UNIVERSITY OF TRENTO

PhD Program in Biomolecular Sciences – 35th Cycle

CIBIO – Department of Cellular, Computational and Integrative Biology

---

The CHRIS Salivary Microbiome

Characterization of the salivary microbiome in a large sample of South Tyrolean adults in relation to lifestyle, environment, and genetics

---

Tutor

Dr. **Christian Fuchsberger**

Institute for Biomedicine, Eurac Research

Advisor

Prof. **Nicola Segata**

CIBIO, University of Trento

Ph.D. Thesis of **Giacomo Antonello**

Institute for Biomedicine, Eurac Research

CIBIO, University of Trento

Academic Year 2021-2022

Declaration

I, Giacomo Antonello, confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Abstract

The oral microbiome is a key component of the human body and has been associated with several habits and diseases. Despite its important role in health, it remains relatively understudied, compared to the gut microbiome.

To deepen our understanding of the oral microbiome and its links to host conditions, the main aim of my PhD thesis was to characterize the lifestyle, environmental and genetic determinants of the salivary microbiome using data from CHRISMB, a convenience sample within the Cooperative Health Research in South Tyrol (CHRIS) study. With more than 1,900 samples, CHRISMB is one of the largest salivary microbiome data resources in the world.

First, I studied the association between the salivary microbiome and smoking status and degree of exposure both from the compositional and predicted metabolism perspective. I found associations with 44 genera, 11 of which were also proportionally affected by the degree of exposure to tobacco. Intriguingly, these associations highlight a novel role of salivary microbiome metabolism in cardiovascular diseases through periodontium degeneration via the nitrate reduction and extracellular matrix degradation pathways.

My second contribution focused on the role of geography, family relatedness, and genetics in shaping CHRISMB diversity. I investigated the associations between household, municipality and altitude of residence, heritability, and genetic marker associations (mbGWAS). I confirmed that cohabitation is a strong driver of microbiome similarity, while municipality and altitude of residence did not show strong associations. Siblings living apart had a more similar microbiota than unrelated and non-cohabiting individuals. Sixteen out of 142 taxa had a significant heritability component, while 34 had a significant household component. A mbGWAS Gene-level analysis resulted in one association between rare variants in the *SRFBP1* and *LOX* genes locus and *Selenomonas noxia*. This work confirmed that host genetics and familial relationships has a modest but significant association with the salivary microbiome composition and that the environment and lifestyle are strongly associated.

In summary, this thesis deepens our understanding of population-level factors associated with salivary microbiome variability, which can help design future hypothesis driven studies.

LIST OF ABBREVIATIONS

1000G	1000 Genomes Project
16S	16S rRNA gene Amplicon Sequencing
4D-SZ	Chinese cohort from the Shenzen, China
ASV	Amplicon Sequence Variant
BH	Benjamini-Hochberg
BMI	Body Mass Index
c^2	Household Component (heritability analysis)
CHRIS	Cooperative Health Research in South Tyrol study
CHRISMB	CHRIS study microbiota sample
DMP	Dutch Microbiome Project
DNA	Deoxyribonucleic acid
eHOMD	Expanded Human Oral Microbiome Database
FDR	False Discovery Rate
FMT	Fecal Material/Matter Transplant
g	Gram/Grams
GLM	Generalized Linear Model
GLS	Gene-Level Significant
GWS	Genome-Wide Significant
h^2	Heritability Component (heritability analysis)
HMP	Human Microbiome Project
HRC	Haplotype Reference Consortium
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
mbGWAS	Microbiome Genome-Wide Association Study/Scan
NGS	Next-Generation Sequencing
NO	Nitric Oxide
NRB	Nitrate-reducing Bacteria
PCR	Polymerase Chain Reaction
PCoA	Principal Coordinate Analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
QTL	Quantitative Trait Locus/Loci
R^2	Proportion of Variance Explained
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
South Tyrol Healthcare System	SABES/ASDAA
S.E.	Standard Error
SGB	Species-Genome Bin
SWS	Study-Wide Significant
TOPMed	Trans-Omics for Precision Medicine Program
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Table of Contents

Abstract 2

List of Abbreviations 3

1 Introduction 9

1.1 The microbiome: definition and evolutionary considerations 9

1.2 Methods to study the microbiome 10

1.2.1 Microbial Culturing 10

1.2.2 Next-Generation Sequencing 11

1.3 The human microbiome in health and disease 13

1.4 Aims of this thesis 14

2 Background 15

2.1 The gut microbiome from ecological considerations to biotechnological applications 15

2.2 The oral cavity harbors a rich microbiome to study 16

2.3 The Cooperative Health Research in South Tyrol study is a rich resource for salivary microbiome 18

2.4 Host genetic factors associated with gut and oral microbiome composition 20

2.5 Genome wide association studies and heritability analysis are tools to explore traits' genetic component in a population 20

2.5.1 Heritability Analysis 21

2.5.2 Genome Wide Association Study 21

2.5.3 Genotype imputation is a cost efficient method to gain detection power in a GWAS 22

2.6 Thesis outline and scientific contribution 23

Chapter 3 – Smoking and salivary microbiota: a cross-sectional analysis of an Italian alpine population 23

Chapter 4 – Geographic and genetic factors on the composition of the adult salivary microbiota of an Italian alpine population 24

Chapter 5 – Additional Contributions 24

3 Smoking and salivary microbiota: a cross-sectional analysis of an Italian alpine population 27

Abstract 27

Highlights 28

3.1 Introduction 29

3.2 Results 30

3.2.1 Characteristics of study population in relation to smoking 30

3.2.2 Qualitative smoking habits are associated with compositional and functional profiles of salivary genera 32

3.2.3 Several microbial genera associated with smoking habits are also associated with the grams of tobacco smoked daily 34

3.2.4 Salivary microbiota of Former smokers who quit 5 years or longer tended to resemble Never smokers' profiles 36

3.2.5 Predicted Functional Profiles Associated with Smoking highlighted a decrease of aerobic and nitrate reducing taxa 38

3.3 Discussion 39

3.3.1 Summary of study and main results 39

3.3.2 Comparison with other studies 39

3.3.3 Salivary microbial genera composition and proportion of aerobes were strongly impacted by smoking 40

3.3.4 Some genera are statistically associated with daily smoking intensity but not with the years since smoking cessation 41

3.3.5 Study Limitations 41

3.3.6 Study Strengths 41

3.4 Conclusions 42

3.5 Acknowledgements 42

3.6 Funding Sources 42

3.7 Data and Analysis Scripts Availability 43

3.8 Materials and Methods 43

3.8.1 Study ethical approval, design, and data collection 43

3.8.2 Salivary microbiota data generation 44

3.8.3 Sequencing data processing 45

3.9 Samples availability and statistical analysis 46

3.9.1 Statistical analysis 46

3.10 Supplementary files (Links) 47

4 Geographic and genetic factors on the composition of the adult salivary microbiota of an Italian alpine population 48

Abstract 48

4.1 Introduction 49

4.2 Results 50

4.2.1 The majority of CHRISMB participants live in households with up to 3 participants and reside mostly in Mals and Graun 50

4.2.2 The salivary microbiota varies largely in relation to cohabitation and marginally to municipality and altitude of residence 52

4.2.3 Familial relationships did not associate with higher salivary microbiota similarity 55

4.2.4 mbGWAS reveals several ASV-variant associations 57

4.3 Discussion 58

4.3.1 The salivary microbiota varies largely in relation to cohabitation and marginally to municipality and altitude of residence 59

4.3.2 The salivary microbiota variability shows contributions of familial relatedness and households 60

4.3.3 The salivary microbiota is mildly associated with host genetics 60

4.3.4 Limitations of the study 61

4.3.5 Strengths of the study 62

4.4 Conclusions 62

4.5 Acknowledgements 62

4.6 Methods 63

4.6.1 Study design, ethical approval, and data collection 63

4.6.2 Epidemiological data generation 63

4.6.3 Salivary microbiota data generation 63

4.6.4 Sample selection and platform for statistical analysis 64

4.6.5 Microbiota analyses 64

4.6.6 Beta diversity Matrix calculation 64

4.6.7 Permutational Multivariate Analysis of Variance (PERMANOVA) 64

4.6.8 Differential abundance analysis 65

4.6.9 Microbiota variation in relation to geographic exposure: household, municipality and altitude of residence 65

4.6.10 Microbiota variability associated with familial relatedness (heritability analysis) 66

4.6.11 Association between salivary microbiota and host genetic variants (mbGWAS) 66

4.6.12 Association between rare variants and microbiota features (Gene-level Analysis) 67

4.7 Supplementary material 68

4.7.1 Supplementary Figures 68

4.7.2 Supplementary Tables 72

5 Additional contributions 81

5.1 Odor identification capabilities and salivary microbiota composition are not significantly associated in CHRISMB 81

5.1.1 Background 81

5.1.2 Methods 81

5.1.3 Results 82

5.1.4 Conclusions 84

5.2 Salivary microbiota exposome scan 84

5.2.1 Background 84

5.2.2 Methods 84

5.2.3 Results 85

5.2.4 Conclusions 86

6 General Conclusions 87

7 Final Acknowledgements /Ringraziamenti Finali 90

Bibliography 92

1 INTRODUCTION

This chapter introduces the reader to the concept of microbiome and the most adopted methods to generate microbiome data. Additionally, it highlights cases in which the microbiome was shown to have pivotal role in biological systems and ends with the aims of my thesis.

1.1 THE MICROBIOME: DEFINITION AND EVOLUTIONARY CONSIDERATIONS

After inventing the microscope around 1663, Antonijn van Leeuwenhoek described tiny “animalcules”, now known as microbes, from several samples, including stool and dental calculus, giving birth to microbiology. In the first two centuries microbiology discoveries with higher resonance were related to pathogenic microbes. In particular, thanks to R. Koch’s contribution, by the second half of the 19th century the bacterial etiology of cholera, anthrax and tuberculosis was proven [1]. Less than 100 years later, it was rather established that only a small fraction of microorganisms was harmful to humans; in fact, most microbes were harmless, and some could even be beneficial. For example, it is now known that the *Clostridium* genus contains species that are highly pathogenic (*C. perfringens* [2]) or even deadly (*C. tetani* [3] and *C. botulinum* [4]). However, several other *Clostridium* species make up a considerable portion of the microbes in the human intestine (*C. coccoides, C. leptum*) [5] and even contribute to useful metabolic features [6]. The usefulness of specific gut bacteria was discovered in 1917, when Alfred Nissle isolated a bacterium from the stool of Prussian soldiers who were resistant to a typhus outbreak in his hospital. More interestingly, he inoculated this bacterium to healthy soldiers, showing that its carriers were protected from typhus: *E. coli* strain “Nissle 1917” was the first probiotic ever reported [7]. Building on top of multifaceted interaction between microbes and their habitats, in 1988 J. M. Whipps and colleagues contributed to a chapter in a book related to the ecology and biocontrol of fungi with the following consideration [8]:

“A convenient ecological framework in which to examine biocontrol systems is that of the microbiome. This may be defined as a characteristic microbial community occupying a reasonably well-defined habitat which has distinct physio-chemical properties. This term thus not only refers to the microorganisms involved but also encompasses their theatre of activity.”

Their contribution was a milestone for microbiome research field because microbes were viewed not only as organisms in an environment, but rather generating a "theatre of activity", a network of molecular interaction between other microbes and with their environment by means of coexistence, antagonism, and synergism [9]. After the year 2000, with the advent of high throughput DNA sequencing – or Next-Generation Sequencing (NGS) – researchers were able to discover a new level of complexity in the microbiome field, with organisms that had never been characterized by traditional cultivation and microscopy techniques in use to that day. Microbiomes do not include only bacteria, but also organisms from other kingdoms. The human gut microbiome, for instance, also comprises unicellular eukaryotes (protists and amoebas) [10], fungi [11], and archaea [12]. Even viruses, especially bacteriophages, are part of gut microbiomes and have a role in controlling community stability [13]. From a human health perspective, it is important to note that our bodies harbor slightly more microbial cells than human cells [14]. This has led to the idea that our microbiome, with its complex biochemical activity, should be considered an additional organ of the body [15]. As a consequence, humans could be considered a sum of human and microbial cells and genetic features: holobionts carrying hologenomes [16,17]. Indeed, research has shown that the metabolic activity of the gut microbiome results in increased energy harvest from food [18]. More interestingly, studies on human and mouse models showed that a high proportion of human blood metabolites, as high as 30%, is derived from gastro-intestinal absorption [19]. A few notable examples of beneficial by-products are short chain fatty acids [20–22], serotonin [23,24], and vitamins [25]. However, some microbiome by-products have been considered harmful, for example Trimethylamine Oxide, Quinolinate, Indole [26], and kynurenine [23].

1.2 METHODS TO STUDY THE MICROBIOME

The study of the microbiome requires methods able to capture its complexity and dynamic variations over time. To date, the two commonly used techniques for observing and studying microbial communities are based on cultivation and DNA sequencing. The latter further separates into 16 ribosomal RNA amplicon sequencing (in brief, 16S) and shotgun metagenomics. A brief description of each method will follow, with a qualitative comparison in **Table 1-1**.

1.2.1 Microbial Culturing

Initially, the field of microbiology relied on culture-based isolation techniques to investigate their physio-chemical features.

Historically, this technique proved effective in the investigation not only of harmful bacteria, like typhus, cholera and those responsible for food spoilage [27], but also of beneficial bacteria like *E. coli* Nissle 1917 [7] and lactic acid bacteria [28], now used as probiotics.

Traditionally, bacterial cultures could be solid or liquid, generally in the presence of atmospheric pressure and oxygen. The more conditions researchers were able to control, the wider was the range of bacteria to culture and characterize [29]. With the microbiome field, not only physio-chemical conditions were exploited, but also inter-species interactions were studied [30,31]. Culture-based characterization continues to be the gold-standard method to validate *in silico* predictions of the biochemical phenotype of a single microbial isolate [32–34] or simple microbial communities [35,36]. The throughput of such methods, however, is generally low. To tackle this limitation, several methods have been developed, for instance the Biolog [37,38] and high throughput culturing [39,40].

1.2.2 Next-Generation Sequencing

The field of the microbiome took a turn with the advent of NGS technologies, which allowed researchers to discover that the vast majority of micro-organisms in a sample were unculturable, to that day, under traditional laboratory conditions [41], opening the field for culture-free microbial ecology studies [42]. There have been two major NGS-based methods to study microbiomes, in chronological order: (1) amplicon sequencing and (2) shotgun metagenomics.

16S ribosomal RNA gene amplicon Sequencing (16S)

Amplicon sequencing was first conceptualized by Woese and Fox in 1977 as a method to classify bacteria and a few archaea [43] using genetic instead of phenotypic features [44]. The method relies on the architecture of the 16S ribosomal RNA gene, which contains regions at high mutation rates flanked by highly conserved ones. Amplification with a Polymerase Chain Reaction (PCR) between the conserved regions results in a library of variable sequences which, coupled with NGS, allows to discriminate bacteria and some archaea up to the genus level [45]. With recent advances in PCR and sequencing accuracy, as well as computational algorithms, species level discrimination can be sometimes reached [46]. This technique has two main limitations. First, the absence of conserved genes across all kingdoms makes amplicon sequencing unable to capture the community complexity of microbiomes, in particular eukaryotes, which have different genes, and viruses, characterized by a vast phylogenetic and compositional variability across individuals [47,48].

Second, the polymerase enzymes can introduce point mutations, which increase noise and technical biases [49,50].

Shotgun Metagenomics

Shotgun metagenomics is a method to sequence virtually all DNA sequences in a sample, instead of a pre-amplified region like 16S. After DNA extraction, differently from 16S, DNA in a sample is fragmented and labeled. Depending on the protocol, labeling is achieved with or without a PCR step. Protocols up to this point have been benchmarked by Tourlousse et al. in 2021 [51]. Given the unbiased labeling of DNA sequences, contaminations are likely. For this reason, it is crucial that sequencing is performed at a high depth, that is with several cycles, to ensure high signal-to-noise ratios. With the proper processing and analysis tools of the sequencing data, shotgun metagenomic sequencing allows the discrimination of bacteria at taxonomic resolution impossible for 16S [52,53]. Remarkably, 16S cannot discriminate *Escherichia and Shigella* genera of the Enterobacteriaceae family, due to a high genomic similarity [54]. Conversely, shotgun metagenomics not only differentiates species within each of these genera, but can also differentiate strains within the *E. coli* species, which enables the distinction between the probiotic Nissle 1917 [55] and other pathogenic *E. coli* [56,57].

Table 1-1. Comparison of the 3 most popular methods to investigate microbes and microbiomes to date. The table presents a macroscopic comparison of the methods, highlighting key differences for easy comprehension.

	Traditional Culturing	**16S Amplicon Sequences**	**Shotgun Metagenomics**
Cost (US $)	NA	~ 50	75+
Taxonomic resolution	Species and Strains	Genus, sometimes species depending on regions targeted, pipelines and database chosen	Species and strains
Taxonomic identification	All organisms, depending on cultivability	Bacteria and Archaea	All organisms with stable DNA in the sample, including DNA viruses
Functional/Metabolic profiles	Yes, direct metabolic evidence	Only “most likely”, with PICRUSt2 [58]	Potential with many tools [59–61]
Throughput	Low	High	Very high

1.3 THE HUMAN MICROBIOME IN HEALTH AND DISEASE

Given the specific physio-chemical differences between human body sites, the oral cavity [62,63], the vagina [64], the skin [65] and breast milk [66] harbor distinct, stable microbiome communities. Presently, the stool microbiome is the most well-studied in terms of number of publications [67], strength [68,69] and reproducibility [70] of findings. Despite being highly individual specific [71], the gut microbiome of healthy individuals was shown to vary in relation to several factors including age, diet and continent of residence [72,73]. Large cross-sectional studies have generated valuable knowledge about the factors associated with the gut microbiome compositions: the Dutch Microbiome Project (n = 8,800) [74], the American Gut Project (n > 10,000) [75] and several independent Chinese gut microbiome datasets (n > 2,000) [76,77]. Moreover, the stool microbiome was reported to be significantly altered in a plethora of disease and conditions, for instance with gastrointestinal [78–80], metabolic [81–83], and neurological [84,85] conditions.

While the gut microbiome is the most well characterized and studied, the oral cavity harbors what is considered the second most diverse human microbiome [62]. Similarly to the gut microbiome, the oral microbiome is stable [62] and individualized [86], which leads, at population level, to high compositional variability. To date, few oral microbiome cohorts with at least 1,000 participants have been published: in decreasing size, the 4D-SZ tongue dorsum and salivary microbiome cohort (n = 3,504) [76], followed by low-income African Americans salivary microbiome dataset (n = 1,616) [87,88], Japanese Cohort (n ~ 700) [89] and the oral samples in the Human Microbiome Project (n = 242) [90]. Given the rising importance of the oral microbiome in not only oral, but also systemic health (see Background), it is valuable to expand the knowledge on the factors influencing its compositional variability across the widest number of conditions.

1.4 AIMS OF THIS THESIS

In this PhD thesis, I extend the population-level knowledge on the factors associated with salivary microbiome composition of an adult European population. My aims are to compare its composition and response to widely known factors influencing it (smoking, oral health, households…), as well as expanding the knowledge of potential other factors (genetics, geographic distribution…). More in detail, my aims are to:

1. Describe the salivary microbiome variation in relation to smoking not only as a qualitative exposure (Never, Former, Current), but also as a quantitative trait (cigarettes per day and years since quitting) both from the taxonomic and from the predicted metabolic function perspective.
2. Explore the role of place of residence, familial relationships, and genetics in relation to the salivary microbiome composition, with particular focus on household, municipality and altitude of residence (geographic factors), as well as relatedness, and genetic factors.

2 BACKGROUND

This chapter describes the gut and oral microbiome research. It introduces the CHRISMB study, the salivary microbiome convenience sample nested in the Cooperative Research in South Tyrol (CHRIS) study. Since one chapter of my thesis utilizes familial and genetic analysis methods, heritability, genome-wide association studies and genotype imputation are also discussed. This chapter ends with an outline of the research sections of the thesis.

2.1 THE GUT MICROBIOME FROM ECOLOGICAL CONSIDERATIONS TO BIOTECHNOLOGICAL APPLICATIONS

Human microbiomes have been associated with several habits and diseases, leading to the definition of dysbiosis whenever microbiomes in individuals with a disease were significantly different from the healthy counterparts [91]. The gut microbiome was shown to have not only associative, but also causal or concurring roles in the onset and progression of several conditions and diseases [92–94]. To prove causality, researchers have developed several methods [95]. Among these, fecal microbiome/material transplantation (FMT) on germ-free mice provided compelling evidence regarding the role of the gut microbiome on energy harvest and obesity [96]. Moreover, FMT from a colorectal cancer patient induced the development of more tumors than FMT from a healthy donor in a mouse model prone to colorectal and mammarian cancer [97]. More specifically, a set of bacteria - first of all *F. nucleatum* - were found enriched in colorectal cancer patients [98], and immunosuppressing capabilities by this species have been linked to cancer progression [99]. To corroborate the existence of a gut-brain axis, FMT from depressed patients induced a depressive behavior in germ-free mice [100]. Once the microbiome etiology of diseases was proven, researchers explored the therapeutic potential of FMT from healthy donors. For example, a generalized trend of transient improvement of depression scores was reported in a review on pre-clinical and clinical FMT to limit depression [101]. While FMT seems a promising direction towards microbiome-based therapeutics, it has been approved only in chronic infection by *Clostridioides difficile* and not yet in Europe [102]. One of the reasons for the reluctancy is the low success rate outside the *C. difficile* case [69] and the chances to inherit unexpected and unwanted phenotypes along with the desired ones (for instance, pathogens glucose sensitivity, tendency to gain weight, etc.) [103].

2.2 The oral cavity harbors a rich microbiome to study

The oral microbiome harbors a highly diverse and relatively understudied host-microbiome interaction environment, relative to the gut, mainly temperature, moisture content, pH, and oxygen partial pressure [104]. In one of the pioneering 16S oral microbiome profiling experiments, the microbiome of 9 oral niches of 3 healthy Caucasian males were sampled in Amsterdam, showing for one of the first time the average relative composition: phyla Firmicutes (genus *Streptococcus, Granulicatella*), Proteobacteria (genus *Neisseria*, *Haemophilus*) Actinobacteria (genus *Corynebacterium*, *Rothia*, *Actinomyces*), Bacteroidetes (genus *Prevotella*, *Capnocytophaga, Porphyromonas*) and Fusobacteria (genus *Fusobacterium*) [105]. They also highlighted compositional differences between oral niches, which was later confirmed in two other cohorts from the United States [90] and China [106] and reviewed by Li et al. [63]. More specifically, except for saliva, in which microbes are in a liquid solution, the oral microbiome is organized in biofilms, called plaque, adhering to hard and soft surfaces [107]. Different surfaces harbor different microbiome profiles, therefore plaque samples from subgingival, supragingival, tongue dorsum mucosae and teeth surfaces can be distinguished [90]. Plaque assembly follows a colonization based on the metabolic requirements of each bacteria, or bacterial taxa. This was visualized at genus level with fluorescence hybridization microscopy, discovering structures that resembled a hedgehog, a cauliflower or corncob [108] (**Figure 2-1**).

