteson, Katrine L., Barbara Bailey, Megan Bergkessel, Douglas Conrad, Laurence Delhaes, Ben Felts, J. Kirk Harris, et al. 2014. “The Upper Respiratory Tract as a Microbial Source for Pulmonary Infections in Cystic Fibrosis. Parallels from Island Biogeography.” *American Journal of Respiratory and Critical Care Medicine* 189 (11): 1309–15.

Whiteson, Katrine L., Simone Meinardi, Yan Wei Lim, Robert Schmieder, Heather Maughan, Robert Quinn, Donald R. Blake, Douglas Conrad, and Forest Rohwer. 2014. “Breath Gas

Metabolites and Bacterial Metagenomes from Cystic Fibrosis Airways Indicate Active pH Neutral 2,3-Butanedione Fermentation." *The ISME Journal* 8 (6): 1247–58.
Whittaker, Robert Harding, and Others. 1975. *Communities and Ecosystems*. New York: MacMillan Publishing Company, Inc.
Widdel, Friedrich, and Norbert Pfennig. 1982. "Studies on Dissimilatory Sulfate-Reducing Bacteria That Decompose Fatty Acids II. Incomplete Oxidation of Propionate by Desulfobulbus Propionicus Gen. Nov., Sp. Nov." *Archives of Microbiology* 131 (4): 360–65.
Wilbert, Steven A., Jessica L. Mark Welch, and Gary G. Borisy. 2020. "Spatial Ecology of the Human Tongue Dorsum Microbiome." *Cell Reports* 30 (12): 4003.
Willis, Jesse R., and Toni Gabaldón. 2020. "The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems." *Microorganisms* 8 (2). https://doi.org/10.3390/microorganisms8020308.
Willis, Jesse R., Pedro González-Torres, Alexandros A. Pittis, Luis A. Bejarano, Luca Cozzuto, Nuria Andreu-Somavilla, Miriam Alloza-Trabado, et al. 2018. "Citizen Science Charts Two Major 'Stomatotypes' in the Oral Microbiome of Adolescents and Reveals Links with Habits and Drinking Water Composition." *Microbiome* 6 (1): 218.
Willis, Jesse R., Susana Iraola-Guzmán, Ester Saus, Ewa Ksiezopolska, Luca Cozzuto, Luis A. Bejarano, Nuria Andreu-Somavilla, et al. 2020. "Oral Microbiome in down Syndrome and Its Implications on Oral Health." *Journal of Oral Microbiology* 13 (1): 1865690.
Willis, Jesse R., Ester Saus, Susana Iraola-Guzmán, Elena Cabello-Yeves, Ewa Ksiezopolska, Luca Cozzuto, Luis A. Bejarano, et al. 2021. "Citizen-Science Based Study of the Oral Microbiome in Cystic Fibrosis and Matched Controls Reveals Major Differences in Diversity and Abundance of Bacterial and Fungal Species." *Journal of Oral Microbiology* 13 (1): 1897328.
Willner, Dana, Mike Furlan, Matthew Haynes, Robert Schmieder, Florent E. Angly, Joas Silva, Sassan Tammadoni, Bahador Nosrat, Douglas Conrad, and Forest Rohwer. 2009. "Metagenomic Analysis of Respiratory Tract DNA Viral Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals." *PloS One* 4 (10): e7370.

Wilson, Michael. 2009. *Bacteriology of Humans: An Ecological Perspective*. Hoboken, NJ: John Wiley & Sons.

Winter-de Groot, Karin M. de, and Cornelis K. van der Ent. 2005. “Nitric Oxide in Cystic Fibrosis.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 4 Suppl 2 (August): 25–29.

Wolff, B., S. Boutin, H-M Lorenz, H. Ueffing, A. Dalpke, and D. Wolff. 2017. “FRI0698 Prevotella and Alloprevotella Species Characterize the Oral Microbiome of Early Rheumatoid Arthritis.” *Annals of the Rheumatic Diseases* 76 (Suppl 2): 754–754.

Wolf, Kyle J., Joseph G. Daft, Scott M. Tanner, Riley Hartmann, Ehsan Khafipour, and Robin G. Lorenz. 2014. “Consumption of Acidic Water Alters the Gut Microbiome and Decreases the Risk of Diabetes in NOD Mice.” *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society* 62 (4): 237–50.

Woo, Jennifer S., and David Y. Lu. 2019. “Procurement, Transportation, and Storage of Saliva, Buccal Swab, and Oral Wash Specimens.” *Methods in Molecular Biology* 1897: 99–105.

Worlitzsch, D., C. Rintelen, K. Böhm, B. Wollschläger, N. Merkel, M. Borneff-Lipp, and G. Döring. 2009. “Antibiotic-Resistant Obligate Anaerobes during Exacerbations of Cystic Fibrosis Patients.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 15 (5): 454–60.