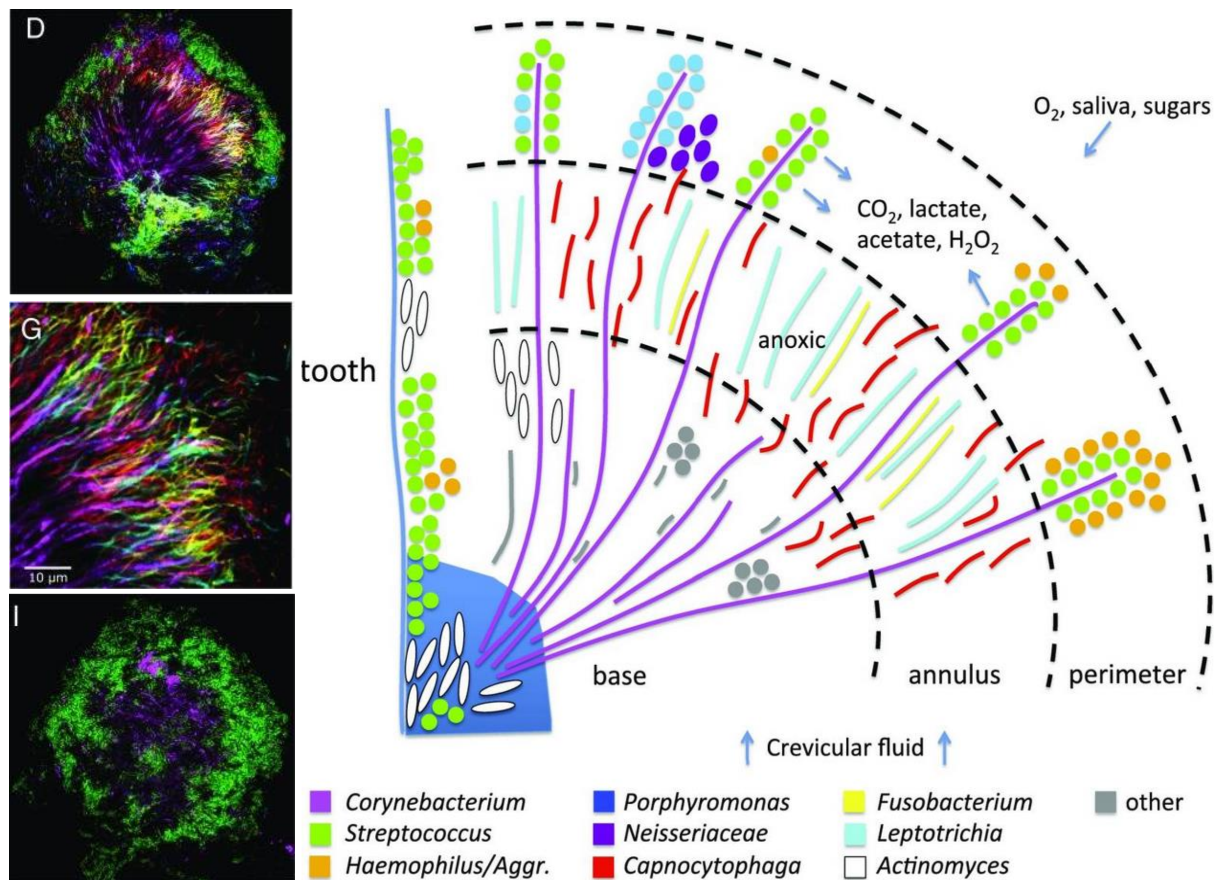


Figure 2-1. Bidimensional organization of bacterial genera in tooth plaque samples. This figure was adapted from figure 3 D, G, I and figure 9 in Mark-Welch et al., "Biogeography of a human oral microbiome at the micron scale", published on PNAS [108].

Despite the high variability, in the last 10 years several associations between the oral microbiome and oral condition like halitosis [109] and diseases like periodontitis and squamous cells carcinoma of the oral cavity [110] have been discussed. Oral microbiome compositional variations were also associated with cardiovascular diseases [111,112], pancreatic [113] and colorectal cancer [114], many of which were further linked to the immune system's ability to control bacterial overgrowth and inflammation [115,116].

Causal links between several diseases and the oral microbiome were found as well. It is established that tooth plaque leads to higher risks of dental caries and periodontitis, respectively attributed mostly to *Streptococcus mutans* [117] and the red complex triad (*Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*) [118]. In recent years, the etiology of these diseases was re-evaluated under the perspective of the microbiome-wide dysbiosis hypothesis, which suggests that diseases may be mostly a consequence of whole-community shifts towards a harmful composition [119,120], although a consensus on a definition is yet to be reached [121].

Given its role in complex diseases, it was hypothesized that, similarly to the gut microbiome, the oral microbiome could potentially be modulated to favor disease remission, reduce inflammation (e.g., periodontitis [122]), or simply improve digestibility of some food molecules (e.g., gluten) [123]. In one study, 30 participants who ate cheese containing *Lactobacillus casei* showed reduced relative abundance of *S. mutans* than participants eating a control cheese [124], opening to the hypothesis that *L. casei* could behave like an oral probiotic against dental caries. Subsequently, *L. reuteri* [125] and *B. animalis* [126] were shown to improve periodontal phenotypes after probiotic intervention. However, like many other probiotic studies, a meta-analysis confirmed only part of these findings, mostly because of inter-study variability of the setup and the intrinsic limited efficacy of probiotic treatments [127]. Following the promising developments in FMT, efforts have been made to explore the feasibility of oral microbiome transplantation as well, mostly through saliva mouthwash. As a proof of concept, a study reported salivary microbiome transplantation from healthy mouse donors to mice undergoing radiotherapy. In this study, mice transplanted with a donor's salivary microbiome developed fewer signs of oral mucositis than non-transplanted controls [128].

Despite the large body of evidence of the role and genesis of the adult oral microbiome, a high proportion of variability needs to be explained. To begin with, understanding the factors determining presence and abundance of microbial species will aid research towards microbiome modulation strategies to improve health. In this context, the salivary microbiome is of particular interest for its ease of collection in a clinical context and its presence on all surfaces of the oral cavity.

2.3 The Cooperative Health Research in South Tyrol study is a rich resource for salivary microbiome

The CHRIS study was designed to explore the molecular and genetic basis of common chronic diseases and their interplay with lifestyle and environmental factors. The aim of the CHRIS study is to provide novel strategies for the prevention, diagnosis, and therapy of chronic diseases [129], in particular cardiovascular, metabolic and neurological diseases, which are among the leading causes of death in high-income countries [130–132]. Between 2011 and 2018 more than 13,000 adults were recruited from 13 municipalities of Vinschgau/Val Venosta, an area considered to have a homogeneous lifestyle and ancestry [129].

The study data collection protocol was outlined by Pattaro et al. in 2015 [129]. During the examination, each participants underwent anthropometric measurements, blood and urine collection, and questionnaire interviews. Additional questionnaires, for example the food frequency questionnaire, were self-administered. Below an overview of the macro areas of the data collection protocol is listed:

- **Personal Information**
- **Anthropometry** (height, weight, blood pressure)
- **Electrocardiogram at rest**
- **Barcode scan of medications**
- **Odor identification test** (Sniffin' Sticks)
- **Pain sensitivity** (Algometer)
- **Self-Administered Questionnaires** (Physical activity, Sleep quality, Mental health, Food Frequency)
- **Interview with CHRIS study nurse** (Occupation, Environmental exposure, Alcohol, Smoking, Vaccinations, Diagnosed chronic diseases)
- **Biochemical data** (Bloodwork, Urine, Metabolomics)
- **Genotype data** (DNA extracted from blood)

In addition to the CHRIS study data collection protocol, between January 2017 and February 2018 participants were requested to complete an oral health questionnaire and provide saliva samples that were used to generate 16S microbiome data. This protocol, named CHRISMB, was designed to increase the knowledge about factors associated with the adult salivary microbiome composition. With 1,923 participants, CHRISMB is, to our knowledge, the second largest salivary microbiome study with both salivary microbiome and host genetic data worldwide, after the 4D-SZ Shenzen cohort [133].

2.4 HOST GENETIC FACTORS ASSOCIATED WITH GUT AND ORAL MICROBIOME COMPOSITION

Studies on large cohorts reported associations between host's genotypes and gut microbiome composition. Genus *Bifidobacterium* was associated with the lactase expression in adults on the LCT genome locus: rs4988235 [134,135] and rs182549, the latter being the strongest signal found also in a 16S gut microbiome meta-analysis of more than 18,000 individuals [136]. When an individual is homozygous for the G variant, lactase is not expressed in adulthood. This results in higher amounts of undigested lactose in the gastrointestinal tract [137], which can serve as a sugar source for bacteria in the genus Bifidobacterium, a clade that metabolizes milk sugars [138].

The oral microbiome was also associated with host genetic variants. Using the 4D-SZ cohort, Liu and colleagues reported several tongue dorsum microbiome-genetic variant associations using 1600 highly prevalent (> 90%) Species-Genome Bins (SGBs). They identified several variants associated with tongue dorsum taxa: rs1196764 in the gene APPL2 was associated with *Prevotella jejuni*, *Oribacterium uSGB 3339* and *Solobacterium uSGB 315*; rs3775944 in gene SLC2A9 with *Oribacterium uSGB 1215*, Oribacterium uSGB 489 and *Lachnoanaerobaculum umeaense*; rs4911713 near the olfactory receptor gene OR11H1 with F0422 *uSGB 392*; rs36186689 in locus LOC105371703 with *Eggerthia*. Their cohort structure allowed them to evaluate the extent of validation of those signals between tongue dorsum and saliva, as well as between the same sample type but in two cohort batches. They concluded that at least 84% of their top signals could be replicated within the body site, but only 6 were shared between saliva and tongue dorsum [139].

2.5 GENOME WIDE ASSOCIATION STUDIES AND HERITABILITY ANALYSIS ARE TOOLS TO EXPLORE TRAITS' GENETIC COMPONENT IN A POPULATION

The links between host genetics and the gut and oral microbiome were conducted with population-wide statistical genetic approaches with the help of methods like heritability analysis genome wide association study (GWAS). The following paragraphs will describe the principles and the success of these methods.

In terms of base pair similarity, more than 99.9% of genomes are identical across all *Homo* sapiens individuals [140]. The remaining 0.1% is enough to determine a wide variety of traits (e.g., eye color, ABO and Rhesus blood type [141], hair and skin color [142], baldness [143])

and diseases (e.g., sickle cell anemia [144], cystic fibrosis [145], Huntington's disease [146]). Genetics play a role in quantitative traits as well: height is known to be highly influenced by genetics [147], but not entirely. Genes and their surrounding genomic regions (herein loci) contributing to a quantitative trait are called quantitative trait loci (QTL) [148]. Studying the overall effect of genetic and familial inheritance components on a trait helps disentangle the factors associated with its variability in a population.

2.5.1 Heritability Analysis

Heritability analysis was first developed by Ronald Fisher in 1918 [149] and it consisted of partitioning a trait's variance into environmental and genetic components [150], the latter inferred from family pedigrees [149]. He suggested that the total variance of a trait, V_T, was the sum of variance attributed to genetics/relatedness, V_G, and the environmental variance, V_E, also called the residual variance, as follows:

$$V_T = V_G + V_E$$

If a trait is strongly associated with genetic determinants, the V_G parameter is higher. Heritability, as later called by Lush [151] was defined as the proportion of variance ascribed to genetics, and it can range between 0 (no role of genetics) and 1 (only genetic) as follows:

$$0 < h^2 = \frac{V_G}{V_T} < 1$$

With this framework, high heritability indicates that the majority of the observed variation in a trait is due to genetics [150]. Importantly, heritability estimation depends on the population tested, since a variation in the genetic background, age, lifestyle and environmental exposure alters h^2 estimates, as previously reviewed [150]. More recently, computational implementations, namely SOLAR [152] and GCTA-GREML [153] were developed to perform these calculations.

2.5.2 Genome Wide Association Study

Visscher et al. showed that high heritability does not necessarily correlate with strong genetic signals for a trait [150,154]. To study the association between genetics and a trait, GWAS are a more suitable, complementary approach [155]. This technique tests all measured variants for association with a trait using linear models while allowing for confounders that may explain part of the trait variability [156]. The human genome is composed of slightly more than 3×10^9 base pairs [157], with an estimated number of almost 20,000 protein coding genes and even more transcripts [158]. To model such highly dimensional data, complex coding and

heuristics strategies have been implemented into popular tools like PLINK [159], GCTA [153] and GEMMA [160].

GWAS tools output summary statistics [161] with chromosome number, position, variant identifier, minor allele frequency (MAF), the strength of the association (β) and the P-value for each variant. Given the likelihood of false discoveries in millions of statistical tests in a GWAS, researchers have proposed multiple methods to account for Type I error rates without being too stringent with Bonferroni correction [162]. The Haplotype Map consortium mapped more than 1 million common variants independent from each other [163,164]. Therefore, it was suggested to correct the significance threshold by the number of genome-wide independent variants to $\alpha = \frac{0.05}{1\times10^6} = 5 \times 10^{-8}$. Since then, this threshold has been widely accepted [164,165].

2.5.3 Genotype imputation is a cost efficient method to gain detection power in a GWAS

The chances of novel discoveries rely on the number of variants tested. Nowadays, the gold-standard method to obtain the most dense and accurate set of variants of a population is whole genome sequencing (WGS) [166]. Despite the low cost of sequencing nowadays, it remains rather expensive to sequence thousands of samples [167]. Chip genotyping, in practice, enables typing of a large and customizable set of variants at a fraction of the cost of WGS. To further increase the cost/benefit ratio of chip genotyping, an *in silico* method was developed to expand the number of variants based on a smaller set of hard-called variants: genotype imputation [168]. In brief, imputation fills the gap between variants in an individual's genotype based on variants that are normally inherited together, named haplotypes [169]. Variants that are generally found in the same haplotype but are not present are "imputed" into the original haplotype of the sample (**Figure 2-2**) [168]. This procedure increases power by enlarging the set of variants tested. For example, a meta-analysis used genotype imputation to unveil novel genetic loci associated with colorectal cancer [170]. Chip-typed Genetic variants (hard calls) in the CHRIS study, including CHRISMB, were imputed using the publicly available Haplotype Reference Consortium (HRC) and the Trans-Omics for Precision Medicine (TOPMed) panels, which expanded the number of testable variants from 579,112 to 19,749,560 and 35,061,390 respectively. Additionally, whole exome sequencing data were generated for 3,294 individuals [171], resulting in 1,034,420 variants, 643,001 imputed exclusively using whole exomes as imputation panel.

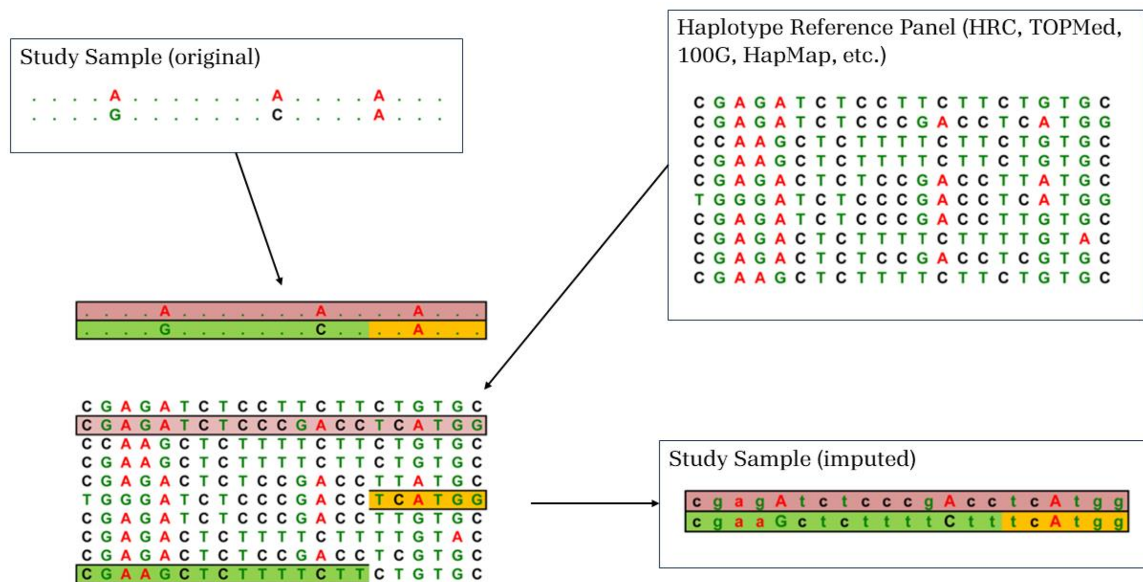


Figure 2-2. Example of a simple imputation process. A study sample is matched against a haplotype reference panel. For each allele, the haplotype with the most variants in the same position is chosen as the most likely haplotype. The variants in the reference haplotype that are missing in the sample are then added to the sample and labelled as “imputed” (see lowercase bases, bottom right). Figure adapted from Figure 1 in Li et al. (2009) [168]

2.6 THESIS OUTLINE AND SCIENTIFIC CONTRIBUTION

This section outlines the research chapters of the thesis, in which I adopt the term "microbiota" instead of "microbiome" for precision: "Microbiota" focuses on sample composition ("who is there?") and "microbiome" encompasses the genetic pool of the "microbiota" ("what can they do?") [172] . Since only 16S compositional data was used in this thesis, the term "microbiota" is adopted in research-oriented chapters, while "microbiome" is utilized in sections speculating on aspects beyond composition.

Chapter 3 – Smoking and salivary microbiota: a cross-sectional analysis of an Italian alpine population

This chapter explores the association of the salivary microbiome composition, predicted metabolic profiles and oxygen metabolism with smoking status (Never, Former, Current), smoking dose per day and years since quitting. Our study adds to the current knowledge of salivary microbiome dynamics in relation to smoking in a European population, with particular emphasis on the mediating role that the microbiome could have in the higher periodontal and cardiovascular disease risk among smokers via, among others, the nitrate reduction pathway and periodontal connective tissue degradation. My contribution to this

project was performing metadata curation and exploration, designing the analyses, generating the microbiome predicted pathways, performing the statistical analyses, writing the manuscript and corresponding with the journal editor.

Chapter 4 – Geographic and genetic factors on the composition of the adult salivary microbiota of an Italian alpine population

This chapter explores the association between salivary microbiome compositional variability in relation to geographic distribution (household, municipality, and altitude of residence) and familial relationships (heritability analysis and microbiome GWAS) in an unprecedented single-population salivary microbiota study. Our study highlights that a large proportion of salivary microbiota is heritable, and that cohabitation fosters higher similarity. Additionally, we found a small compositional shift in relation to the municipality of residence. Our microbiota GWAS highlighted four study-wide significant variants, supporting a small but significant role of genetics. My contribution to this project was designing and performing statistical analyses, interpreting results, and writing the manuscript. Submission is planned in the first quarter of 2024. In that occasion, I planned to correspond with the editor.

Chapter 5 – Additional Contributions

As introduced in the Background section, CHRISMB is a rich resource with data ranging from self-reported questionnaires to biochemical characterization of blood and urine. My main work has focused on detailed smoking habits, host genetics and the CHRISMB Alpine geographic peculiarities. However, I conducted 2 relevant analyses not yet included in a publication. This chapter presents these analyses as follows:

Odor identification capabilities and salivary microbiota composition are not significantly associated in CHRISMB

In this analysis I explored the association between the salivary microbiota composition and odor identification performance assessed with 16 felt-tip pens imbued with odorous molecules. I found that the salivary microbiota did majorly vary in relation to smell identification performance, pointing to a weak association show strong enough associations with smell performance. However, at a finer scale, I found three differentially abundant ASVs in the hyposmic group, two less abundant and one more abundant, suggestive of a marginal statistical association between olfaction performance and the salivary microbiota.

Salivary microbiota exposome scan

Population cross-sectional studies are generally rich in phenotypic data, derived with objective techniques, professional assessments, or questionnaires. Performing hypothesis-driven analyses is a more refined but time-consuming and case-specific method. I therefore performed an exposome scanning to identify which phenotypes significantly associated with microbiota variability in the population. To do so, I performed a non-parametric ANOVA-like model [173] which estimates the proportion of variance attributed to each phenotype separately. The analysis showed that the strongest associations were related to variables related to smoking, sample processing batch, age, oral health, occupation, education, and a few food frequency questionnaire items related to sugar consumption. I did not see associations with variables related to bloodwork, disease questionnaires, neuropsychiatric questionnaires, nor environmental phenotypes. To the best of my knowledge, this is the first exposome scan performed on the salivary microbiota data in a large population. My findings could foster future research on the salivary microbiome, as well as serving as a panel of variables to consider in designing future experiments and statistical models in similar experimental designs.

3 Smoking and salivary microbiota: a cross-sectional analysis of an Italian alpine population

Giacomo Antonello (1,2,#), Freida Blostein (3), Deesha Bhaumik (3), Elyse Davis (3), Martin Gögele (1), Roberto Melotti (1), Peter Pramstaller (1), Cristian Pattaro (1), Nicola Segata (2), Betsy Foxman (3, #), Christian Fuchsberger (1,#)

(1) Institute for Biomedicine, Eurac Research - Affiliated Institute of the University of Lübeck, Bolzano, Italy
(2) Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy
(3) School of Public Health – Epidemiology, University of Michigan, Ann Arbor, MI
Corresponding

Abstract

The oral microbiota plays an important role in the exogenous nitrate reduction pathway and is associated with heart and periodontal disease and cigarette smoking. We describe smoking-related changes in oral microbiota composition and resulting potential metabolic pathway changes that may explain smoking-related changes in disease risk.

We analyzed health information and salivary microbiota composition among 1,601 Cooperative Health Research in South Tyrol (CHRIS) participants collected 2017 to 2018. Salivary microbiota taxa were assigned from amplicon sequences of the 16S-V4 rRNA and used to describe microbiota composition and predict metabolic pathways.

Aerobic taxa relative abundance decreased with daily smoking intensity and increased with years since cessation, as did inferred nitrate reduction. Former smokers tended to be more similar to Never smokers than to Current smokers, especially those who had quit for longer than 5 years.

Cigarette smoking has a consistent, generalizable association on oral microbiota composition and predicted metabolic pathways, some of which associate in a dose-dependent fashion. Smokers who quit for longer than 5 years tend to have salivary microbiota profiles comparable to never smokers.

HIGHLIGHTS

- Cigarette smoking has a consistent, generalizable association on oral microbiota composition associated with smoking. Several taxa showed a relative abundance proportional to the number of cigarettes smoked per day.
- The salivary microbiota of people who quit smoking longer than 5 years resembled Never smokers' profiles, consistent with reports of changes in disease risk following smoking cessation.
- Decreased microbial nitrate reduction pathway abundance in smokers may provide an additional explanation for the effect of smoking on cardiovascular and periodontal diseases risk, a hypothesis which should be tested in future studies.

3.1 Introduction

Smoking is a risk factor for several complex, chronic diseases including but not limited to respiratory diseases [174], periodontitis [175–177], oropharyngeal cancers [178,179] and cardiovascular diseases [180]. Recently, alterations to oral microbiota composition have been observed in cases of periodontitis [181–184], squamous cells carcinoma [185], cardiovascular diseases [186,187] and in cigarette smokers [188–192] (Supplementary File 1, Table 1). Therefore, it is possible that smoking related changes in the oral microbiota contribute to the etiology of one or more chronic health conditions. The oral microbiota performs several functions, including playing an important role in the exogenous nitrate reduction pathway and hence blood pressure regulation via nitric oxide (NO) [193–196]. Diets high in nitrate increase the presence of oral nitrate-reducing bacteria (NRB), the most prevalent of which are species in the *Neisseria*, *Prevotella* and *Actinomyces* genera [197]. when NRB are present, salivary nitrate reduction increases [196,198]. Whether tobacco consumption directly or indirectly alters the relative abundance of nitrate reducing bacteria remains to be explored; however, smoking was reported to inhibit uptake of blood-circulating nitrate into saliva [199].