Wright, Erik S. 2016. “Using DECIPHER v2. 0 to Analyze Big Biological Sequence Data in R.” *The R Journal* 8 (1). https://journal.r-project.org/archive/2016/RJ-2016-025/index.html.

Wu, Guangxi, He Zhao, Chenhao Li, Menaka Priyadarsani Rajapakse, Wing Cheong Wong, Jun Xu, Charles W. Saunders, et al. 2015. “Genus-Wide Comparative Genomics of Malassezia Delineates Its Phylogeny, Physiology, and Niche Adaptation on Human Skin.” *PLoS Genetics* 11 (11): e1005614.

Wu, Jing, Brandilyn A. Peters, Christine Dominianni, Yilong Zhang, Zhiheng Pei, Liying Yang, Yingfei Ma, et al. 2016. “Cigarette Smoking and the Oral Microbiome in a Large Study of American Adults.” *The ISME Journal* 10 (10): 2435–46.

Wu, T., and L. P. Samaranayake. 1999. “The Expression of Secreted Aspartyl Proteinases of Candida Species in Human Whole Saliva.” *Journal of Medical Microbiology* 48 (8): 711–20.

Wylie, Kristine M., George M. Weinstock, and Gregory A. Storch. 2012. “Emerging View of the Human Virome.” *Translational Research: The Journal of Laboratory and Clinical Medicine* 160 (4): 283–90.

Xiao, E., Marcelo Mattos, Gustavo Henrique Apolinário Vieira, Shanshan Chen, Jôice Dias Corrêa, Yingying Wu, Mayra Laino Albiero, Kyle Bittinger, and Dana T. Graves. 2017. “Diabetes Enhances IL-17 Expression and Alters the Oral Microbiome to Increase Its Pathogenicity.” *Cell Host & Microbe* 22 (1): 120–28.e4.

Xiao, Jin, Xuelian Huang, Naemah Alkhers, Hassan Alzamil, Sari Alzoubi, Tong Tong Wu, Daniel A. Castillo, et al. 2018. “Candida Albicans and Early Childhood Caries: A Systematic Review and Meta-Analysis.” *Caries Research* 52 (1-2): 102–12.

“XIT Genomic DNA from Buccal Cells: For Extraction of Genomic DNA from Buccal/Cheek Cells.” 2012. G-Biosciences. https://www.gbiosciences.com/image/pdfs/protocol/786-341_protocol.pdf.

Xu, He, Wenjing Hao, Qiong Zhou, Wenhong Wang, Zhongkui Xia, Chuan Liu, Xiaochi Chen, Man Qin, and Feng Chen. 2014. “Plaque Bacterial Microbiome Diversity in Children Younger than 30 Months with or without Caries prior to Eruption of Second Primary Molars.” *PloS One* 9 (2): e89269.

Xu, Lei, Xi Chen, Yuan Wang, Wen Jiang, Sa Wang, Zongxin Ling, and Hui Chen. 2018. “Dynamic Alterations in Salivary Microbiota Related to Dental Caries and Age in Preschool Children With Deciduous Dentition: A 2-Year Follow-Up Study.” *Frontiers in Physiology* 9 (April): 342.

Xu, Xin, Jinzhi He, Jing Xue, Yan Wang, Kun Li, Keke Zhang, Qiang Guo, et al. 2015. “Oral Cavity Contains Distinct Niches with Dynamic Microbial Communities.” *Environmental Microbiology* 17 (3): 699–710.

Yamashita, Yoshihisa, and Toru Takeshita. 2017. “The Oral Microbiome and Human Health.” *Journal of Oral Science* 59 (2): 201–6.

Yang, Andrian, Michael Troup, and Joshua W. K. Ho. 2017. “Scalability and Validation of Big Data Bioinformatics

Software." *Computational and Structural Biotechnology Journal* 15 (July): 379–86.

Yatsunenko, Tanya, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, et al. 2012. "Human Gut Microbiome Viewed across Age and Geography." *Nature* 486 (7402): 222–27.

Yoon, Hwan Su, Dana C. Price, Ramunas Stepanauskas, Veeran D. Rajah, Michael E. Sieracki, William H. Wilson, Eun Chan Yang, Siobain Duffy, and Debashish Bhattacharya. 2011. "Single-Cell Genomics Reveals Organismal Interactions in Uncultivated Marine Protists." *Science* 332 (6030): 714–17.

Yuan, Zhihong, Dipti Panchal, Mansoor Ali Syed, Hiren Mehta, Myungsoo Joo, Walid Hadid, and Ruxana T. Sadikot. 2013. "Induction of Cyclooxygenase-2 Signaling by Stomatococcus Mucilaginosus Highlights the Pathogenic Potential of an Oral Commensal." *Journal of Immunology* 191 (7): 3810–17.

Yu, Guoqin, Stephen Phillips, Mitchell H. Gail, James J. Goedert, Michael S. Humphrys, Jacques Ravel, Yanfang Ren, and Neil E. Caporaso. 2017. "The Effect of Cigarette Smoking on the Oral and Nasal Microbiota." *Microbiome* 5 (1): 3.