The salivary microbiota composition varies by smoking habits. A 2016 meta-analysis of 1204 USA citizens from two national cohorts found that compared to former or never smokers, smokers had a decreased relative abundance of Proteobacteria, an increase of Actinobacteria and a lower proportion of aerobic taxa after adjustment for age and sex [188]. A 2019 study set in New York city confirmed and extended those findings showing that, in contrast to former or never smokers (N= 86), the salivary microbiota of smokers (N = 86) showed higher abundance of genera *Stomatobaculum*, *Megasphaera, Veillonella, Leptotrichia, Campylobacter* and *Treponema,* and lower abundance of *Neisseria*, *Lautropia*, *Haemophilus*, *Capnocytophaga* [189]. Studies conducted in Saudi Arabia [190], Asia [89,200] and Europe [192] reported comparable findings (Supplementary File 1, Figure 1). In a meta-analysis with 1,204 Americans, Wu and colleagues uniquely found that the relative abundance of classes Betaproteobacteria, Gammaproteobacteria and Flavobacteriia was inversely correlated with the number of cigarettes smoked daily and directly correlated with the years since quitting smoking [188]. While associations between smoking status and salivary microbial composition have been previously characterized in Americans, no study has described associations of the salivary microbiota composition and metabolic potential with daily smoking intensity or years since quitting in a European population.

This study adds to our understanding of the associations of the salivary microbiota taxonomic and predicted metabolic functional composition with smoking status, intensity (grams/day) and history (years since cessation) in a large, novel, homogeneous Italian cohort aged 18 to 91: the Cooperative Health Research in South Tyrol (CHRIS) [201] Microbiome study (CHRISMB). We hypothesized that we would observe results consistent with the literature and some novel insights attributable to the unique characteristics of CHRISMB and the large sample size. We additionally hypothesized that the nitrate reduction pathways could be less abundant in smokers, given the previous findings of decreases of taxa in the *Neisseria* and *Haemophilus* genera, which harbor several NRB species [202].

3.2 Results

3.2.1 Characteristics of study population in relation to smoking

After exclusions (see methods and Supplementary File 1, Tables 2 and 3 for details), CHRISMB consisted of 1601 individuals with an average age of 45 years (range 18 - 91) and had slightly more females (52.9%) than males. Most had 20 or more natural teeth (72.1%). Almost half (45%) were Current or Former smokers; cigarettes were the primary source of tobacco for all but 5 participants. Smokers were more frequently males and younger than Never or Former smokers (**Table *3-1***). Former smokers quit smoking 17.96 years, on average (Range 0 - 61; median 16). When stratified by age group, Current and Former smokers aged 41 to 60 years with higher lifetime exposure to smoke tended to have fewer teeth than smokers with a lower cumulative exposure (Supplementary File 1, Figure 2).

Salivary microbiota DNA sequencing of selected samples consisted of almost 36 million reads, with a median read count per sample of 22,308 (interquartile range: 11,884, full range 5283 - 65,837). After filtering by prevalence and minimum detection (see Methods), the dataset included 627 ASVs assigned to 82 genera (Supplementary File 1, Table 4).

Table 3-1. Distribution of selected demographic descriptors in relation to smoking status in the Cooperative Health Research in South Tyrol Microbiome (CHRISMB) study. Per-column percentages were also reported in brackets. The whole cohort is included under the "CHRISMB" column. Significance was calculated as X^2 test for categorical variables. Non-available measures were reported as "Missing".

	Never (N=880)	Former (N=395)	Current (N=326)	CHRISMB (N=1601)	X^2 P-value
Sex					2.7e-07
Male	356 (40.5%)	222 (56.2%)	173 (53.1%)	751 (46.9%)	
Female	524 (59.5%)	173 (43.8%)	153 (46.9%)	850 (53.1%)	
Age Category (years)					3.6e-19
18-30	238 (27.0%)	41 (10.4%)	130 (39.9%)	409 (25.5%)	
31-40	139 (15.8%)	73 (18.5%)	57 (17.5%)	269 (16.8%)	
41-50	196 (22.3%)	75 (19.0%)	64 (19.6%)	335 (20.9%)	
51-60	144 (16.4%)	112 (28.4%)	51 (15.6%)	307 (19.2%)	
61-70	93 (10.6%)	57 (14.4%)	23 (7.1%)	173 (10.8%)	
71+	70 (8.0%)	37 (9.4%)	1 (0.3%)	108 (6.7%)	
N° Teeth (self-reported)					0.07
0	50 (5.7%)	23 (5.8%)	16 (4.9%)	89 (5.6%)	
1-9	57 (6.5%)	41 (10.4%)	20 (6.1%)	118 (7.4%)	
10-19	117 (13.3%)	74 (18.7%)	48 (14.7%)	239 (14.9%)	
20+	656 (74.5%)	257 (65.1%)	242 (74.2%)	1155 (72.1%)	
Gums Health (self-reported)					0.87
Excellent	45 (5.1%)	18 (4.6%)	16 (4.9%)	79 (4.9%)	
Very good	188 (21.4%)	79 (20.0%)	64 (19.6%)	331 (20.7%)	
Good	291 (33.1%)	124 (31.4%)	87 (26.7%)	502 (31.4%)	
Average	229 (26.0%)	84 (21.3%)	99 (30.4%)	412 (25.7%)	
Poor	47 (5.3%)	24 (6.1%)	22 (6.7%)	93 (5.8%)	
Very poor	6 (0.7%)	2 (0.5%)	3 (0.9%)	11 (0.7%)	
Missing	74 (8.4%)	64 (16.2%)	35 (10.7%)	173 (10.8%)	

3.2.2 Qualitative smoking habits are associated with compositional and functional profiles of salivary genera

The microbiota composition of CHRISMB at phylum level was dominated by Firmicutes, followed by Bacteroidetes Proteobacteria, Fusobacteria and Actinomycetes; while at Genus level it was dominated by *Prevotella, Streptococcus, Veillonella, Haemophilus, Neisseria* (Supplementary File 1, Figure 3). The salivary microbiota was significantly associated with smoking (**Figure 3-1** A, PERMANOVA $R^2 = 0.04$, $p = 0.001$, 2,000 permutations) as well as sex, age group and number of teeth, considering the marginal effect of all variables together (Supplementary File 1, Table 5). Alpha diversity was not significantly associated with smoking status (Supplementary File 1, Figure 4). Principal coordinate analysis and differential abundance analysis together suggested that the salivary microbiota of Former smokers was highly similar to Never smokers. Consensus-based differential abundance analysis identified 44 genera that were significantly different between Current smokers and Never smokers after adjusting for age, sex, and number of teeth (**Figure 3-1** B). To investigate sex-dependent associations, we repeated the same consensus differential abundance analysis separately by sex, again adjusting for age and number of teeth. Despite finding sex-specific differentially abundant genera, all were in the set of 44 differentially abundant genera of the model adjusted for sex, age group, and number of teeth (Supplementary File 1, Figure 5). We annotated genera based on their oxygen requirements from a manually curated table by Calgaro et al. [203], and observed that the relative abundance of aerobic taxa decreased consistently in smokers (from a median of 7% to 3%), in favor of anaerobes (**Figure 3-1** C).

0

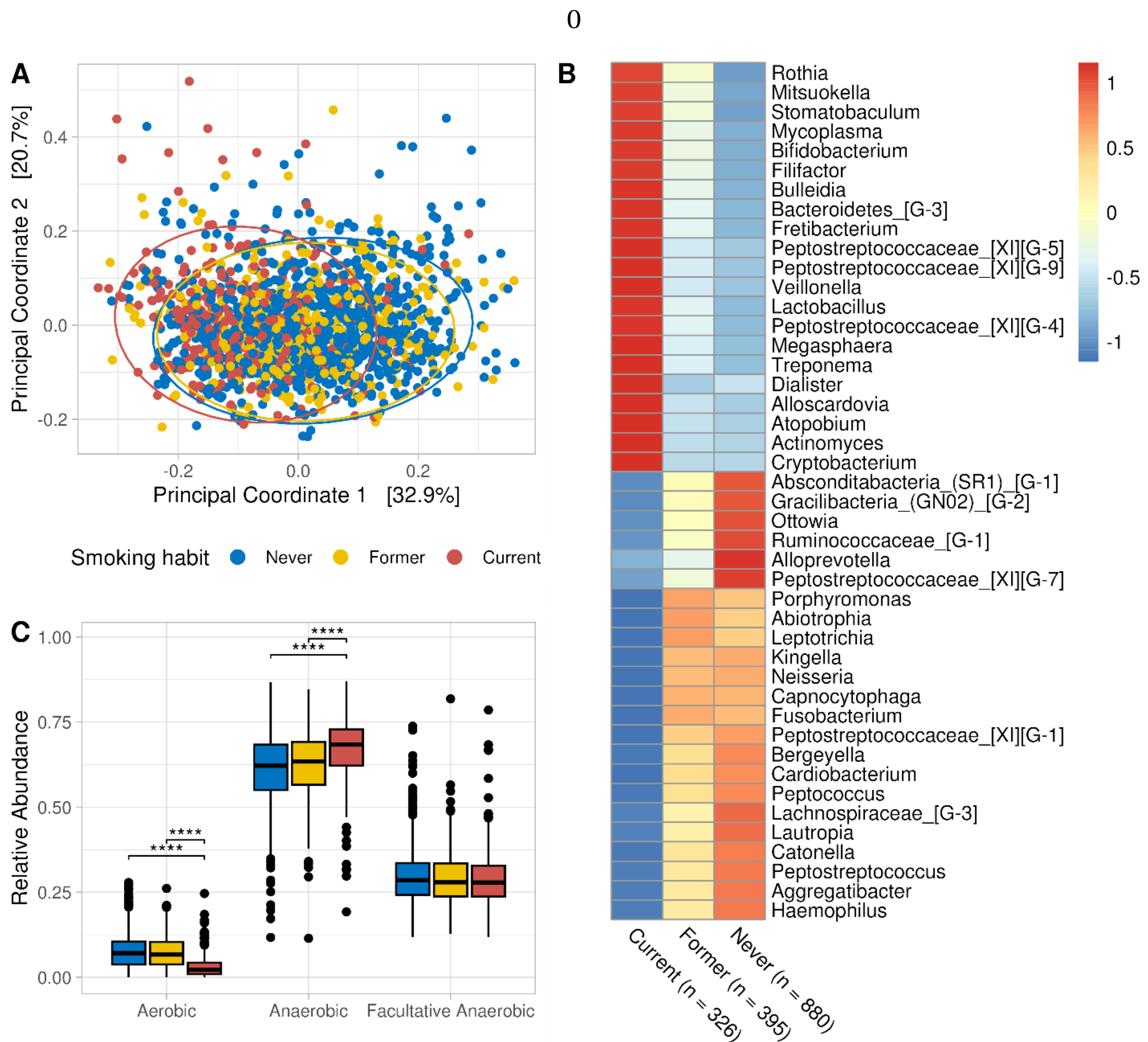


Figure 3-1. Association between qualitative smoking habits (Never, Former and Current) and the salivary microbiota in the CHRISMB cohort. (A) Principal Coordinate Analysis on the Bray-Curtis dissimilarity at genus level. confidence areas (95%) were drawn as ellipses. Group separations were mild but significant (PERMANOVA R^2 = 0.04, p = 0.001, beta-dispersity p = 0.104). Axes x and y were chosen as the principal components which explained most of the overall microbiota variability, which is shown in square brackets. **(B)** Heatmap of the 44 genera differentially abundant between Current and Never smokers. Each genus was transformed to relative abundance and Z-score scaled. Red and blue colors indicate a higher and lower mean abundance, respectively, while yellow colors indicate no difference. Genera reported in the figure were differentially abundant (Benjamini-Hochberg Q-value < 0.05, False Discovery Rate (FDR) = 5%, ALDEx2 Holm Q-value < 0.05) in at least 4 out of 5 differential abundance methods (DESeq2, LinDA, MaAsLin2, ALDEx2, ANCOM-BC), adjusting for age (categorical), sex (binary) and number of teeth (categorical). **(C)** Relative abundance of aerobes, anaerobes, and facultative anaerobes in relation to smoking status. Statistical significance was calculated with pairwise Wilcoxon test adjusting P-values (Q-values) for a 5% FDR with the Benjamini-Hochberg method (** q < 0.05; *** q < 0.001, **** q < 0.0001).

3.2.3 Several microbial genera associated with smoking habits are also associated with the grams of tobacco smoked daily

We regressed each genus against daily smoking intensity as multiples of 5 grams per day (see Methods). *Fretibacterium* was positively associated with increases in daily smoking intensity and 10 with decreases (**Figure *3-2*** A). Except for *Campylobacter* and *Selenomonas*, the remaining 9 genera were also differentially abundant comparing Current against Never smokers (Figure 1B). Additionally, the effect sizes estimated in the daily smoking intensity regression were highly correlated with the estimates obtained comparing Current against Never smokers (Pearson ρ = 0.87, Supplementary File 1, Figure 6), suggesting that some genera associated with smoking against non-smoking were additionally associated with daily smoking intensity. The complete linkage hierarchical clustering in the *pheatmap* function tended to cluster heavier smokers together, further suggesting a dose effect (**Figure *3-2*** A). The mean relative abundance and variance of aerobes significantly decreased at the increasing daily smoking intensity (linear regression $\beta(\frac{1}{grams/day}) = 0.027$, P-value = 4.6×10^{-4}; Supplementary File 1, Tables 6, 7, adjusted for age as continuous variable, sex and number of teeth) with a plateau at more than 10 grams (**Figure 3-2** B). Conversely, the relative abundance of anaerobes and facultative anaerobes slightly increased.

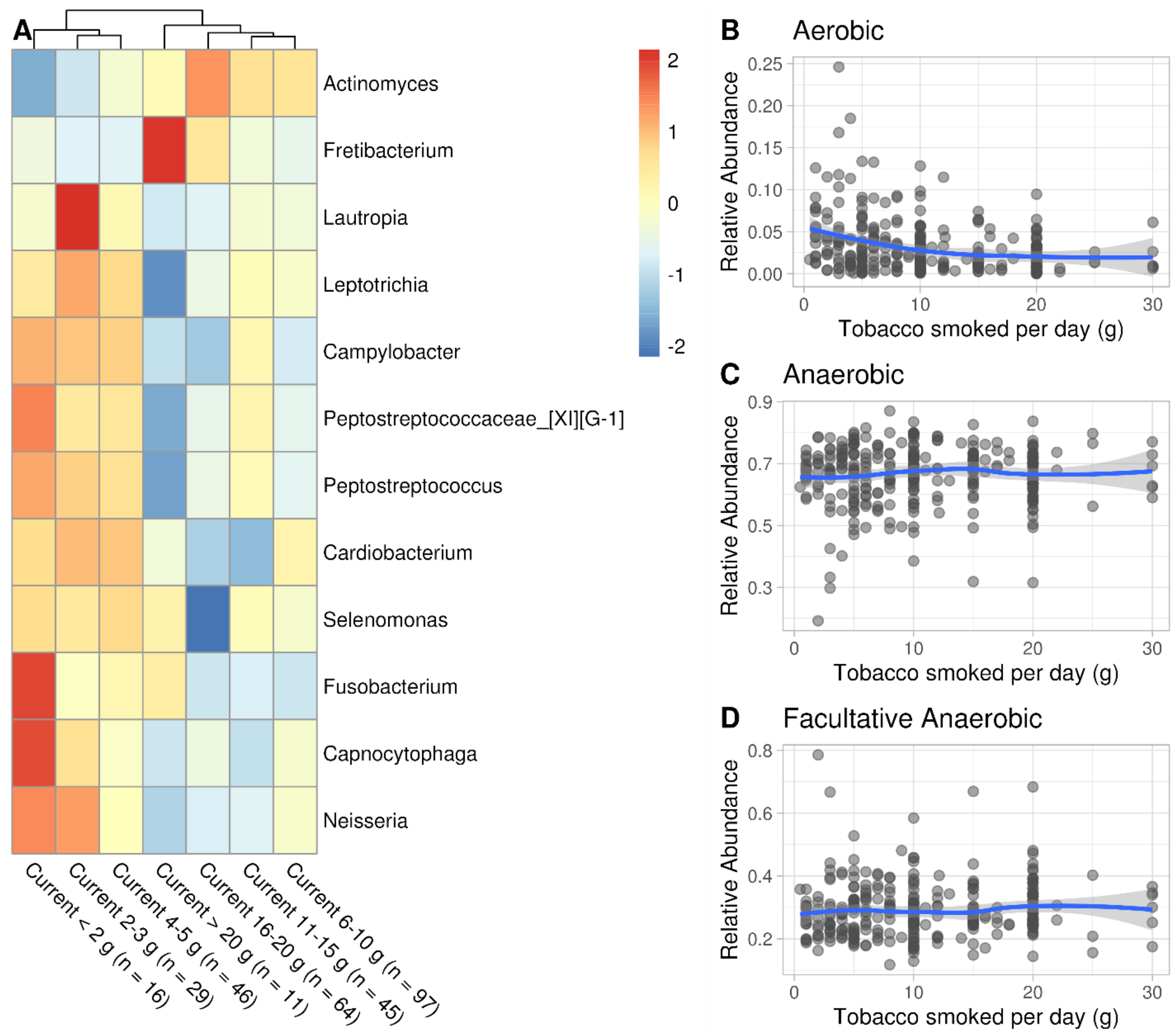


Figure 3-2. Smokers' (n = 308) daily smoking intensity is associated with relative abundance shifts of several genera and a decrease of aerobic taxa relative abundance. (A) Heatmap of genera significantly affected by daily smoking intensity. Genera were transformed to relative abundance and Z-score scaled to highlight relative differences in mean abundance in relation to the smoking intensity. Significant genera (Benjamini-Hochberg Q-value < 0.05, FDR = 5%) were obtained modeling each genus in response to daily smoking intensity as multiples of 5 grams per day as a semi-continuous variable, adjusting for age (continuous), sex and number of teeth in the DESeq2 negative binomial generalized linear model framework. **(B, C, D)** Relative abundance of aerobes, anaerobes and facultative anaerobes, respectively, in relation to the grams of tobacco smoked daily.

3.2.4 Salivary microbiota of Former smokers who quit 5 years or longer tended to resemble Never smokers' profiles

We studied the association between salivary genera of former smokers and the years since smoking cessation using the same model framework as the daily intensity regression (**Figure *3-2***), with 1 year scale, finding no statistically significant association. We visualized the mean relative abundance of genera associated with smoking (**Figure 3-1** B) in the Former smokers' group with 20 or more natural teeth, grouping them by bins of years since quitting. We limited the visualization to individuals with 20 or more teeth to minimize the effect of tooth loss on the microbiota of Former smokers, who tended to be older than Current and Never smokers (**Figure 3-3** A). Looking at the complete linkage hierarchical clustering, we noticed a gradual increase of similarity of Never smokers to Former smokers who quit for more years, except for the "Former 2-3 y" group (**Figure 3-3** A). The relative abundance of aerobes mildly increased in the first 20 years since quitting ($\beta_{0 \leq years \leq 20} = 0.001$, P-value 0.052, adjusted for age, sex and number of teeth; Supplementary File 1, Tables 8, 9) (**Figure 3-3** B).

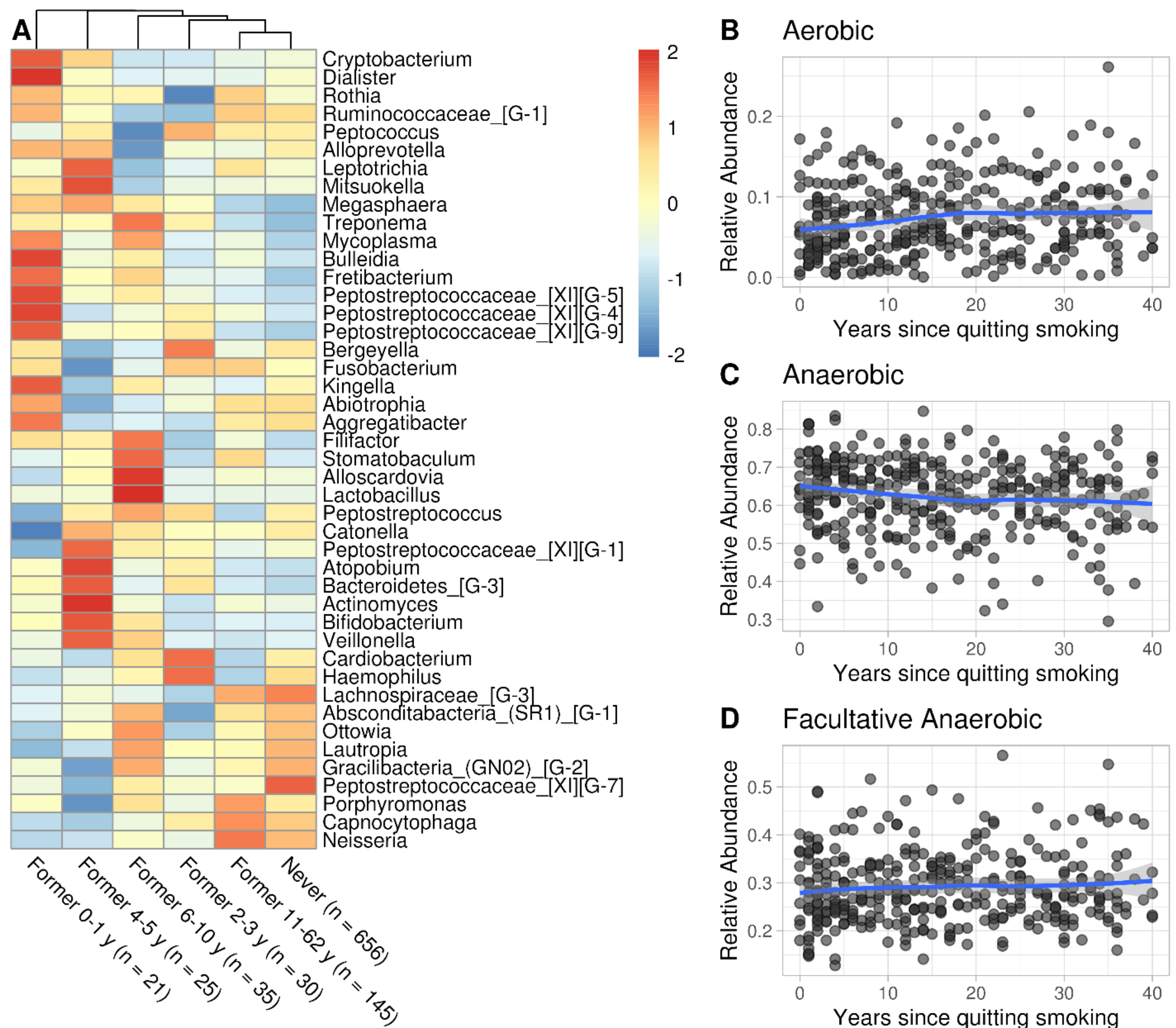


Figure 3-3. The salivary microbiota of individuals who quit smoking (n = 369) showed multiple-year perturbation and tends to resemble Never smokers' profiles within 5 years. (A) Heatmap of the relationship between the years since quitting smoking and the mean relative abundance of genera previously found significantly associated with smoking (see Figure 1). Taxa were transformed to relative abundance and scaled by row, to highlight differences in mean abundance in relation to bins of years since quitting to limit the low sample size of some categories. Complete linkage hierarchical clustering was used to cluster columns. Since Former smokers tend to be older and given the tendency of the elderly to lose teeth, we limited the visualization to people with 20 or more teeth**. (B, C, D)** Relative abundance of anaerobes, aerobes, and facultative anaerobes in relation to years since quitting smoking.

3.2.5 Predicted Functional Profiles Associated with Smoking highlighted a decrease of aerobic and nitrate reducing taxa

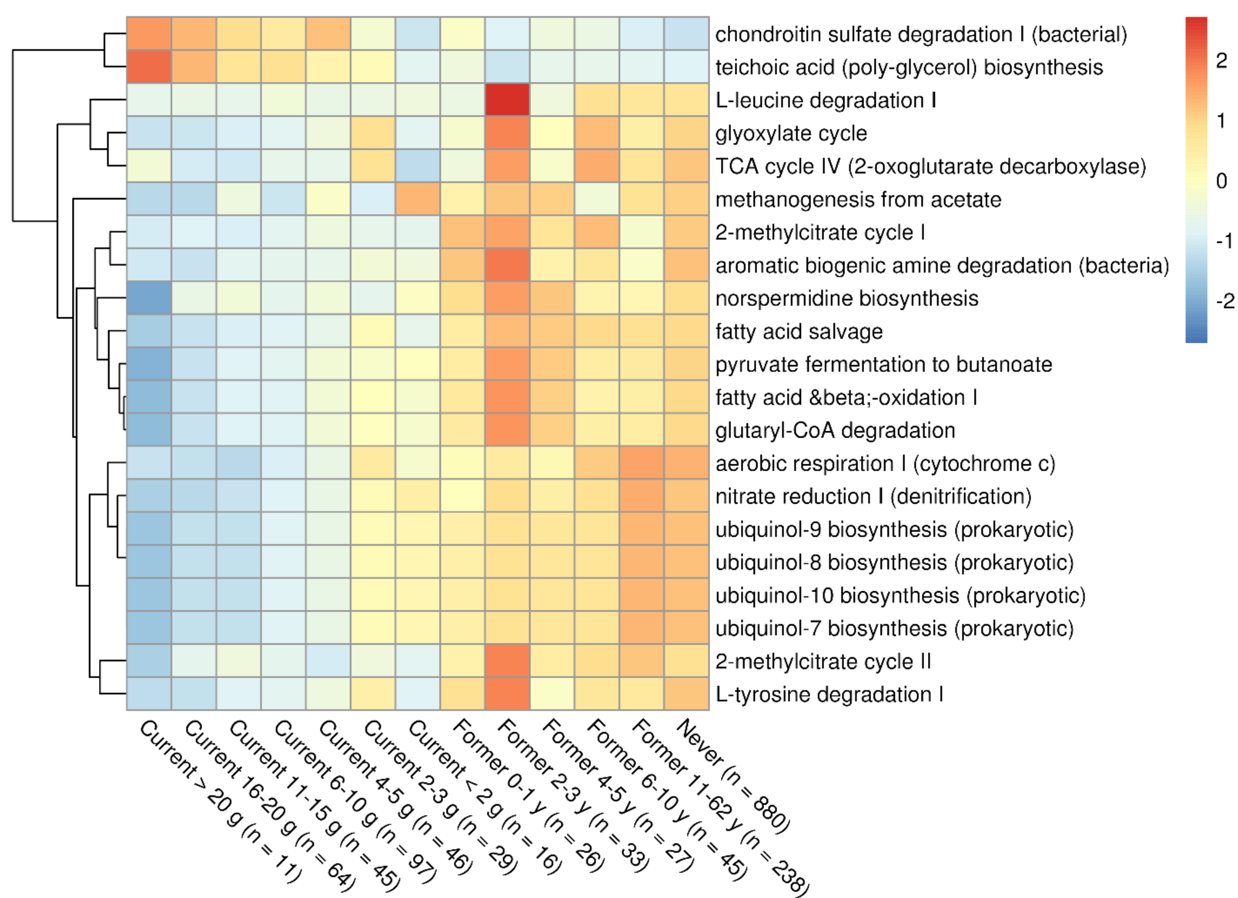


Figure 3-4. Microbial metabolic pathways inferred with PICRUSt2 that were differentially abundant in relation to smoking exposure, adjusting for age, sex and number of teeth. Heatmap of the 21 differentially abundant pathways in Current against Never smokers contrasts. Each pathway was transformed to relative abundance and Z-score scaled. Groups were ordered based on decreasing exposure to smoking, from heavier smokers to Former smokers who quit for the most years. As a reference for absence of exposure to smoking, Never smokers were included in the rightmost column. Red and blue colors indicate a higher and lower mean abundance, respectively, while yellow colors indicate no difference. Differential abundance analysis was performed with a consensus-based approach of 5 differential abundance methods (DESeq2, LinDA, MaAsLin2, ALDEx2, ANCOM-BC), modeling each pathway against smoking status and adjusting for age (categorical), sex (binary) and number of teeth (categorical). Pathways reported in the figure were differentially abundant (Benjamini-Hochberg Q-value < 0.05, False Discovery Rate = 5%, ALDEx2 Holm Q-value < 0.05) in at least 4 methods with an absolute effect size larger than 0.5.

After predicting microbial pathway abundance with PICRUSt2, we identified pathways that were differentially abundant between Current and Never smokers using the same consensus method used for genus-level taxonomy. We identified 21 pathways, which we later visualized in relation to a gradient of smoking exposure, without clustering (**Figure 3-4**). It should be noted that some of these were reconstructed from the same sets of predicted enzymes, therefore their correlation was 1 (e.g., Ubiquinol pathways). To avoid selection bias, we performed the analysis on all pathways regardless of their correlation and reported the correlation matrix of the significant ones in Supplementary File 1, Figure 7.

3.3 Discussion

3.3.1 Summary of study and main results

We investigated the associations between salivary microbial genera and predicted metabolic pathways and smoking status, daily smoking intensity and years since cessation in CHRISMB, a convenience sample of 1601 adult participants in the CHRIS study in South Tyrol, Italy [201]. We confirmed previous findings regarding salivary microbiota compositional differences by smoking behavior. Additionally, we demonstrated that aerobic taxa varied with the frequency and intensity of smoking exposure, and that the salivary microbiota of Former smokers is generally more similar to the salivary microbiota of Never smokers, especially of those who quit longer than 5 years. Several aerobic or oxygen-requiring predicted microbial pathways decreased in smokers. The nitrate reduction pathway was significantly lower in smokers than in non-smokers. The decreases in nitrate reduction pathways among current smokers and increases in these pathways among former smokers is consistent with previous reports of decreases in cardiovascular events among former smokers [204]. This suggests that oral microbiota functional changes with smoking may be an additional explanation for changes in cardiovascular risk with changes in smoking habits.

3.3.2 Comparison with other studies

The relative abundance of salivary microbiota phyla of CHRISMB participants was comparable with the mean composition of a Japanese [89] and Middle Eastern [190]: Firmicutes were the most abundant, followed by Bacteroidetes, Proteobacteria, Actinobacteria and Fusobacteria. Consistent with previous analyses in Americans of Caucasian, African and Hispanic ancestry [188,189], and cohorts of middle [190] and eastern Asian Ancestry [89,191], Italian smokers had decreased abundance of *Neisseria*, *Lautropia*, *Haemophilus*, *Capnocytophaga,* and increased abundance of *Atopobium*, *Megasphaera,*

and *Veillonella* when compared to Never smokers (**Figure 3-1** B). This suggests that cigarette smoking has a consistent and generalizable effect on the oral microbiota. We also identified 12 novel differentially abundant genera between Current and Never smokers: *Alloscardovia,* Bacteroidetes *Genus 3, Bulleidia*, *Cryptobacterium*, *Fretibacterium, Mitsuokella, Parvimonas,* Peptostreptococcaceae XI *Genus 9* and *Stomatobaculum* were increased, while Absconditabacteria (SR1) *Genus 1, Ottowia* and *Peptidiphaga* were decreased (Supplementary File 1, Figure 1). Further work is required to determine whether these changes are specific to this work.

3.3.3 Salivary microbial genera composition and proportion of aerobes were strongly impacted by smoking

Of the 44 differentially abundant genera in smokers, compared to Never smokers, genera in the phylum Proteobacteria (N=7) were decreased and Actinobacteria (N=6) were increased among smokers. These two phyla harbor mostly aerobic and anaerobic taxa, respectively. Indeed, the proportion of aerobes was inversely proportional to the frequency and intensity of exposure to smoking (**Figure 3-1**, **Figure 3-2**, **Figure 3-3**). We also predicted functional profiles based on our compositional data, observing an increase of Gram-positive associated pathways in smokers, in particular teichoic acid biosynthesis (**Figure 3-4**), which we confirmed looking at the relative abundances of Gram staining of bacteria across smoking status (Supplementary File 1, Figure 8). Moreover, we observed a decrease in pathways associated with aerobes, such as nitrate reduction and ubiquinol synthesis, which is pivotal in the electron transport chain [205], and a decrease of pathways that require oxygen and/or produce an excess reducing power, such as fatty acid oxidation. These findings support the hypothesis that smoking induces a hypoxic environment in the oral cavity. A decreased abundance of the nitrate reduction pathway in smokers could be an effect of the decrease of genera *Neisseria, Haemophilus, Kingella*, which harbor several NRB. A decrease of NRB may have a detrimental effect on enterosalivary nitrate reduction [206], which is a considerable source of blood nitrites for endogenous NO synthesis. Decreases in NO, which is a vasodilator [207], might hinder gingival blood flow and increase stress over time, which could lead to higher chances of gingival recession and periodontal diseases [208]. Indeed, chondroitin sulfate degradation was increase in heavier smokers, which may be indicative of higher stress to the gingival connective tissue and increase the risk of periodontal diseases. NO deficiency has also been suggested as a risk factor for developing cardiovascular diseases [209–211]. Taken together, microbiota-derived NO depletion may increase the chance of developing periodontal and cardiovascular diseases in smokers, as recently reviewed [212].

3.3.4 Some genera are statistically associated with daily smoking intensity but not with the years since smoking cessation

In addition to examining quantitative differences by Current smoking status, we tested for differences in bacterial composition by daily intensity of tobacco exposure (g/day) (**Figure 3-2**). Extending observations by Wu et al. [188] at lower taxonomic level and higher resolution of exposure variables, genera belonging to classes Betaproteobacteria (*Lautropia, Neisseria)*, Gammaproteobacteria (*Cardiobacterium*) and Flavobacteriia (*Capnocytophaga*) were significantly decreased at increasing grams of tobacco smoked per day. Additionally, we found negative correlation with grams of tobacco smoked per day for genera in classes Clostridia (Peptostreptococcaceae Family XI - *Genus 1, Peptostreptococcus*), Epsilonproteobacteria (*Campylobacter*), Fusobacteriia (*Fusobacterium, Leptotrichia*) and Negativicutes (*Selenomonas*). Genera *Actinomyces* (class Actinobacteria) and *Fretibacterium* class Sinergistia) were significantly increased. The subsiding of smoking-related microbial taxa was in line with the observation of full recovery of cardiovascular health risk within 5 years since quitting [213]. It is possible that smoking induced oral microbiota alterations may last longer than 5 years (**Figure 3-3** B, C, D), which would align with the subsiding of periodontal disease risks in smokers within 10 years [214].

3.3.5 Study Limitations

Major limitations of this study include the cross-sectional design and the lack of a professional assessment of the number of decayed, missing and filled teeth and gum health. While we controlled for age, sex, and number of teeth as potential confounders in our models, residual confounding is still possible due to, for instance, medications usage, diet, and alcohol intake. Furthermore, some subgroup strata were small, and structural non-positivity could exist. Bacterial metabolic pathways inference was based solely on salivary microbiota composition. While it is encouraging that our results regarding changes in salivary microbiota composition with smoking habits are consistent with those of previous studies conducted among very different populations, prospective studies are required to more directly address whether oral microbiota play a mediating role in the onset of smoking-related chronic diseases.

3.3.6 Study Strengths

Our analysis also has several strengths. The smoking questionnaire was detailed, allowing for high-resolution qualitative and quantitative characterization of smoking habits. We tested for a dose-dependent relationship between smoking and perturbation to the oral microbiota, supporting a causal relationship according to the Bradford Hill criteria [215].

This study cohort was particularly homogenous from the perspective of ethnicity, lifestyle and microbiota data generation, which should significantly limit confounding effects.

Our sample size was the largest to date to examine associations between smoking and the oral microbiota in a European population. While the salivary microbiota is a composite of multiple oral communities, saliva samples are easy to collect, making them ideal for large epidemiological cohorts and for future diagnostics and prognostics.

3.4 CONCLUSIONS

Smoking is associated with changed in the salivary microbiota composition often in a dose-dependent fashion. The salivary microbiota of people who quit smoking longer than 5 years resembled Never smokers' profiles. Irrespective of the phylogeny, aerobic taxa are the most sensitive to smoking exposure. Decreased microbial nitrate reduction pathway abundance in smokers may provide an additional explanation for the effect of smoking on cardiovascular and periodontal diseases risk, a hypothesis which should be tested in future studies.

3.5 ACKNOWLEDGEMENTS

We thank the CHRIS study coordinators and the personnel of the South Tyrol Healthcare Organization (SABES) for carrying out the data and sample collection in the CHRIS visit centre in Schlanders/Silandro. Special thanks to Deborah Mascalzoni and Massimiliano Pellegrini for providing the ethical and legal framework for CHRIS. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano for covering the Open Access publication costs.

3.6 FUNDING SOURCES

The CHRIS study was funded by the Department of Innovation, Research and University of the Autonomous Province of Bolzano-South Tyrol and supported by the European Regional Development Fund (FESR1157). The CHRIS biobank was assigned the “Bioresource Research Impact Factor” code BRIF6107. The CHRISMB microbiota data generation was funded by the National Institute of Dental and Craniofacial Research grant number R01 DE014899.

3.7 Data and Analysis Scripts Availability

CHRIS and CHRISMB data cannot be shared openly due to Italian laws on personal data protection. However, CHRIS data can be downloaded from https://chrisportal.eurac.edu/, after approval of the researcher's proposal by the CHRIS data access committee. Analysis scripts are freely accessible at https://github.com/g-antonello/CHRISMB-smoking-epidemiology.

3.8 Materials and Methods

3.8.1 Study ethical approval, design, and data collection

The CHRIS study was approved by the local Ethical Committee within the South Tyrol healthcare on April 19, 2011, and registered with code 21.2011. The legal base for personal data handling and protection was the informed consent explained to and signed by each participant. The personal data protection warrant of CHRIS constantly ensures that all data are handled and protected in full compliance with the European Regulation (EU 2016/679) and Italian law (D.L.vo 196/2003).

The CHRIS study includes adults of both sexes aged 18 and older. Participants were recruited starting in 2011 with extensive outreach including advertisements, electronic and paper mail to cover most people residing in the Vinschgau/Val Venosta district (South Tyrol, Italy). On the day of visit, participants answered lifestyle, dietary, general health, and socio-economic status questionnaires [201]. The CHRIS Salivary microbiota (CHRISMB) project is a convenience sample of CHRIS participants recruited between January 2017 and February 2018.

Epidemiological data generation. We defined age as the difference between the examination date and the birth date, rounded to the closest integer, and categorized age into six groups as shown in Table 1. CHRISMB participants filled in an adapted version of the World Health Organization oral health survey [216], from which we extracted information about the number of natural teeth in 4 ranges: 0, 1-9, 10-19 and 20 or more. We derived smoking variables from smoking questionnaires harmonized from the European Community Respiratory Health Survey III questionnaire [217]. We defined qualitative smoking habits - "Never", "Former", "Current with reduction" - Current (R), and "Current without reduction" - Current (NR) - according to Murgia and colleagues [218]. Former smokers were smokers who quit for longer than 1 month prior to the visit.

Current (R) were individuals who reported being smokers at the day of examination but that reduced the daily smoking intensity at least 1 month prior to the visit. Since we did not observe differences in the microbiota composition of Current (R) and Current (NR) (Supplementary File 1, Figure 9), we decided to aggregate the two smoking groups. For completeness, included in the supplement is a description of the study population showing the separate characteristics of the Current and Former smoker groups (Supplementary File 1, Table 10). Cigarettes were the primary source of tobacco, except for 5 participants. To include all sources of tobacco as one variable of smoking intensity, we converted the number of cigarettes, cigars, and cigarillos into grams of tobacco equivalents, respectively 1, 5 and 3 grams (g), and converted g/week to g/day as previously proposed [218,219]. We defined "smoking history" as the difference between the age of the participant to CHRIS and the reported age at which the participant quit smoking, rounded to the closest integer.

3.8.2 Salivary microbiota data generation

Saliva sample collection and storage. CHRISMB participants were required to drink only water and fast from the night before. Additionally, they were required not to drink, eat or smoke within 1 hour prior to the visit. During the visit, they provided 2-5 mL unstimulated saliva samples into Oragene OG-500 tubes. Within a few hours after the collection, samples were vortexed, split into 0.5 mL aliquots, and promptly stored at -80 °C.

DNA extraction and sequencing. Salivary DNA extraction and sequencing were conducted by the University of Michigan microbiome core. DNA was extracted using the Eppendorf epMotion liquid handling system and Qiagen MagAttract PowerMicrobiome Kit protocol and quantified with the PicoGreen dsDNA Assay kit (Thermo Fisher Quant-iT, cat# P7589).We amplified the V4 hypervariable region of the 16S rRNA gene by polymerase chain reaction (PCR) using a dual indexing strategy [220]. PCR products were visualized using E-Gel 96 with 2% SYBR Safe DNA Gel Stain (Life Technologies cat# G7208-02). PCR products were then pooled and normalized using SequalPrep Normalization Plate Kit (Life Technologies, cat# A10510-01) following the manufacturer's protocol for sequential elution.

The final pooled library consisted of equimolar amounts of each plate, normalized to the pooled plate at the lowest concentration. Sequencing libraries were prepared according to the Illumina MiSeq guidelines, adding phiX phage genome to ease diversity and quality control. Each of the 5 libraries contained 2 negative and 2 positive controls, respectively using water from the extraction step and commercially available DNA from communities of known

composition from the PCR step (Zymo Research, cat# D6306). We sequenced reads on an Illumina MiSeq machine.

3.8.3 Sequencing data processing

Sequencing data processing. We assessed the sequencing quality of the 69,286,448 obtained reads using "MultiQC" (v. 1.7) to visually determine read trimming length. We performed FASTQ read trimming, filtering, and taxonomic assignment with the "DADA2" package (v. 1.14) [221] in R (v. 3.6.0) [222]. This method generates a high-resolution sequence table of Amplicon Sequence Variants (ASVs), each differing by at least one nucleotide. We removed the first 20 and last 8 nucleotides to eliminate primer and barcode sequences and to ensure homogeneity of ASV calling across batches. After these steps, we submitted 59,331,563 reads to the *LearErrorRates* step, separately for each run, using $1x10^8$ bases as the learning rate parameter, which helps infer technical and real sequence differences. Then, we merged paired ends, resulting in 57,122,521 reads. Removal of chimeras using the consensus method resulted in an additional loss of 1.05% and 44,136,182 total reads used for taxonomic assignment. We assigned taxonomy from kingdom to genus level using the Bayesian classifier and the expanded Human Oral microbiome Database (eHOMD), while the species level was assigned using the 100% identity *addSpecies* strategy. To increase the likelihood of assignment at the species level, we enriched the eHOMD database with publicly available 16S rRNA FASTA sequences from known oral species in the genera *Lactobacillus*, *Streptococcus*, and *Prevotella* (Supplementary File 1, Table 11). We confirmed homogeneity across batches based on positive compositional profiles (Supplementary File 2).

Microbiota data preparation for analysis. We generated a phyloseq object starting from the counts table, taxonomic table and taxonomy tree using the Bioconductor package "phyloseq" (v. 1.42.0) [223] and "ape" (v. 5.7). We retained only those taxa that were present with at least 10 reads in at least 1% of samples with the function *core* of the "microbiome" package (v. 1.20.0) [224]. We aggregated ASVs at the genus level with the *tax_glom* function in the GitHub package "speedyseq" ("mikemc/speedyseq"), a faster version of phyloseq for microbiome data manipulation.

3.9 Samples availability and statistical analysis

Participants with missing data on smoking habits (N = 4), number of teeth (N = 44) and antibiotic usage within 3 months prior to the visit (N = 83) or who reported taking antibiotics within 3 months prior to saliva collection (N = 191) were excluded, leaving 1601 analytic samples. Additionally, we excluded 17 smokers from the "Regression of microbial genera against smoking intensity" due to missing or inconsistent grams of tobacco smoked per day and 1 participant who declared smoking 60 cigarettes per day, which was far beyond the range of the rest of the data (0.5-30). We further excluded 4 participants from the analysis "Regression of microbial genera against smoking history" due to inconsistent or missing answers.

3.9.1 Statistical analysis

Unless reported otherwise, we performed all statistical analyses using R (v. 4.2.2) and RStudio Server (v. 2022.07.2).

Pairwise relationship between demographics. We tested the independence of smoking habits from age groups, sex, self-reported gum health and self-reported number of natural teeth using a χ^2 test of independence with Yates's correction for low-frequency groups. We considered traits with a P-value lower than 0.05 as statistically non-independent.

Beta diversity and dimensionality reduction visualization. We estimated between-sample microbiota dissimilarity transforming genera counts to relative abundance and calculating the Bray-Curtis dissimilarity with the *distance* function in "phyloseq". We obtained eigenvectors with *ordinate* and visualized the two vectors explaining the most variance with *plot_ordination*, drawing 95% confidence interval ellipses with *stat_ellipse* in the "ggplot2" package (v. 3.4.0). We estimated the impact of smoking habits, number of teeth, sex and age group on the beta diversity using a permutational multivariate analysis of variance (PERMANOVA) [225] with *adonis2* in the "vegan" package (v. 2.6-4), with 2,000 permutations considering the marginal effect of all variables. We ensured even intraclass dispersion of smoking status groups (Never, Former and Current) using *betadisper* followed by *permutest*, with 2,000 permutations (Supplementary File 1, Table 5).

Differential abundance analysis in relation to smoking. To study the association between each oral genus abundance and smoking, we performed differential abundance analysis comparing Current with Never smokers adjusting for age group, sex and number of teeth. We performed a consensus based differential abundance analysis, as advised by Nearing et al.

[226], using 5 different methods having: DESeq2 (v. 1.38.2) [227], LinDA (v. 1.1) [228], MaAsLin2 (v. 1.12) [229], ALDEx2 (v. 1.30) [230] and ANCOM-BC (v. 1.6.4) [231]. We defined significant differentially abundant genera if Benjamini-Hochberg (BH) corrected Q-values were below 0.05 in at least 4 out of 5 methods with a false discovery rate (FDR) = 5% [232].

We used Holm multiple testing correction in ALDEx2 as it was the only method implemented in its generalized linear model (GLM) framework.

Regression of microbial genera against smoking intensity. To study the compositional changes of microbial genera in response to the grams of tobacco smoked per day, we modeled each genus in against the grams of tobacco per day as a continuous variable in a Negative binomial GLM (DESeq2). We binned the daily tobacco smoked into multiples of 5 g as those were the most frequent answers (Supplementary File 1, Figure 10). We considered genera as significant when BH-corrected Q-values were lower than 0.05 with FDR = 5%.

Regression of microbial genera against smoking history. To study the compositional changes of microbial genera in response to smoking history, we modeled each genus in response to years since smoking cessation as a continuous variable, at 1-year interval in a Negative binomial GLM (DESeq2). We considered genera as significant when BH-corrected Q-values were lower than 0.05 with FDR = 5%.