Zakharkina, Tetyana, Elke Heinzel, Rembert A. Koczulla, Timm Greulich, Katharina Rentz, Josch K. Pauling, Jan Baumbach, et al. 2013. "Analysis of the Airway Microbiota of Healthy Individuals and Patients with Chronic Obstructive Pulmonary Disease by T-RFLP and Clone Sequencing." *PloS One* 8 (7): e68302.

Zambirinis, Constantinos P., Elliot Levie, Susanna Nguy, Antonina Avanzi, Rocky Barilla, Yijie Xu, Lena Seifert, et al. 2015. "TLR9 Ligation in Pancreatic Stellate Cells Promotes Tumorigenesis." *The Journal of Experimental Medicine* 212 (12): 2077–94.

Zaura, Egija, Bernd W. Brandt, Andrei Prodan, Maarten Joost Teixeira de Mattos, Sultan Imangaliyev, Jolanda Kool, Mark J. Buijs, et al. 2017. "On the Ecosystemic Network of Saliva in Healthy Young Adults." *The ISME Journal* 11 (5): 1218–31.

Zaura, Egija, Bart J. F. Keijser, Susan M. Huse, and Wim Crielaard. 2009. "Defining the Healthy 'Core Microbiome' of Oral Microbial Communities." *BMC Microbiology* 9 (December): 259.

Zemanick, Edith T., Brandie D. Wagner, Charles E. Robertson, Richard C. Ahrens, James F. Chmiel, John P. Clancy, Ronald L. Gibson, et al. 2017. "Airway Microbiota across Age and

Disease Spectrum in Cystic Fibrosis.” *The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology* 50 (5). https://doi.org/10.1183/13993003.00832-2017.

Zetterquist, Wilhelm, Helena Marteus, Pia Kalm-Stephens, Elisabeth Näs, Lennart Nordvall, Marie Johannesson, and Kjell Alving. 2009. “Oral Bacteria--the Missing Link to Ambiguous Findings of Exhaled Nitrogen Oxides in Cystic Fibrosis.” *Respiratory Medicine* 103 (2): 187–93.

Zhang, Chen-Zi, Xing-Qun Cheng, Ji-Yao Li, Ping Zhang, Ping Yi, Xin Xu, and Xue-Dong Zhou. 2016. “Saliva in the Diagnosis of Diseases.” *International Journal of Oral Science* 8 (3): 133–37.

Zhang, Xiao-Bing, Shuang-Jiang Yu, Jun-Xiao Yu, Ya-Li Gong, Wei Feng, and Feng-Jun Sun. 2012. “Retrospective Analysis of Epidemiology and Prognostic Factors for Candidemia at a Hospital in China, 2000-2009.” *Japanese Journal of Infectious Diseases* 65 (6): 510–15.

Zhang, Xuan, Dongya Zhang, Huijue Jia, Qiang Feng, Donghui Wang, Di Liang, Xiangni Wu, et al. 2015. “The Oral and Gut Microbiomes Are Perturbed in Rheumatoid Arthritis and Partly Normalized after Treatment.” *Nature Medicine* 21 (8): 895–905.

Zhang, Xuyun, Huiwen Chen, Weili Lu, Lei Zhu, Wei Zhou, and Zhongchen Song. 2020. “Characterization of the Subgingival Microbiota in the Peritoneal Dialysis Patients with Periodontitis.” *Archives of Oral Biology* 115 (July): 104742.

Zhang, Y., B. Doranz, J. R. Yankaskas, and J. F. Engelhardt. 1995. “Genotypic Analysis of Respiratory Mucous Sulfation Defects in Cystic Fibrosis.” *The Journal of Clinical Investigation* 96 (6): 2997–3004.

Zhao, Jiangchao, Patrick D. Schloss, Linda M. Kalikin, Lisa A. Carmody, Bridget K. Foster, Joseph F. Petrosino, James D. Cavalcoli, et al. 2012. “Decade-Long Bacterial Community Dynamics in Cystic Fibrosis Airways.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (15): 5809–14.

Zhou, Jianye, Nan Jiang, Shaoguo Wang, Xiaopan Hu, Kangli Jiao, Xiangyi He, Zhiqiang Li, and Jizeng Wang. 2016. “Exploration of Human Salivary Microbiomes--Insights into the Novel Characteristics of Microbial Community Structure in Caries and Caries-Free Subjects.” *PloS One* 11 (1): e0147039.

Zhou, Yi-Hui, and Paul Gallins. 2019. “A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction.” *Frontiers in Genetics* 10 (June): 579.

Zielenski, J., R. Rozmahel, D. Bozon, B. Kerem, Z. Grzelczak, J. R. Riordan, J. Rommens, and L. C. Tsui. 1991. “Genomic DNA Sequence of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene.” *Genomics* 10 (1): 214–28.