Insights into the functional potential of the salivary microbiota. We inferred the functional potential of the oral microbiota at the ASV level using *picrust2_pipeline.py* with default parameters implemented in PICRUSt2 (v. 2.5) [58]. We investigated differential abundant pathways with the same strategy used for genera differential abundance. We considered pathways as significant if the absolute effect size was above 0.5 and the Q-value below 0.05 in at least 4 methods. To further confirm the impact of smoking in relation to the proportion of aerobic taxa, we mapped each genus to a table of curated annotations of three oxygen metabolism classes: aerobic, anaerobic and facultative anaerobic [203]. We visualized the relative abundance of aerobes, anaerobes and facultative anaerobes in each sample with respect to smoking status with pairwise Wilcoxon tests, correcting P-values with BH (FDR = 5%).

3.10 Supplementary files (Links)

Supplementary File 1

Supplementary File 2

4 Geographic and genetic factors on the composition of the adult salivary microbiota of an Italian alpine population

Giacomo Antonello (1,2,#), Michele Filosi (1), Eva König (1), David Emmert (1), Deesha Bhaumik (3), Elyse Davis (3), Peter Pramstaller (1), Cristian Pattaro (1), Nicola Segata (2), Betsy Foxman (3), Christian Fuchsberger (1,#) # Corresponding

(1) Institute for Biomedicine, Eurac Research - Affiliated Institute of the University of Lübeck, Bolzano, Italy
(2) Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy
(3) School of Public Health – Epidemiology, University of Michigan, Ann Arbor, MI

Abstract

The oral microbiota has gained traction as a key player in health and disease. While it has been shown that sharing the same household has an important role in determining salivary microbiota similarity, less is known about geographic distribution and genetics.

We generated 16-V4 salivary microbiota data and explored its association with geographic factors (cohabitation, municipality, and altitude of residence) and family factors (familial relatedness, genetics) in 1,782 participants of the Cooperative Health Research in South Tyrol (CHRIS) study.

We found four study-wide significant (P-value < 1.25×10^{-9}) associations between TOPMed imputed variants and ASVs, three of which with minor allele frequency (MAF) less than 1%, after adjusting for age, sex, smoking, number of teeth, antibiotics usage, and 10 genetic principal components: *Selenomonas* ASV 113 – rs9511156 (MAF = 7.6%) ; *Aggregatibacter* ASV 90 – rs916234787 (MAF = 0.38%) and rs535001228 (MAF = 0.48%); *Prevotella* ASV 178 – rs72809470 (MAF = 0.36%). Gene level analysis on rare variants (MAF < 1%) identified *SRFBP1,* and *LOX* associated with *Selenomonas noxia* (P-value < 3.05×10^{-8}).

Our findings yield novel insights into the genetic contributions to the salivary microbiota composition, emphasizing the substantial impact of environmental and lifestyle factors.

4.1 Introduction

The microbial community stably residing in the oral cavity of humans is called oral microbiota. It is primed in the first months to years of life as a consequence of birth method, exposure to microbes through human contact, lactation, food and the surrounding environment [233,234]. The degree of similarity between mother and child oral and gut microbiota in the first months of life is higher than in adulthood [235,236]. The oral microbiota of adults is highly personalized [237] and plays a role in food processing, antagonism against pathogens and modulation of mucosal immunity [238]. Additionally, it was suggested to significantly contribute to the concentration of circulating nitrite via salivary nitrate reduction [239], which is then converted into nitric oxide by the host, a vasodilator with protective effect on blood pressure and cardiovascular health [196]. On the other hand, the oral microbiota plays a role in the likelihood, progression and severity of conditions like halitosis [109,240] and diseases like dental caries [241] and periodontitis [242]. Periodontitis in particular was associated with higher risk of cardiovascular diseases [187]. Given the accumulating evidence of the importance of the oral microbiota for host health and disease, it is becoming a factor to consider in the public health context [243].

Among all oral niches such as periodontium, gums, teeth surfaces, and the tongue dorsum, the salivary microbiota is investigated for its high diversity [90] and ease of sample collection and preservation. The salivary microbiota of adults is composed of few highly prevalent genera, *Prevotella, Streptococcus, Veillonella*, *Haemophilus, Neisseria* account for approximately 50% of the total bacteria in samples [244,245]. The remaining 50% is composed of a larger set of variable taxa, responsible for most of the salivary microbiota diversity within and between populations. Understanding the factors that associate with population-level compositional variability of the salivary microbiota should be considered for potential future microbiota-based diagnostic and therapeutic tool against diseases and conditions. In particular, the association between the salivary microbiota composition and the geographic distribution of individuals requires further investigation. A pilot study on 50 individuals in northern Italy found region-specific signatures [246] but did not discuss within-region similarity. Here, we explore the compositional variability of the salivary microbiota of an Italian Alpine population in relation to household sharing, genetic background, municipality of residence and altitude of residence in CHRISMB, a salivary microbiota convenience sample within the Cooperative Health Research in South Tyrol (CHRIS) study.

We confirmed the cohabitation associates with a large degree of salivary microbiota compositional similarity and that municipality and altitude of residence are not major environmental factors in the salivary microbiota composition. We show that several salivary taxa have a significant pedigree heritability (N = 16) and household effects (N = 34). We found four study-wide significant ASV-variant associations, and none at genus level, supporting a role of genetics to the salivary microbiota composition. Our analyses confirm that, while familial relationships and genetic factors are associated with the salivary microbiota composition, environmental and lifestyle factors - such as smoking and oral health - have a considerably higher impact.

4.2 RESULTS

We show that cohabitation is strongly associated with a higher microbiota similarity, while municipality and altitude showed weak and no association, respectively. Fourteen genera and 16 ASVs had a significant heritability component ($17.4\% \leq h^2 \leq 30.9\%$), while 14 genera and 37 ASVs showed a significant household component ($9.9\% \leq c^2 \leq 30.8\%$). With the microbiota GWAS (mbGWAS), we found no study-wide significant (SWS) Genus-variant associations. However, we found four SWS ASV-variant associations after adjusting for age, sex, smoking, number of teeth, antibiotics usage, and 10 genetic principal components: *Selenomonas* ASV 113 – rs9511156; *Aggregatibacter* ASV 90 – rs916234787 and rs535001228; *Prevotella* ASV 178 – rs72809470. Gene level analysis on rare variants (minor allele frequency < 1%) further identified *SRFBP1,* and *LOX* associated with *Selenomonas noxia* (P-value < 3.05×10^{-8}).

4.2.1 The majority of CHRISMB participants live in households with up to 3 participants and reside mostly in Mals and Graun

After applying sample inclusion criteria and read quality filtering (see Methods), the population included 1,782 participants, with a median age of 45 years (interquartile range: 31-57, range: 18-93) and 53% females. The majority of participants resided in Mals (N = 853) and Graun (N = 562). Smaller municipalities had generally younger participants (age ≤ 41 years) and smokers proportions varied across municipality, ranging from 19% in Mals to 34.4% in Laas. Households with at least 2 participants were 38.6%, while households with one participants were 32.8%. The remaining CHRISMB participants resided in households with 3 to 9 participants (**Table 4-1**).

Table 4-1. Demographics of CHRISMB in South Tyrol, Italy, with respect to Municipality of residence. Municipalities with less than 100 partcipants were grouped together as "Other". Per-column percentages were also reported in brackets. Population sample statistics were included under the "CHRISMB" column. Household size is the number of people residing at a particular address. Altitudes were categorized as "Low" (< 1,500 m) and "Moderate" (between 1,500 and 2,500 m), according to Parati et al. [247].

	Mals (N=853)	Graun (N=562)	Other (N=367)	CHRISMB (N=1782)
Sex				
Male	402 (47.1%)	263 (46.8%)	173 (47.1%)	838 (47.0%)
Female	451 (52.9%)	299 (53.2%)	194 (52.9%)	944 (53.0%)
Age (years)				
Mean (SD)	45.1 (16.2)	47.7 (15.7)	41.0 (18.5)	45.1 (16.7)
Median	46.0	48.0	36.0	45.0
Range	18.0 – 91.0	18.0 – 86.0	18.0 – 93.0	18.0 – d 93.0
Number of Teeth (self-reported)				
0	52 (6.1%)	37 (6.6%)	14 (3.8%)	103 (5.8%)
1-9	52 (6.1%)	48 (8.5%)	28 (7.6%)	128 (7.2%)
10-19	138 (16.2%)	98 (17.4%)	36 (9.8%)	272 (15.3%)
20+	611 (71.6%)	379 (67.4%)	289 (78.7%)	1279 (71.8%)
Smoking status				
Never	476 (55.8%)	305 (54.3%)	195 (53.1%)	976 (54.8%)
Former	215 (25.2%)	136 (24.2%)	95 (25.9%)	446 (25.0%)
Current	162 (19.0%)	121 (21.5%)	77 (21.0%)	360 (20.2%)
Number of participants per household				
1	250 (29.3%)	131 (23.3%)	204 (55.6%)	585 (32.8%)
2	364 (42.7%)	210 (37.4%)	114 (31.1%)	688 (38.6%)
3	147 (17.2%)	108 (19.2%)	27 (7.4%)	282 (15.8%)
4	72 (8.4%)	96 (17.1%)	16 (4.4%)	184 (10.3%)
5-9	20 (2.3%)	17 (3.0%)	6 (1.6%)	43 (2.4%)
Altitude category				
Low	689 (80.8%)	182 (32.4%)	352 (95.9%)	1223 (68.6%)
Moderate	164 (19.2%)	380 (67.6%)	15 (4.1%)	559 (31.4%)

4.2.2 The salivary microbiota varies largely in relation to cohabitation and marginally to municipality and altitude of residence

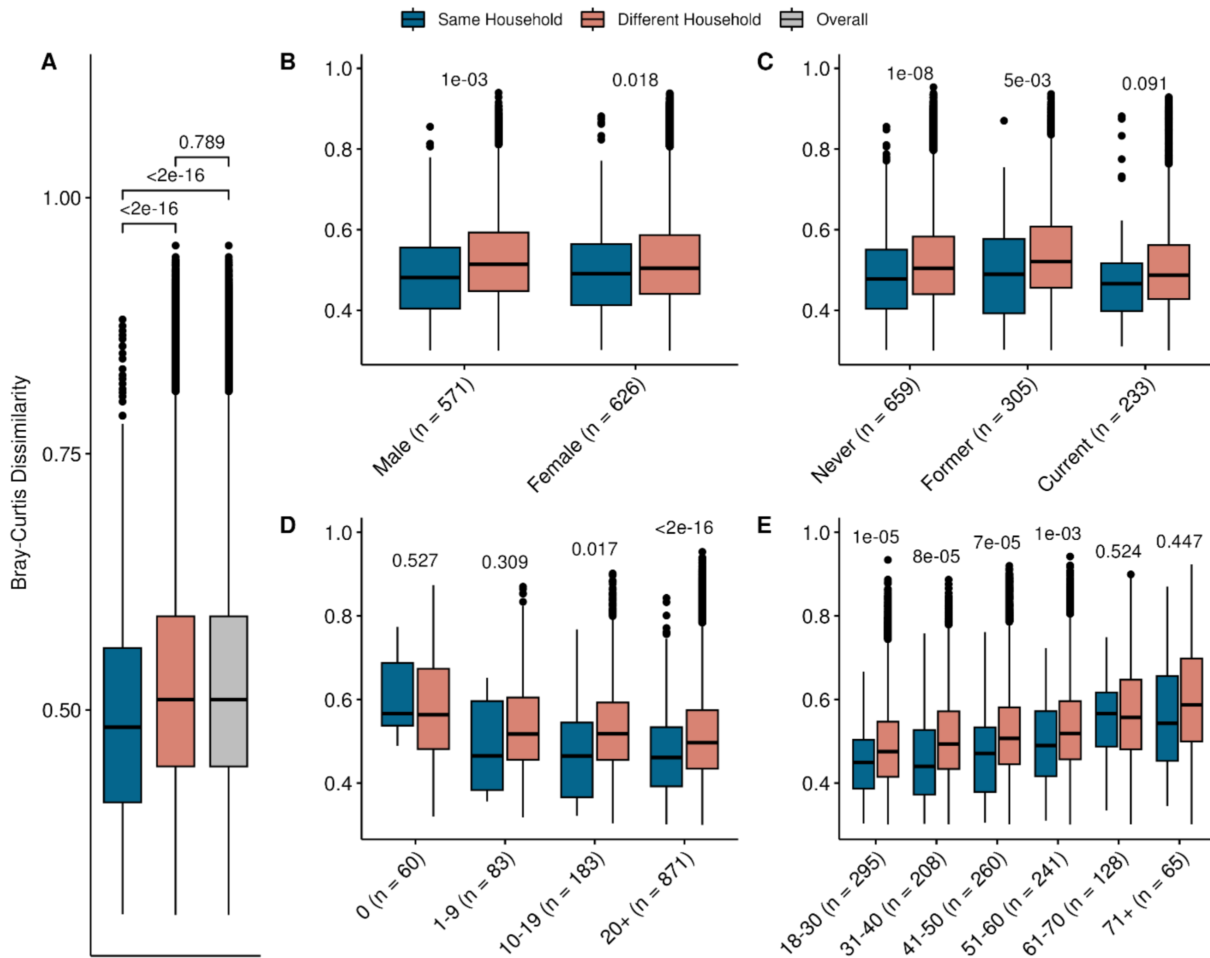


Figure 4-1. Pairs of participants living within households have a more similar salivary microbiota than pairs living between households. Boxplots of pairwise Bray-Curtis dissimilarity (1 – similarity), indices based on 622 ASVs. Lower values indicate a more similar microbiota. The analysis was performed on participants living with at least 1 other participant (N = 1,197). We show the household effect in the overall population **(A)**, as well as split by sex **(B)**, smoking **(C)**, number of teeth category **(D)** and age group **(E)**. Per-group sample sizes were added in the x axis labels. Significance was calculated with pairwise t-test, adjusting P-values with Benjamini-Hochberg under a false discovery rate of 5%.

We quantified the association between household and salivary microbiota variability by modeling the Bray-Curtis distance between the 1,197 people that lived with at least one other person, including sex, smoking, number of teeth, antibiotics usage and age group as covariates in a permutational multivariate analysis of variance (PERMANOVA) model with the *by = "margin"* parameter and 2,000 permutations. The proportion of variance (R^2) associated

with the household variable was 53%. We then visualized the Bray-Curtis dissimilarity index in relation to cohabitation, splitting groups by other factors associated with microbiota variability (age, sex, smoking, and number of teeth), which confirmed that cohabitation was strongly associated with higher microbiota similarity (**Figure 4-1**). We additionally observed a gradual decrease of microbiota similarity at decreasing number of natural teeth and increasing age.

Then we investigated whether the municipality of residence and the altitude of residence associated with higher microbiota similarity. First, we tested the association between salivary microbiota variability and municipality and altitude using PERMANOVA including the marginal effect of age, sex, smoking, number of teeth, antibiotics usage, municipality, and altitude categorized according to Parati et al. [247]. To estimate the role of municipality and altitude on the salivary microbiota limiting the confounding effect that households and small municipalities could have, we tested four PERMANOVA analyses (**Figure 4-2** A). We observed that the R^2 associated with municipality was close to 1% when rare municipalities were included in the model, while it was 0.5% when they were excluded. Sampling 1 individual per household 500 times resulted in comparable R^2 estimates, but insufficient statistical power. Altitude was never significantly associated with salivary microbiota variability. We visualized the population level distribution in relation to municipality (**Figure 4-2** A) and altitude of residence (**Figure 4-2** B) with a PCoA. Next, we tested for differentially abundant ASVs between municipalities adjusting for age, sex, number of teeth, smoking, and antibiotics usage. We used Mals as baseline, as it was the municipality with the highest number of CHRISMB participants and in the low altitude category. We considered significant association at Q-values less than 0.1 divided by the 7 contrasts tested (α = 0.014). We observed 5 ASVs significantly more abundant in participants from other municipalities, with prevalence ranging between 3% and 25%, and $\log_2$ fold change estimates between 0.5 and 1.5 (Supplementary Figure 1). *Treponema* ASV 239 was significant in Graun, *Treponema* HMT 225 and *Treponema* ASV 626 in Schlanders, *Veillonella* ASV 186, and *Capnocytophaga* ASV 387 in Taufers. We applied the same model to find differentially abundant ASVs comparing moderate against low altitude residents, adjusting for age, sex, number of teeth, smoking, and antibiotics usage. We reported four ASVs significantly more abundant in moderate altitudes (Q-value less than 0.1), with a $\log_2$ fold change less than 1. We repeated the model regressing ASVs against continuous altitude values (range 302 – 2296 m), finding 12 significantly more abundant ASVs. All four ASVs found in the categorized differential abundance model were significant in the regression model (Supplementary Figure 2).

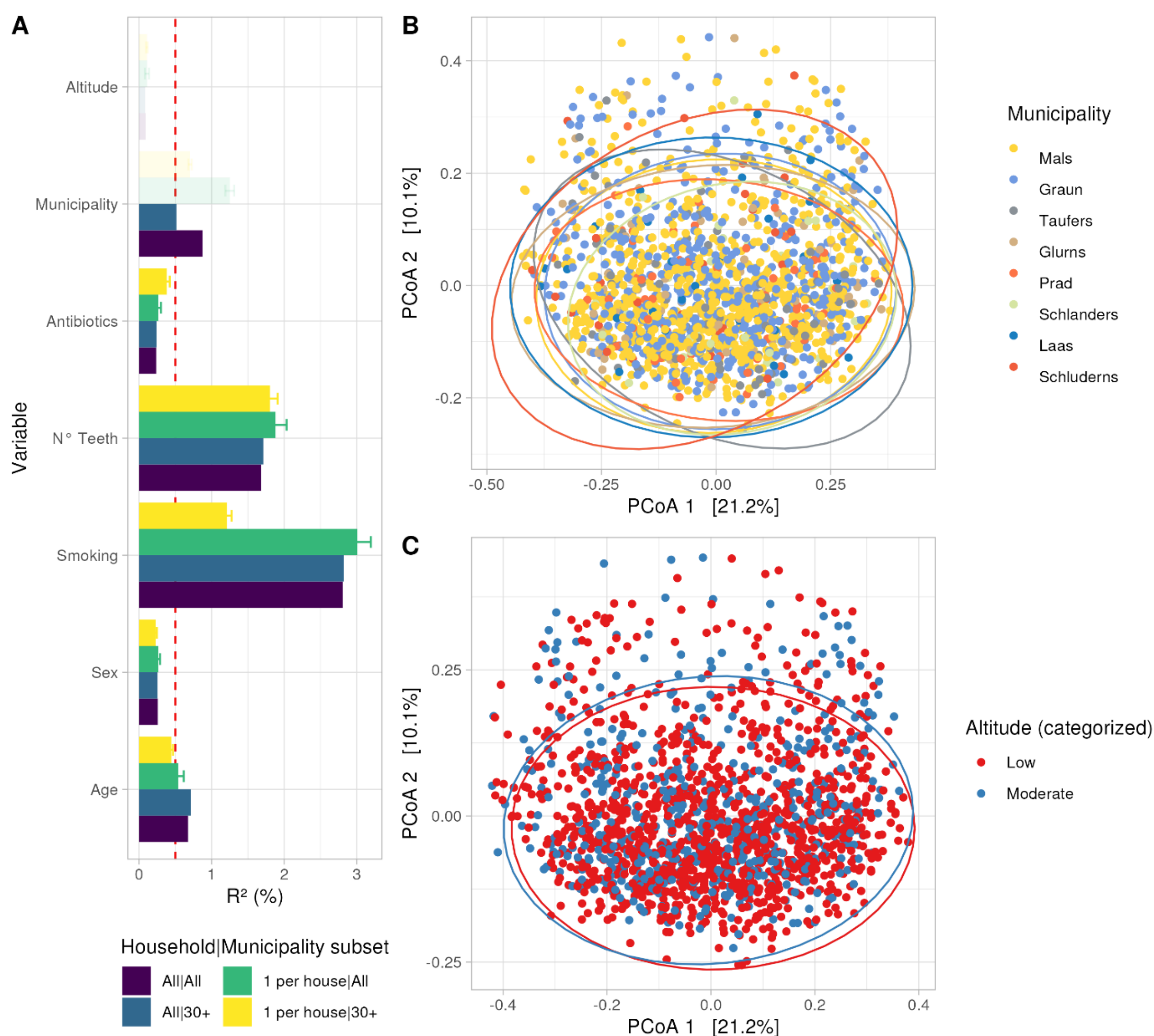


Figure 4-2. The salivary microbiota shows little variation in relation to municipality and altitude of residence on municipalities with at least 30 participants (N = 1,737). (A) PERMANOVA model with percentage of variance explained (R^2) on the x axis for each variable considered (y axis). Transparency of the bars indicates a statistically non-significant association (p-value < 0.05). Four models were compared: "All|All": all households and all municipalities were included; "All|30+": All households were included, along with municipalities with at least 30 participants; "1 per house|All": One participant per household was randomly selected, all municipalities were included; "1 per house|30+": One participant per household was randomly selected, along with municipalities with at least 30 participants. The last two models were repeated with 500 random samples and R^2 and P-value estimates were averaged and standard deviations were reported as error bars. **(B, C)** Principal coordinate analysis on the Bray-Curtis distance. Samples were colored by municipality (B) and altitude of residence (C), the latter categorized according to Parati et al. [247] ("Low" below 1,500 m; "Moderate" between 1,500 and 2,500 m). Ellipses represent the 95% confidence intervals around the samples within each municipality and altitude category.

4.2.3 Familial relationships did not associate with higher salivary microbiota similarity

Because CHRIS encouraged family participation, cohabitation was linked to higher likelihood of relatedness (Supplementary Figure 3). To disentangle the effect of cohabitation from relatedness, we extracted pairwise pedigree relationships of the 1,197 participants selected for the household effect analysis (**Figure 4-1**). Since only 5 grandparents-grandchildren pairs shared the same household (Supplementary Table 1), we retained only first-degree relationships. We observed that the salivary microbiota similarity between siblings was significantly higher than children with their parents and unrelated individuals, both cohabiting and not (**Figure 4-3** A). Additionally, parent-child similarity was not significantly different in relation to cohabitation (**Figure 4-3** B). We performed heritability analysis using the SOLAR algorithm [248], which partitions a trait's variance in relation to relatedness (heritability, h^2) and household (c^2) components, after removing the portion due to fixed effects. We modeled the inverse rank normal transformed relative abundances of 142 ASVs and 48 genera with minimum 20% prevalence, adjusting for sex, age, smoking, antibiotics usage, and number of teeth on all available samples (N = 1,782). We considered h^2 or c^2 as significant if the BH adjusted Q-value was lower than 0.1. We found 16 ASVs with significant h^2 (range: 18% – 30.8%) and 37 with significant c^2 (range: 10.7% – 26.9%). Two ASVs, *Prevotella* ASV 67 and *Stomatobaculum* HMT 97 had both h^2 and c^2 significant components (Supplementary Table 2). ASVs with the highest h^2 components were *M. muciniformis, Lachnospiraceae Genus 2 bacterium HMT_096*, and *Atopobium ASV 28*; while those with the highest c^2 components were *Prevotella* ASV 21 (26.8%), *Porphyromonas gingivalis* (22.9%), and *Gemella* ASV 12 (22.2%) (**Figure 4-3** C). We repeated the same analysis at genus level, finding 14 genera with significant h^2 and 14 with significant c^2 (Supplementary Table 3).

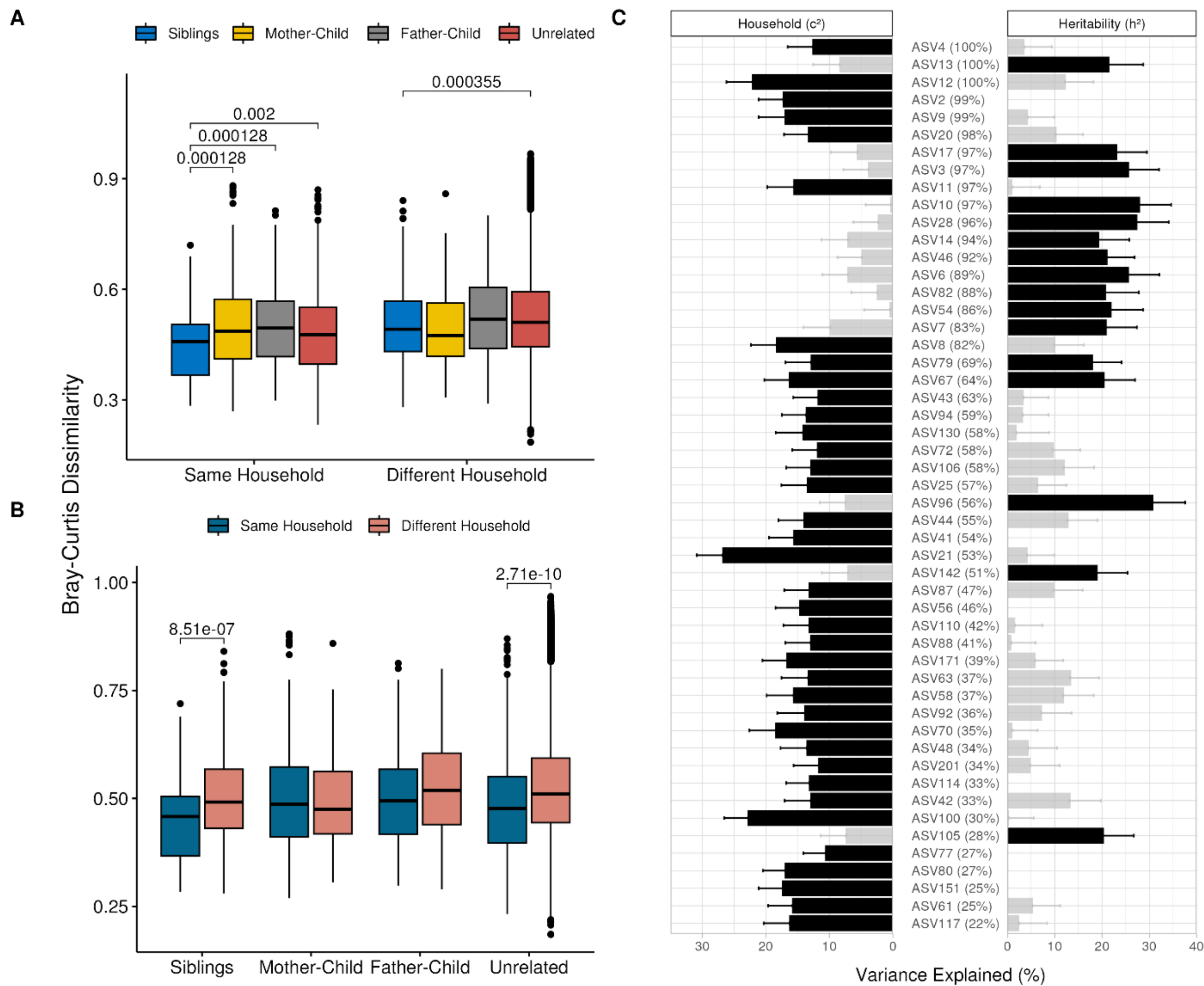


Figure 4-3. Microbiota similarity is associated with household and not with degree of relatedness or age difference in CHRISMB. (A) Bray-Curtis dissimilarity measures in pairs of relatives. Lower values indicate a more similar microbiota. **(B)** Bray-Curtis dissimilarity measures in relation to cohabitaiton, grouped by familial relationship. Lower values indicate a more similar microbiota. **(C)** Heritability (h^2) analysis of 142 ASVs with a minimum prevalence of 20% adjusting for age, sex, smoking, antibiotics usage, and number of teeth in the SOLAR framework. The household contribution (c^2) was also included. Taxa were plotted by decreasing prevalence from top to bottom. Representation was limited to associations with Benjamini-Hochberg adjusted Q-values below 0.1 in h^2 or c^2 components. Components with Q-values > 0.1 were colored in grey.

4.2.4 mbGWAS reveals several ASV-variant associations

Next, we investigated the association between taxa with minimum 20% prevalence at genus and ASV level and genotypes imputed from the TOPMed panel using the GRCh38 genome build. We modeled each taxon against 11 million variants adjusting for sex, antibiotics usage, smoking, number of teeth and the first 10 genetic principal components with the nf-gwas pipeline, a Nextflow pipeline [249] which exploites the Regenie algorithm [250] for association testing. We did not find study wide significant (SWS) associations at genus level (P-value < 3.85×10^{-9}, see Methods), but we found four SWS ASV-variant associations (P-value < 1.25×10^{-9}, see Methods for details) (**Table 4-2**). We investigated these signals by visualizing neighbour signals in each locus with LocalZoom.

Table 4-2. Study-wide significant genotype variants associated with microbiota features at ASV level. Study wide significance at ASV level was set to 1.25×10^{-9} (see Methods for details). The association was performed with the "nf-gwas" pipeline [249], adjusting for age, sex, smoking, number of teeth, usage of antibiotics, and the first 10 genetic principal components using TOPMed imputed genotypes. ASV: Amplicon Sequence Variant; rsID: Reference SNP cluster ID; Chr: Chromosome; Pos: Position on reference genome; MAF: Minor Allele Frequency.

ASV	rsID Chr. Pos	MAF (%)	Nearest Gene	Distance	$-\log_{10}(P)$	Beta	Std. Err.	Imputation quality
ASV 113	rs951115613 13 24223648	7.6	AL359736.1 – SPATA13	0	9.68337	0.369	0.058	0.801
ASV 90	rs916234787 14 70110307	0.38	SLC8A3	0	9.35422	1.264	0.203	0.856
ASV 148	rs72809470 16 88611992	0.48	ZC3H18	0	8.96617	1.510	0.248	0.601
ASV 90	rs535001228 14 70191018	0.36	SLC8A3	1948	8.9074	1.137	0.187	0.885

Conditional analysis on these variants did not identify additional signals. We tested the cumulative effect of rare variants (MAF ≤ 1%) with gene level analysis defining 3 variant masks in relation to the predicted impact of the variants in each locus - low, moderate, and high impact - as annotated in the Variant Effect Predictor (VEP) [251]. We found 2 loci associated with *Selenomonas noxia* ASV 172: *SRFBP1*, (low impact variants P-value = 2.04×10^{-8};

moderate impact variants P-value = 2.04×10^{-8}) and the low impact mask of the neighboring *LOX* gene (P-value = 1.90×10^{-8}).

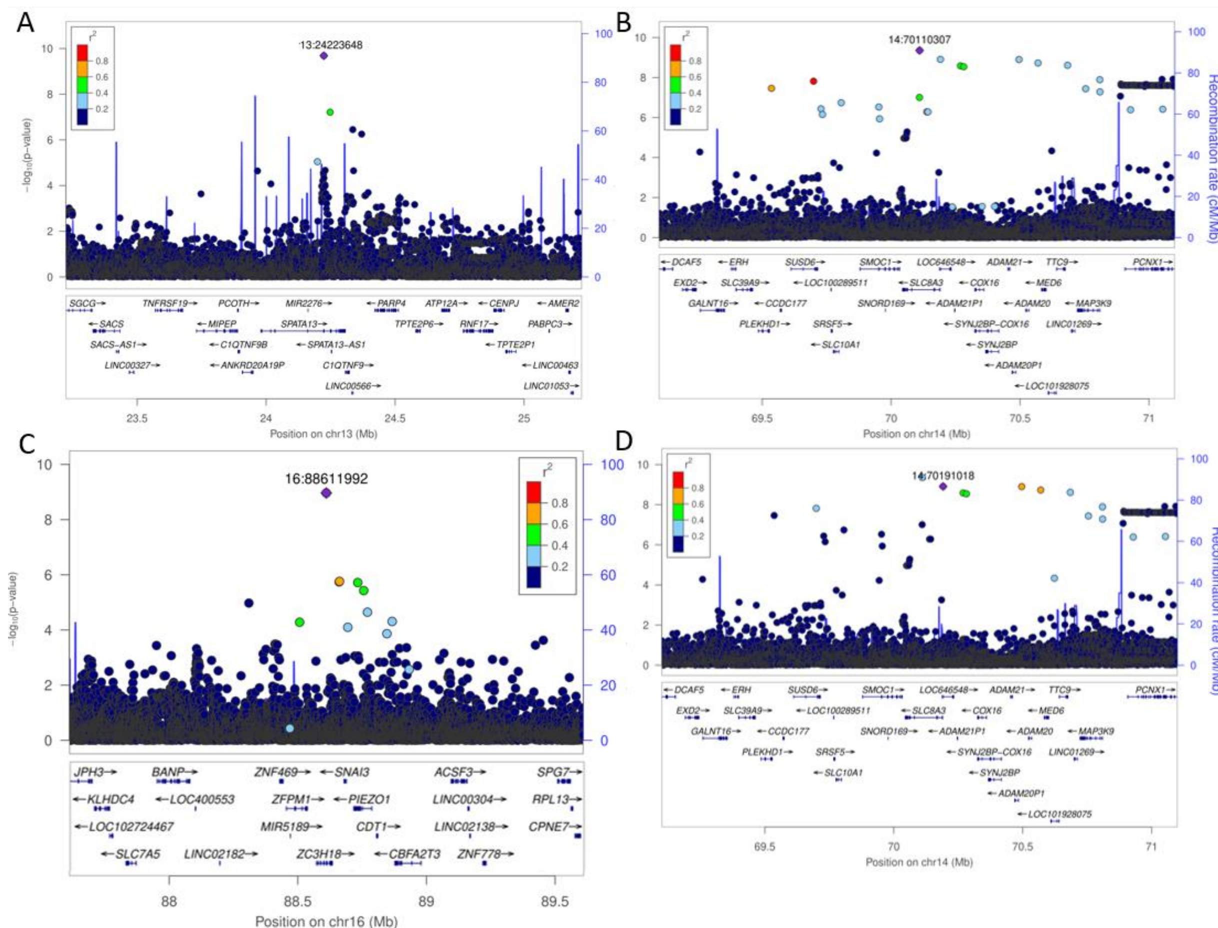


Figure 4-4. Locus Zoom plots of the four ASV-gene associations found with the mbGWAS. (A) ASV 113 – rs9511156. **(B)** ASV 90 – rs916234787. **(C)** ASV148 – rs72809470. **(D)** ASV 90 - rs535001228. Plots were generated with LocusZoom standalone. Each point is a variant, colored by its linkage disequilibrium (LD) with respect to the leading variant for each locus (purple diamond). Variants with missing LD are colored in grey. The bright blue line indicates the recombination rate, expressed in centimorgan/megabase, as seen on the right y axis. Recombination Rate and linkage disequilibrium were calculated from the CHRISMB data.

4.3 Discussion

In this work we explored the association between geographic and genetic factors and salivary microbiota composition of an Italian Alpine population in CHRISMB, a convenience sample within the CHRIS study with focus on microbiota variability in relation to household sharing, relatedness and genetics, municipality, and altitude of residence. We found that the genetic

component is limited but significant, while the environmental component, first of all cohabitation, was strongly associated, followed by age and lifestyle such as smoking and oral health.

4.3.1 The salivary microbiota varies largely in relation to cohabitation and marginally to municipality and altitude of residence

People cohabiting had a higher salivary microbiota similarity than people not cohabiting, which also reflected in the number of ASVs with a significant household component (**Figure 4-1** C). This result, also previously observed in the gut [235,236,252] and in the oral microbiome [236], is generally attributed to a higher likelihood of sharing lifestyle and environments, but more importantly to a higher interindividual transmission of microorganism. Intimate kissing is considered the major player for microbial exchange events [253]. Social interaction, including talking and touching, also increase the sharing of microbes, as shown in gut microbiomes of mice [254] and humans [255]. Household effects were not visible in participants with 9 or fewer natural teeth (**Figure 4-1** D) and in older age groups (**Figure 4-1** E). These two observations are likely linked ($X^2 = 721.95$, df = 15, P-value $< 2.2 \times 10^{-16}$), since tooth loss generally occurs at older age [256–258]. If this finding is replicated with more individuals, it will further highlight the role of natural teeth as a surface for oral microbes to grow [259]: tooth loss, associated with decreased alpha diversity and increased compositional variability [260] could decrease chances of stable microbial exchange between individuals.

Municipality of residence is associated with small differences in microbiota profiles as seen in PERMANOVA and PCoA (**Figure 4-2**). Few low-prevalence (< 5%) taxa were differentially abundant between municipalities with fewer participants, with Mals as a baseline (Supplementary Figure 1). Previous literature highlighted that microbiota dissimilarity across different cities or municipalities of residence are possible for both the stool [252] and the salivary microbiota [261,262]. However, they also highlight the role of physical contact between individuals and lifestyle similarities as microbiota similarity enhancers. CHRISMB participants are a homogeneous population with respect to ethnicity, traditions and lifestyle, as well as living area, which could all explain the low microbiota variability in relation to municipality and altitude of residence. More extreme contrasts would facilitate isolation of municipalities, with lower chances of social contact with the outside. We found *F. nucleatum* subsp. Vincentii [263] to be significantly more abundant in participants living in the moderate altitude group. *F. nucleatum* is acknowledged as a risk factor for several oral and non-oral

diseases [264]. A study reported that people inhabiting high altitude areas (3,200 m) in a Saudi Arabia region showed more dysbiotic oral microbiomes than sea-level controls [265]. Another small study on 167 Tibetans [266] reported salivary microbiota differences between ultra-high (> 4,000 m) and high-altitude (< 3,650 m) residents. Taken together, salivary microbiome variation in relation to altitude could be a sign of physiological and lifestyle adaptation.

4.3.2 The salivary microbiota variability shows contributions of familial relatedness and households

We observed that siblings and unrelated individuals had a significantly higher salivary microbiota similarity when cohabiting (**Figure 4-3** A), while mother-child and father-child pairs did not (**Figure 4-3** A, B). We also investigated the heritability and household components of the salivary microbiota on a large number of participants with ages ranging from 18 to 92 years old, finding 17 significant h^2 components and 37 c^2 components (**Figure 4-3** C). We argue that it is possible that some salivary microbial signatures persist among family members living in different households, possibly due to a higher genetic similarity. Multiple twins studies, reported heritable microbiome components both for the gut and the salivary microbiome [267–270]. Demmit et al., in particular, performed a large heritability analysis of the salivary microbiota, finding *Granulicatella* h^2 = 55.8% in a cohort of adult twins in Colorado, United States [268]. In our study, genus *Granulicatella* was the 5th genus with the highest h^2 component (23.9%, BH Q-value = 0.004). Heritability differences are known to be mainly to population differences, sampling and analysis methods and characteristics of the population [150]. Regarding the latter, studies on younger individuals, especially twin studies, generally show higher heritability estimates [267–270].

4.3.3 The salivary microbiota is mildly associated with host genetics

We found support for a role of genetics in determining the salivary microbiota composition, with 4 SWS variants: *Selenomonas* ASV 113 – rs9511156; *Aggregatibacter* ASV 90 – rs916234787; Absconditabacteria (SR1) [G-1] bacterium HMT 874 ASV 148 – rs72809470; ASV 90 – rs535001228. Variant rs9511156 was previously reported in studies regarding lung oxygen uptake performance [271,272], but not related to *Selenomonas* taxa. Variants rs916234787 and rs535001228 are rare variants annotated in the *SLC8A3* gene locus. The former was a rare variant (MAF = 0.7%), the latter is an intergenic variant, almost 2,000 base pairs away from the *SLC8A3* coding region. *SLC8A3* is a gene involved in calcium-mediated signaling in mitochondria and excitable neurons [273]. Variant rs72809470 is a predicted intron variant in the *ZC3H18* zinc-finger gene, which was previously reported to enhance

homologous recombination in ovarian cancer models [274]. We additionally found one gene-level association between *Selenomonas noxia* ASV 172 and the *SRFBP1-LOX* locus. Serum Response Factor Binding Protein 1 (*SRFBP1*), a predicted rRNA maturation protein, which was reported highly and moderately expressed in salivary glands and oral mucosa, respectively [275]. Lysil oxidase gene (*LOX*) was significantly associated in the gene level analysis, indicating a cumulative effect of rare variants on that locus. Some variants on this locus were associated with aortic aneurysm, the weakening of arterial walls [276]. *S. noxia* was also previously associated with increased cardiovascular disease risk [277], a connection which requires further exploration. Although intriguing, our findings highlight borderline associations with imputed variants, stressing the need to replicate our findings. Secondly, locus zoom plots did not highlight variants in LD with the leading variant. This phenomenon, indicative of higher likelihood of a false positive association, was already described in mbGWAS studies [278].

Our mbGWAS is of comparable size with the work performed on a Chinese population of around 2,000 individuals, finding associations between tongue dorsum and salivary microbiota composition and host genetics [133]. Our results were incomparable with theirs, mainly due to different data generation from the genetic, lifestyle and microbiota standpoint. Additionally, while both studies contribute with a remarkable sample size from a single population, they are both underpowered. In fact, it was calculated that a minimum of 25,000 samples would be required to detect taxon-variant associations with a bacterium prevalence of 10% and a variant with a contribution of 0.4% of the phenotype variance [278]. Our study, in addition to previous work [133,279,280], supports a genetic component to the salivary microbiota composition, which should be investigated further with larger sequencing efforts and meta-analyses. CHRISMB, for example, may be extended with the salivary microbiota and genetic data from the COHRA2 cohort, in which microbiota and host genetic data were generated with protocols comparable to CHRISMB [120].

4.3.4 Limitations of the study

Our study has a few limitations. First, oral health and status were self-reported by the participants, rather than by a professional. Amplicon sequencing technology is a cost-effective method to obtain microbiota data at the expenses of taxonomic resolution. In the mbGWAS context, microbiota data is challenging to model due to its sparsity, which further interacts with the sparsity of genomic variants, which increase changes of structural non-positivity of some groups. Traditional heritability analysis methods, such as SOLAR, tend to

inflate h^2 estimates on compositional data [281]. While we tried to minimize the bias with data transformation, microbiome-tailored heritability analysis methods could be developed.

4.3.5 Strengths of the study

Our study has a strong setup compared to previously published studies: to the best of our knowledge, it is among the largest data sets worldwide able to investigate salivary microbiota heritability and mbGWAS. Moreover, the population is culturally and geographically homogeneous, which reduces the effect of confounders. The household and pedigree data in the CHRIS study are manually curated and validated. Amplicon sequencing data were performed at a high depth and with automated robot handling, which minimize human error and contaminations.

4.4 Conclusions

We confirmed the cohabitation associates with a large degree of salivary microbiota compositional similarity. We show that several salivary taxa have a significant pedigree heritability and household effects. With heritability analysis, we found 14 ASVs with significant heritability and 34 with significant household components. We found four study-wide significant ASV-variant associations, and none at genus level, supporting a role of genetics to the salivary microbiota composition. While the municipality and altitude of residence may affect the salivary microbiota composition as proxies of environmental exposure, the highly uniform living environment of CHRISMB participants showed a limited contribution. Our analyses show that other factors, mainly smoking and oral health, have a much larger effect on salivary microbiota composition than the region of residence, in CHRISMB. Future studies could investigate the dynamics of divergence of microbiomes over time and space.

4.5 Acknowledgements

We thank the CHRIS study personnel of the South Tyrol Healthcare System (SABES/ASDAA) for carrying out the data and sample collection in the CHRIS visit center in Schlanders/Silandro. Special thanks to Deborah Mascalzoni and Massimiliano Pellegrini for providing the ethical and legal framework for CHRIS. We thank Dr. Hannes Gatterer for providing insights into interpretation of the results related to the altitude. The CHRIS study was funded by the Department of Innovation, Research and University of the Autonomous Province of Bolzano-South Tyrol and supported by the European Regional Development Fund

(FESR1157). The CHRIS biobank was assigned the "Bioresource Research Impact Factor" code BRIF6107. The CHRISMB microbiota data generation was funded by the National Institute of Dental and Craniofacial Research grant number R01 DE014899.

4.6 METHODS

4.6.1 Study design, ethical approval, and data collection

The CHRIS study was approved by the local Ethical Committee within the South Tyrol healthcare on April 19, 2011, and registered with code 21.2011. Participants, all at least 18 years old, were recruited from the Vinschgau/Val Venosta district (South Tyrol, Italy). On the day of visit, participants answered lifestyle, dietary, general health, and socio-economic status questionnaires [201]. The CHRIS Salivary microbiota (CHRISMB) project was a convenience sample of CHRIS participants recruited between January 2017 and February 2018. Selection criteria and sampling and data generation strategies were thoroughly described elsewhere [282]. In brief, unstimulated saliva was collected in Oragene OG-500 tubes. Salivary DNA was extracted with an automatic liquid handler, the V4 region of the 16S rRNA gene was amplified and indexed with the dual indexing strategy. Sequencing was performed on an Illumina MiSeq, fastq reads were processed with DADA2 using the eHOMD database, further expanded with oral-specific bacterial genomes.

4.6.2 Epidemiological data generation

We defined age as the difference between the examination date and the birth date, rounded to the closest integer. CHRISMB participants completed an adapted version of the World Health Organization oral health survey [216], which contained a question related to the number of natural teeth grouped in four ranges: 0, 1-9, 10-19 and 20 or more. We derived smoking variables from smoking questionnaires harmonized from the European Community Respiratory Health Survey III questionnaire [217]. We used participants' residence address to derive household identifiers, municipality, and altitude of residence. We grouped altitude in categories according to Parati et al. [247]: Low altitude was up to 1,500 m, Medium was between 1,501 and 2,500, no other categories were present.

4.6.3 Salivary microbiota data generation

The sample generation protocol was described elsewhere [282]. In brief, unstimulated saliva samples were processed with a robot liquid handler to generate 16-V4 amplicon sequencing data.

Fastq file filtering, trimming clustering and ASV calling were performed based on perfect identity of sequences and taxonomic calling with DADA2 v.1.12 using the extended Human Microbiome Database with additional oral species included as reference database.

We generated a phyloseq object starting from the counts table, taxonomic table and taxonomy tree using the Bioconductor package “phyloseq” (v. 1.42.0) [223] and “ape” (v. 5.7). We retained only those taxa that were present with at least 10 counts in 1% of individuals with the function *core* of the “microbiome” package (v. 1.20.0) [224]. When required, we aggregated ASVs at the genus level with the *tax_glom* function in the GitHub package “speedyseq” (“mikemc/speedyseq”).

4.6.4 Sample selection and platform for statistical analysis

We retained samples with non-missing data for sex, age, smoking status, number of natural teeth, household, municipality, altitude and usage of antibiotics in the 3 months prior to the visit. Unless stated otherwise, we performed within the R (v. 4.3.1) [222] and RStudio Server (v. 2023.03.0 Build 386) framework.

4.6.5 Microbiota analyses

4.6.6 Beta diversity Matrix calculation

To generate the pairwise dissimilarity matrix (beta-diversity), we transformed the 622 filtered ASVs to relative abundance and calculated the Bray-Curtis dissimilarity for all samples using the *distance* function in the “phyloseq” package (v. 1.44.0). We subsequently used the same matrix for all statistical analyses and visualizations below to limit sample subsets similarity biases.

4.6.7 Permutational Multivariate Analysis of Variance (PERMANOVA)

To test microbiota variability in relation to multiple variables, we modeled the beta diversity matrix with a Permutational Multivariate Analysis of Variance (PERMANOVA), adding desired variables into the model. We used the *“by = “margin”* parameter to estimate the proportion of variance explained (R^2) of each variable against the remaining variables in the model. We used the *“permutations = 2000”* parameter to generate a null distribution by permuting the investigated variable 2,000 times and to generate a Monte Carlo simulated P-value [173,225]. We used the *adonis2* implementation in the “vegan” package v.2.6-4 to perform the calculations.

4.6.8 Differential abundance analysis

We performed differential abundance analyses using the LinDA algorithm [228] using the adaptive option, and quantile winsorization equal to 0.97. We set the significance threshold to $\alpha = 0.1$.

4.6.9 Microbiota variation in relation to geographic exposure: household, municipality and altitude of residence

To assess the association between microbiota variability and household sharing, we selected participants living with at least 1 other participant and made a PERMANOVA model including age, sex, number of teeth, smoking status, and antibiotics usage, and household identifier. We then visualized within-household and between-household beta diversity distributions separating participants by sex, smoking, number of teeth and age groups.

We performed a PERMANOVA model including age, sex, number of teeth, smoking status, antibiotics usage, and municipality to assess the impact of the latter on the microbiota variability. To account for the effect of cohabitation and smaller municipalities, we performed four models as follows: (1) All households and municipalities; (2) One participant per household, all municipalities; (3) All households, municipalities with at least 30 participants; (4) One participant per household, municipalities with at least 30 participants. In models (2) and (4) we performed PERMANOVAs on 500 random samples of 1 participant per household.

We then performed a differential abundance analysis on the sample set (3), comparing each municipality against Mals as baseline and adjusting for age, sex, number of teeth, smoking status, and antibiotics usage. We adjusted the significance threshold by the number municipalities tested against Mals (n , leading to α = 0.014.

We applied the same differential abundance model to investigate the association of microbiota variation in relation to the altitude adjusting for age, sex, number of teeth, smoking status, and antibiotics usage.

To visualize the salivary microbiota variation in relation to municipality and altitude of residence, we generated principal coordinate analysis (PCoA) plots, from the first two eigenvectors of the dissimilarity matrix with *ordinate* in “phyloseq” and visualized them with *plot_ordination* drawing 95% confidence interval ellipses with *stat_ellipse* in the “ggplot2” package (v. 3.4.0).

4.6.10 Microbiota variability associated with familial relatedness (heritability analysis)

We investigated the association between participants' relatedness and the variability of each genus and ASV with heritability analysis using SOLAR-Eclipse (or SOLAR) [248,283], a variance decomposition method which takes a pedigree as source of relatedness and allows partitioning of the household component. We transformed the relative abundance of the taxa above 20% prevalence with the inverse rank normal transformation to approximate normality. We added age, sex, smoking, number of teeth and antibiotics usage as confounders, and added household identifiers to calculate the household variance component. A detailed description of the protocol and scripts can be found on Github.

4.6.11 Association between salivary microbiota and host genetic variants (mbGWAS)

We investigated the association between the salivary microbiota composition and host genetic variants with a microbiota GWAS (mbGWAS). Host genetic variants were genotyped using Illumina Human Omni2.5 Exome and Illumina OmniEURHD chips. Raw genotypes were processed with Illumina GenomeStudio to exclude samples with a call rate less than 0.985 and variants with a GenTrain score below 0.7. Subsequently, variants were filtered according to the following criteria: (1) B allele corruption using BAFRegress; (2) identification of inconsistent relatedness estimates using KING; (3) identification of inconsistent sex estimates using PLINK 1.9 "sex check". The batches were merged sequentially, keeping variants present on both array chips and samples with 5% missingness at most. We expanded the number of testable variants by imputing hard-called genotypes using the TOPMed imputation server. Given the explorative nature of the analysis, we retained the 11 million variants with MAF larger than 0.2% for the analysis. We modeled the inverse rank normal transformed taxa with minimum prevalence of 20% against the 11 million filtered variants adjusting for age, sex, number of teeth groups, smoking (binary) and antibiotics usage (binary). We retained participants who took antibiotics to leverage on the sample size and because we observed that antibiotics affected the salivary microbiota composition significantly but marginally (R^2 = 0.2%, P-value = 0.001, Supplementary Table 4), as confirmed in previous studies [284,285]. We further included the first 10 genetic principal components as covariates to account for population stratification and relatedness. We considered nominal genome-wide significant (GWS) variants with a P-value less than the significance threshold $\alpha = 5x10^{-8}$. We defined a study-wide significance (SWS) threshold by dividing the GWS threshold by the number of taxa needed to explain 95% variability in a Bray-Curtis based principal component ordination.

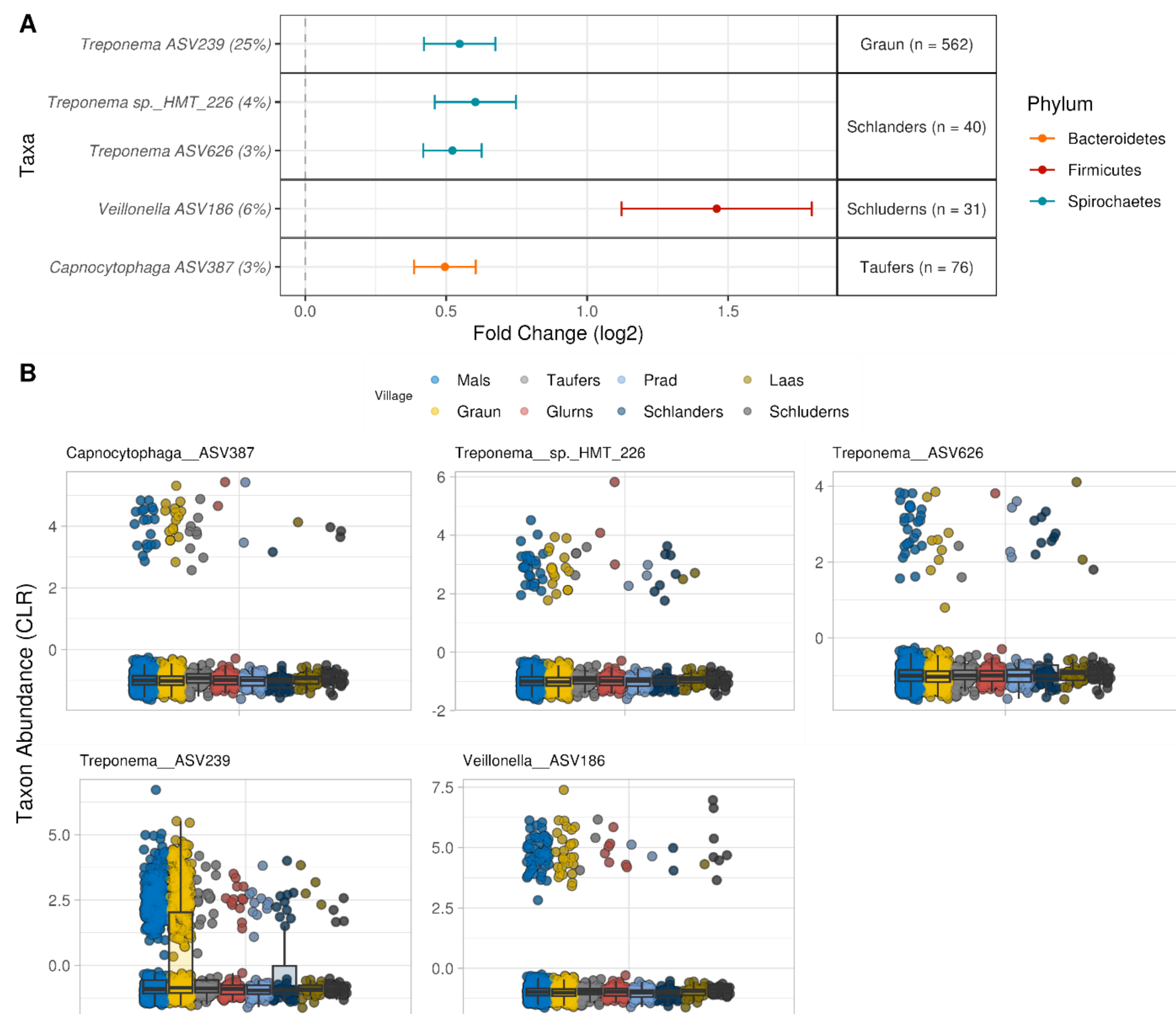
For the genus and ASV level mbGWAS, the SWS threshold was $\alpha < \frac{5x10^{-8}}{13} = 3.85 \times 10^{-9}$ and $\alpha < \frac{5x10^{-8}}{40} = 1.25 \times 10^{-9}$, respectively. We visualized SWS loci using LocusZoom standalone version 1.4 [286], flanking the leading variants with upstream and downstream. We used CHRISMB imputed genotype data to calculate linkage disequilibrium information using the "--ld-vcf" parameter. To assess the strength of association between the most significant variants and microbiota features, we further performed a GWAS conditioning on the variant with the most significant P-value in each locus by extracting their dosage from the genotype files and including them as covariates into the model, as proposed previously [287].

4.6.12 Association between rare variants and microbiota features (Gene-level Analysis)

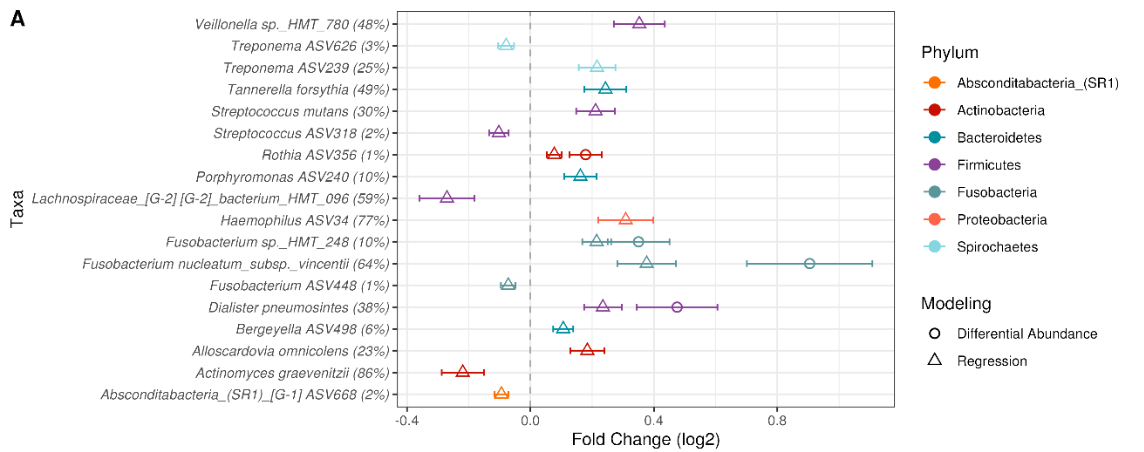
To increase power for rare variant association tests, we performed gene level analysis with the SKATO-ACAT test [288] implemented in Regenie using three variant impact masks: low, moderate, and high impact as defined by Variant Effect Predictor [251]. We defined nominal gene-level-significant (GLS) signals if $P < \frac{0.05}{N°Loci\ tested} = \frac{0.05}{41{,}005} = 1.22 \times 10^{-6}$, which we further adjusted for multiple with the same strategy as the mbGWAS, leading to a Genus-GLS and ASV-GLS threshold of $\frac{1.22\times10^{-6}}{13} = 9.38 \times 10^{-8}$and $\frac{1.22\times10^{-6}}{40} = 3.05 \times 10^{-8}$, respectively.

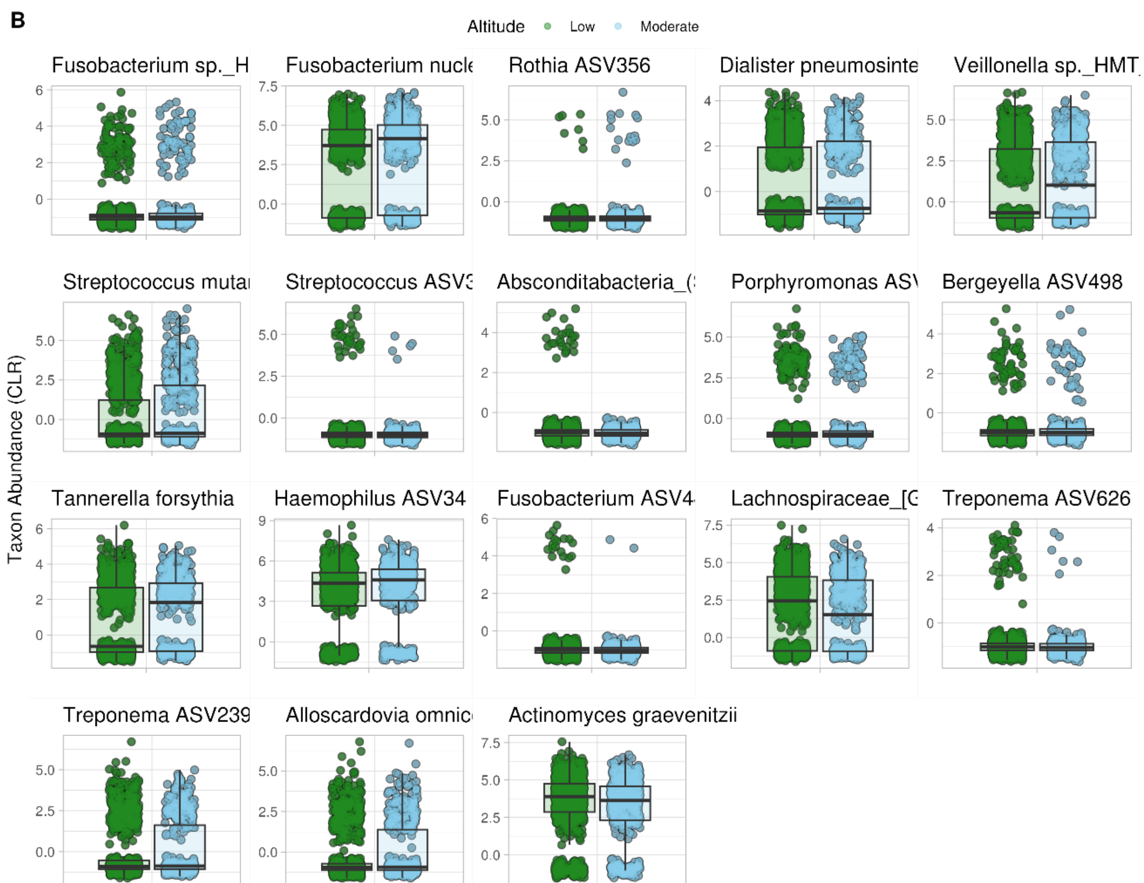
4.7 SUPPLEMENTARY MATERIAL

4.7.1 Supplementary Figures

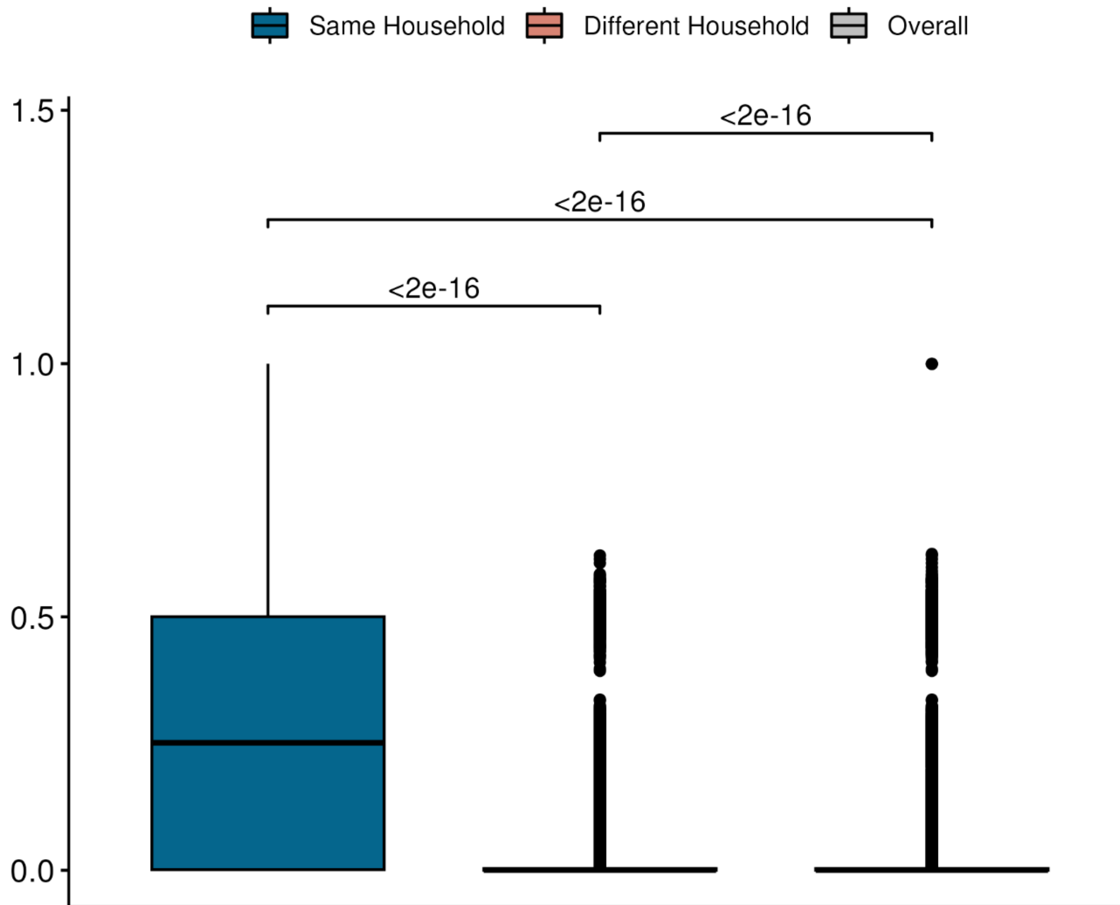


Supplementary Figure 1. Differentially abundant ASVs across municipalities with at least 30 CHRISMB participants (Mals as baseline). The model was adjusted for age, sex, smoking and number of teeth, and antibiotics usage with the LinDA differential abundance algorithm [228]. Significance was defined as for Benjamini-Hochberg (FDR 5%) adjusted Q-values less than 0.1 divided by the number of municipalities tested against Mals, the baseline (n = 7). **(A)** Forest plot of the differentially abundant taxa as a result of the differential abundance. ASV prevalence was added in brackets after each ASV name on the y axis. **(B)** Center-log ratio transformed abundances of differentially abundant ASVs in any of the 5 municipalities tested against Mals.

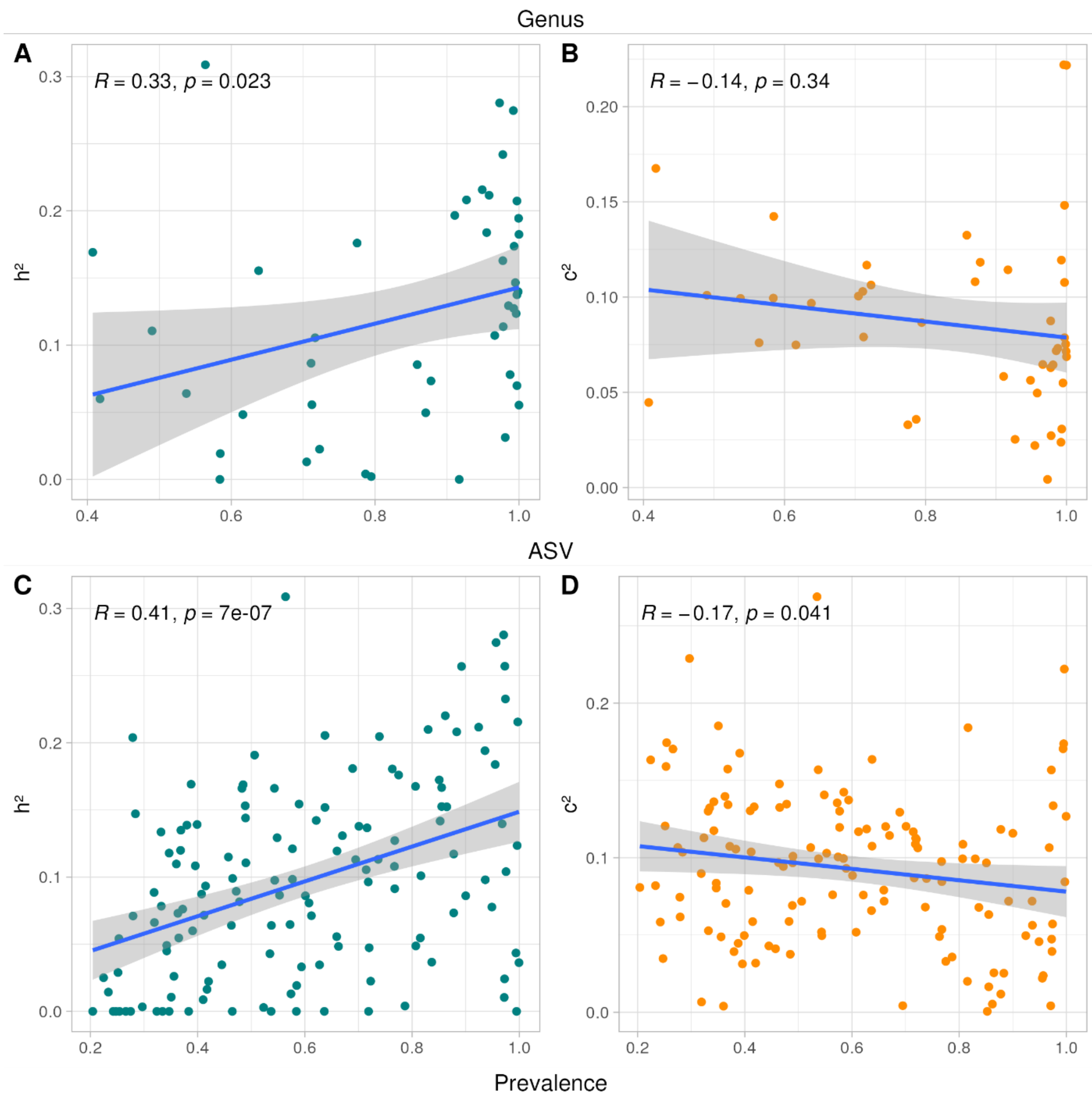




Supplementary Figure 2. Differentially abundant ASVs comparing residents at moderate altitude (1,500-2,500 m) with residents at low altitude (< 1,500 m) The model was adjusted for age, sex, smoking and number of teeth in the LinDA differential abundance algorithm [228]. Significance was defined as for Benjamini-Hochberg (FDR 5%) adjusted Q-values less than 0.1. **(A)** Forest plot of the differentially abundant taxa as a result of a LinDA model, adjusting for age, sex, number of teeth and smoking. **(B)** Center-log ratio transformed abundances of differentially abundant ASVs found in (A).



Supplementary Figure 3. Pairwise kinship values in relation to cohabitation. Kinship values were calculated with Plink based on CHRISMB genotypes. Higher values indicate a higher genetic similarity. Significance was calculated with pairwise t-test with Benjamini-Hochberg P-value adjustment (FDR = 5%).



Supplementary Figure 4. SOLAR heritability estimates mildly correlate with prevalence, while household estimates do not. The trend is visible both at Genus **(A, B)** and ASV level **(C, D).** Spearman's correlation coefficient (R) and significance (p) are reported in the top left of each plot. The analysis was performed with SOLAR on taxa at genus and ASV level with minimum prevalence of 20%. The models were adjusted for fixed effects of age, sex, smoking, number of teeth and antibiotics usage. The household component was included as separate parameter, as indicated in the SOLAR documentation.

4.7.2 Supplementary Tables

Supplementary Table 1. Pairwise familial relationships available in CHRISMB (N = 1782). Relationships were derived from the CHRIS study pedigree using the "FamAgg" package. Siblings and vertical relationships were derived. Unrelated individuals were defined as having kinship equal to 0.

	Same Household	Different Household
Siblings	108	419
Mother	221	87
Father	171	45
Grandmother (Mother's)	1	14
Grandfather (Mother's)	1	5
Grandmother (Father's)	2	14
Grandfather (Father's)	1	3
Unrelated	413	1,397,937

Supplementary Table 2. Heritability (h^2) and household effect (c^2) on ASV-level taxa with prevalence above 20% Samples were scaled by relative abundance and ASVs were normalized with the inverse rank normal transformation. Only ASVs with all parameters correctly estimated were kept.

ASV (Prevalence %)	h^2		c^2	
	Estimate (Std.Err.)	Q-value (BH)	Estimate (Std.Err.)	Q-value (BH)
Lachnospiraceae_[G-2] [G-2] bacterium HMT 096 (56.4%)	0.309 (0.067)	7.00E-05	0.076 (0.039)	0.038
Megasphaera micronuciformis (97.1%)	0.28 (0.066)	2.00E-04	0.004 (0.039)	0.46
Atopobium ASV28 (95.7%)	0.275 (0.066)	2.00E-04	0.024 (0.039)	0.291
Veillonella ASV3 (97.3%)	0.257 (0.063)	2.00E-04	0.039 (0.039)	0.176
Fusobacterium periodonticum (89.2%)	0.257 (0.064)	2.00E-04	0.072 (0.039)	0.048
Prevotella salivae (97.4%)	0.233 (0.062)	6.00E-04	0.057 (0.041)	0.106
Lachnoanaerobaculum orale (86.2%)	0.22 (0.067)	0.003	0.005 (0.04)	0.46
Granulicatella adiacens (99.7%)	0.216 (0.072)	0.005	0.084 (0.041)	0.031

Bergeyella sp. HMT 322 (92.4%)	0.212 (0.057)	6.00E-04	0.05 (0.038)	0.117
Prevotella histicola (83%)	0.21 (0.064)	0.003	0.099 (0.042)	0.015
Mogibacterium ASV82 (88.3%)	0.208 (0.07)	0.005	0.025 (0.04)	0.289
Prevotella ASV67 (63.7%)	0.205 (0.064)	0.003	0.164 (0.039)	1.00E-04
Actinomyces sp. HMT 172 (73.9%)	0.205 (0.076)	0.011	0.087 (0.041)	0.026
Leptotrichia ASV105 (27.9%)	0.204 (0.063)	0.003	0.074 (0.039)	0.04
Veillonella parvula (93.6%)	0.194 (0.064)	0.005	0.072 (0.041)	0.054
Kingella oralis (50.6%)	0.191 (0.063)	0.005	0.072 (0.04)	0.05
Solobacterium moorei (95.5%)	0.184 (0.065)	0.008	0.022 (0.04)	0.313
Stomatobaculum sp. HMT 097 (68.9%)	0.181 (0.06)	0.005	0.129 (0.04)	0.002
Corynebacterium durum (76.3%)	0.181 (0.062)	0.006	0.049 (0.038)	0.125
Peptostreptococcaceae [XI] [G-1] [XI][G-1]_sulci (77.5%)	0.176 (0.069)	0.014	0.033 (0.046)	0.261
Neisseria ASV5 (85.1%)	0.172 (0.065)	0.011	0.097 (0.04)	0.014
Mitsuokella sp. HMT 521 (38.8%)	0.169 (0.069)	0.017	0.045 (0.038)	0.146
Veillonella sp. HMT 917 (48.5%)	0.169 (0.058)	0.006	0.038 (0.037)	0.177
Corynebacterium matruchotii (80.6%)	0.168 (0.063)	0.011	0.099 (0.04)	0.012
Porphyromonas ASV22 (85.5%)	0.167 (0.06)	0.008	0.063 (0.039)	0.072
Actinomyces lingnae [Not_Validly_Published] (54.3%)	0.166 (0.062)	0.011	0.052 (0.041)	0.125
Tannerella forsythia (48.3%)	0.166 (0.061)	0.01	0.059 (0.035)	0.059
Alloprevotella sp. HMT 473 (58.9%)	0.154 (0.065)	0.02	0.093 (0.039)	0.015
Rothia ASV55 (48.9%)	0.153 (0.056)	0.009	0.069 (0.039)	0.053
Capnocytophaga leadbetteri (85.5%)	0.153 (0.057)	0.011	0.017 (0.038)	0.349
Rothia dentocariosa (86.5%)	0.152 (0.058)	0.012	0.026 (0.037)	0.27
Stomatobaculum longum (63.7%)	0.152 (0.061)	0.016	0.107 (0.04)	0.007
Prevotella shahii (28.4%)	0.147 (0.061)	0.018	0.104 (0.04)	0.01

Fusobacterium nucleatum subsp. animalis (48.9%)	0.144 (0.061)	0.02	0.097 (0.039)	0.013
Veillonella dispar (62.1%)	0.142 (0.057)	0.016	0.076 (0.039)	0.039
Alloprevotella sp. HMT 308 (85.2%)	0.142 (0.063)	0.027	0.001 (0.039)	0.493
Actinomyces ASV16 (96.8%)	0.14 (0.066)	0.034	0.107 (0.042)	0.011
Fusobacterium ASV85 (39.9%)	0.139 (0.062)	0.029	0.05 (0.04)	0.132
Leptotrichia hongkongensis (38%)	0.139 (0.061)	0.026	0.039 (0.039)	0.177
Capnocytophaga gingivalis (70.1%)	0.138 (0.063)	0.03	0.12 (0.041)	0.005
Actinomyces ASV62 (71.6%)	0.137 (0.065)	0.034	0.087 (0.04)	0.025
Haemophilus ASV63 (36.9%)	0.135 (0.059)	0.023	0.134 (0.041)	0.002
Alloprevotella sp. HMT 914 (33.2%)	0.134 (0.064)	0.035	0.13 (0.04)	0.002
Capnocytophaga sputigena (67%)	0.131 (0.061)	0.033	0.114 (0.039)	0.005
Streptococcus ASV44 (54.8%)	0.129 (0.061)	0.034	0.141 (0.04)	0.001
Haemophilus ASV34 (76.8%)	0.127 (0.065)	0.048	0.085 (0.038)	0.022
Gemella ASV12 (99.6%)	0.124 (0.058)	0.034	0.222 (0.04)	6.00E-07
Granulicatella elegans (57.7%)	0.121 (0.062)	0.05	0.13 (0.038)	0.001
Aggregatibacter ASV58 (36.8%)	0.12 (0.063)	0.051	0.157 (0.041)	5.00E-04
Prevotella denticola (66%)	0.119 (0.058)	0.036	0.072 (0.039)	0.048
Campylobacter sp. HMT 044 (34.7%)	0.118 (0.063)	0.053	0.08 (0.039)	0.031
Leptotrichia sp. HMT 417 (87.8%)	0.117 (0.062)	0.053	0.012 (0.039)	0.399
Prevotella sp. HMT 305 (45.7%)	0.115 (0.065)	0.065	0.041 (0.037)	0.157
Neisseria ASV29 (73.7%)	0.113 (0.059)	0.051	0.068 (0.038)	0.05
Fusobacterium ASV59 (69.5%)	0.113 (0.06)	0.053	0.004 (0.036)	0.46
Peptococcus ASV182 (49%)	0.111 (0.061)	0.056	0.101 (0.04)	0.011
Prevotella sp. HMT 317 (36%)	0.11 (0.055)	0.046	0.004 (0.036)	0.46
Haemophilus ASV47 (39.6%)	0.108 (0.059)	0.056	0.031 (0.036)	0.209

Lachnoanaerobaculum umeaense (76.7%)	0.108 (0.057)	0.053	0.098 (0.039)	0.012
Ruminococcaceae_[G-1] [G-1]_bacterium_HMT 075 (71.4%)	0.105 (0.061)	0.071	0.117 (0.042)	0.006
Campylobacter ASV20 (97.5%)	0.104 (0.056)	0.053	0.134 (0.038)	0.001
Prevotella pallens (81.6%)	0.101 (0.061)	0.075	0.184 (0.039)	3.00E-05
Parvimonas ASV87 (46.5%)	0.099 (0.06)	0.075	0.133 (0.039)	0.002
Capnocytophaga granulosa (57.7%)	0.098 (0.056)	0.066	0.12 (0.039)	0.003
Oribacterium sinus (93.7%)	0.098 (0.058)	0.073	0.056 (0.04)	0.103
Campylobacter ASV118 (54.4%)	0.098 (0.06)	0.078	0.05 (0.037)	0.116
Porphyromonas ASV66 (71.8%)	0.096 (0.059)	0.077	0.109 (0.041)	0.009
Rothia aeria (41.5%)	0.093 (0.059)	0.083	0.059 (0.04)	0.095
Streptococcus ASV30 (76.8%)	0.091 (0.055)	0.075	0.054 (0.039)	0.113
Veillonella sp. HMT 780 (47.2%)	0.089 (0.056)	0.081	0.094 (0.039)	0.014
Streptococcus sanguinis (31.9%)	0.089 (0.06)	0.102	0.09 (0.04)	0.022
Aggregatibacter ASV90 (40.7%)	0.087 (0.06)	0.109	0.079 (0.037)	0.025
Lautropia mirabilis (55.3%)	0.087 (0.061)	0.117	0.103 (0.039)	0.008
Prevotella nigrescens (90%)	0.086 (0.057)	0.098	0.116 (0.04)	0.005
Prevotella sp. HMT 306 (60.1%)	0.086 (0.065)	0.136	0.089 (0.041)	0.024
Alloprevotella rava (47.8%)	0.082 (0.074)	0.178	0.135 (0.046)	0.005
Kingella ASV69 (60.8%)	0.081 (0.059)	0.126	0.052 (0.039)	0.116
Prevotella veroralis (33.2%)	0.078 (0.071)	0.177	0.053 (0.045)	0.143
Selenomonas ASV24 (94.9%)	0.078 (0.066)	0.162	0.046 (0.04)	0.149
Oribacterium parvum (37.2%)	0.076 (0.063)	0.156	0.107 (0.038)	0.006
Peptostreptococcus stomatis (87.8%)	0.073 (0.061)	0.156	0.118 (0.042)	0.006
Absconditabacteria_(SR1)_[G-1] (SR1)_[G-1] bacterium_HMT 345 (36.3%)	0.073 (0.063)	0.168	0.14 (0.042)	0.002

Rothia ASV60 (41.2%)	0.072 (0.055)	0.139	0.104 (0.039)	0.008
Selenomonas ASV113 (61.2%)	0.071 (0.063)	0.17	0.117 (0.04)	0.005
Fretibacterium ASV188 (27.9%)	0.071 (0.057)	0.151	0.062 (0.039)	0.076
Selenomonas ASV143 (31.9%)	0.066 (0.054)	0.152	0.007 (0.035)	0.441
Prevotella ASV25 (57.2%)	0.065 (0.06)	0.184	0.135 (0.04)	0.002
Leptotrichia sp. HMT 212 (46.3%)	0.064 (0.062)	0.194	0.097 (0.041)	0.016
Abiotrophia defectiva (53.8%)	0.064 (0.053)	0.159	0.099 (0.039)	0.011
Filifactor alocis (39.1%)	0.06 (0.058)	0.196	0.168 (0.038)	6.00E-05
Catonella morbi (65.9%)	0.056 (0.06)	0.222	0.079 (0.041)	0.041
Absconditabacteria_(SR1)_[G-1] (SR1)_[G-1] bacterium_HMT 874 (36.5%)	0.055 (0.06)	0.225	0.07 (0.038)	0.048
Aggregatibacter ASV33 (81.5%)	0.055 (0.054)	0.2	0.02 (0.037)	0.315
Absconditabacteria_(SR1)_[G-1] ASV61 (25.3%)	0.054 (0.057)	0.216	0.159 (0.038)	1.00E-04
Capnocytophaga sp. HMT 326 (34.2%)	0.049 (0.061)	0.264	0.118 (0.039)	0.004
Veillonella rogosae (80.7%)	0.049 (0.062)	0.264	0.109 (0.038)	0.006
Prevotella nanceiensis (66.3%)	0.048 (0.062)	0.264	0.12 (0.04)	0.004
Leptotrichia sp. HMT 221 (72%)	0.047 (0.063)	0.268	0.112 (0.039)	0.005
Fusobacterium ASV48 (34.2%)	0.045 (0.06)	0.268	0.136 (0.041)	0.002
Streptococcus ASV9 (99.4%)	0.043 (0.056)	0.264	0.17 (0.041)	1.00E-04
Prevotella ASV21 (53.5%)	0.043 (0.056)	0.264	0.269 (0.04)	2.00E-09
Leptotrichia sp. HMT 215 (83.7%)	0.037 (0.06)	0.311	0.068 (0.038)	0.05
Streptococcus ASV4 (99.9%)	0.036 (0.057)	0.307	0.127 (0.039)	0.002
Fretibacterium fastidiosum (44.5%)	0.035 (0.059)	0.317	0.043 (0.04)	0.162
Fusobacterium nucleatum subsp. vincentii (62.7%)	0.035 (0.052)	0.298	0.119 (0.038)	0.003
Treponema denticola (59.4%)	0.033 (0.054)	0.311	0.137 (0.038)	7.00E-04

Prevotella oris (25.1%)	0.029 (0.055)	0.336	0.121 (0.04)	0.004
Leptotrichia ASV112 (35.6%)	0.026 (0.061)	0.367	0.049 (0.036)	0.113
Aggregatibacter ASV117 (22.4%)	0.025 (0.059)	0.367	0.163 (0.04)	2.00E-04
Prevotella melaninogenica (97.3%)	0.024 (0.055)	0.367	0.047 (0.039)	0.137
Peptostreptococcaceae_[XI][G-9] [XI][G-9]_brachy (72.3%)	0.022 (0.057)	0.374	0.106 (0.039)	0.007
Treponema socranskii (42%)	0.022 (0.053)	0.367	0.032 (0.035)	0.202
Butyrivibrio sp. HMT 455 (58.5%)	0.019 (0.069)	0.416	0.142 (0.042)	0.002
Prevotella oris (41.8%)	0.016 (0.058)	0.416	0.133 (0.04)	0.002
Prevotella sp. HMT 309 (23.3%)	0.014 (0.055)	0.421	0.082 (0.037)	0.022
Dialister invisus (57.4%)	0.013 (0.057)	0.43	0.101 (0.041)	0.014
Alloprevotella ASV70 (35.1%)	0.011 (0.053)	0.438	0.185 (0.041)	5.00E-05
Streptococcus ASV11 (97.2%)	0.01 (0.058)	0.442	0.157 (0.041)	4.00E-04
Alloprevotella tannerae (41%)	0.009 (0.05)	0.442	0.13 (0.039)	0.002
Ruminococcaceae_[G-2] [G-2] bacterium_HMT 085 (78.7%)	0.004 (0.065)	0.479	0.036 (0.039)	0.203
Porphyromonas gingivalis (29.7%)	0.003 (0.052)	0.479	0.229 (0.037)	8.00E-09
Prevotella ASV51 (52.3%)	0.003 (0.063)	0.482	0.107 (0.04)	0.008

Supplementary Table 3. Heritability (h^2) and household effect (c^2) on genus-level taxa with prevalence above 20% Samples were scaled by relative abundance and ASVs were normalized with the inverse rank normal transformation. Only ASVs with all parameters correctly estimated were kept.

Genus (Prevalence %)	**h^2**		**c^2**	
	Estimate (Std.Err.)	**Q-value (BH)**	**Estimate (Std.Err.)**	**Q-value (BH)**
Lachnospiraceae_[G-2] (56.4%)	0.309 (0.067)	3.00E-05	0.076 (0.039)	0.047
Megasphaera (97.3%)	0.28 (0.066)	4.00E-05	0.004 (0.039)	0.456

Atopobium (99.2%)	0.275 (0.066)	4.00E-05	0.024 (0.039)	0.28
Capnocytophaga (97.8%)	0.242 (0.058)	4.00E-05	0.063 (0.038)	0.069
Stomatobaculum (94.9%)	0.216 (0.059)	3.00E-04	0.056 (0.039)	0.092
Bergeyella (95.8%)	0.212 (0.057)	3.00E-04	0.05 (0.038)	0.112
Mogibacterium (92.7%)	0.208 (0.07)	0.003	0.025 (0.04)	0.28
Granulicatella (99.7%)	0.207 (0.07)	0.003	0.108 (0.041)	0.012
Corynebacterium (91.1%)	0.197 (0.063)	0.003	0.058 (0.04)	0.093
Prevotella (99.9%)	0.194 (0.063)	0.003	0.072 (0.038)	0.049
Solobacterium (95.5%)	0.184 (0.065)	0.004	0.022 (0.04)	0.297
Veillonella (100%)	0.182 (0.061)	0.003	0.069 (0.037)	0.049
Peptostreptococcaceae_[XI][G-1] (77.5%)	0.176 (0.069)	0.01	0.033 (0.046)	0.257
Fusobacterium (99.3%)	0.174 (0.061)	0.004	0.031 (0.039)	0.249
Mitsuokella (40.7%)	0.169 (0.069)	0.012	0.045 (0.038)	0.145
Neisseria (97.8%)	0.163 (0.064)	0.01	0.087 (0.04)	0.03
Absconditabacteria_(SR1)_[G-1] (63.8%)	0.155 (0.062)	0.01	0.097 (0.039)	0.015
Leptotrichia (99.5%)	0.147 (0.067)	0.024	0.055 (0.038)	0.093
Actinomyces (99.9%)	0.14 (0.065)	0.024	0.075 (0.042)	0.052
Rothia (99.7%)	0.137 (0.064)	0.024	0.079 (0.04)	0.047
Alloprevotella (98.5%)	0.129 (0.064)	0.032	0.072 (0.04)	0.052
Campylobacter (99.3%)	0.127 (0.059)	0.024	0.119 (0.038)	0.004
Gemella (99.6%)	0.124 (0.058)	0.024	0.222 (0.04)	3.00E-07
Lachnoanaerobaculum (97.8%)	0.114 (0.06)	0.043	0.027 (0.037)	0.257
Peptococcus (49%)	0.111 (0.061)	0.045	0.101 (0.04)	0.014
Porphyromonas (96.6%)	0.107 (0.057)	0.043	0.065 (0.039)	0.069
Ruminococcaceae_[G-1] (71.7%)	0.105 (0.061)	0.059	0.117 (0.042)	0.01

Lautropia (71.1%)	0.087 (0.061)	0.106	0.103 (0.039)	0.011
Kingella (85.9%)	0.085 (0.061)	0.106	0.133 (0.04)	0.003
Oribacterium (98.8%)	0.078 (0.056)	0.106	0.073 (0.04)	0.052
Peptostreptococcus (87.8%)	0.073 (0.061)	0.146	0.118 (0.042)	0.01
Haemophilus (99.7%)	0.07 (0.063)	0.169	0.148 (0.041)	0.001
Abiotrophia (53.8%)	0.064 (0.053)	0.147	0.099 (0.039)	0.014
Filifactor (41.8%)	0.06 (0.058)	0.187	0.168 (0.038)	4.00E-05
Catonella (71.2%)	0.056 (0.06)	0.207	0.079 (0.041)	0.049
Streptococcus (100%)	0.055 (0.057)	0.199	0.222 (0.04)	3.00E-07
Treponema (87%)	0.05 (0.057)	0.222	0.108 (0.038)	0.009
Fretibacterium (61.6%)	0.048 (0.061)	0.241	0.075 (0.04)	0.049
Selenomonas (98.1%)	0.031 (0.064)	0.348	0.064 (0.04)	0.072
Peptostreptococcaceae_[XI][G-9] (72.3%)	0.022 (0.057)	0.378	0.106 (0.039)	0.011
Butyrivibrio (58.5%)	0.019 (0.069)	0.417	0.142 (0.042)	0.003
Dialister (70.5%)	0.013 (0.057)	0.428	0.101 (0.041)	0.016
Ruminococcaceae_[G-2] (78.7%)	0.004 (0.065)	0.484	0.036 (0.039)	0.212
Tannerella (79.5%)	0.002 (0.053)	0.484	0.087 (0.037)	0.019

Supplementary Table 4 - PERMANOVA modeling the Bray-Curtis dissimilarity matrix in relation to sex, age, number of teeth, smoking, altitude of residence, municipality of residence, and antibiotics usage within 3 months prior to visit. The model was calculated with the *vegan::adonis2* function using the marginal effect of each covariate, age was model as a continuous variable. The total number of samples was 1,782 after keeping samples with available data for each variable and municipalities with a minimum of 20 participants. Deg. Freedom: degrees of freedom, R^2: percentage of variance explained, P-value: Monte Carlo simulated P-value calculated on 2,000 sample permutations.

	Deg. Freedom	R^2 (%)	P-value
Sex	1	0.3	0.001
Age	1	1.2	0.001
Number of Teeth	3	2.1	0.001
Smoking Status	2	2.8	0.001
Altitude of Residence	1	0.07	0.176
Municipality of Residence	7	0.5	0.009
Antibiotics	1	0.2	0.001

5 ADDITIONAL CONTRIBUTIONS

As introduced in the Background section, CHRISMB is a rich data resource that ranges from self-reported questionnaires to biochemical characterization of blood and urine. My main work has focused on detailed smoking habits, host genetics and the CHRISMB alpine geographic peculiarities. In this chapter hereby present several relevant analyses which are not yet published.

5.1 ODOR IDENTIFICATION CAPABILITIES AND SALIVARY MICROBIOTA COMPOSITION ARE NOT SIGNIFICANTLY ASSOCIATED IN CHRISMB

5.1.1 Background

Taste perception is partially determined by genetics [289] and the salivary microbiota [290]. Odor sensitivity and identification ability has a genetic [291] and nasal microbiota component [292]. While it is known that the oral microbiome's metabolism generates a plethora of volatile compounds [22], some of which responsible for breath smell (i.e. Ammonia and hydrogen sulfide) [109,240], the association between salivary microbiota and smell has not yet been investigated. We tested the association between the salivary microbiota composition and smell identification capabilities in CHRISMB, the largest resource with both phenotypes available.

5.1.2 Methods

Participants were presented with 16 felt-tip pens imbued with different solutions of odor molecules (ODOFIN Burghart's smell test) and were required to choose the smell recognized among 4 options. Participants who were aware of smell conditions, including pre-diagnosed infectious or chronic respiratory diseases, hay fever, and septum deviation did not take undergo the test (N = 313). We calculated the smell identification score (SIS) as the sum of the smells correctly recognized. We categorized SIS, which ranged from 0 (none correct) to 16 (all correct), into anosmic/hyposmics/normosmics score (AHN) according to the following criteria: anosmics ($SIS \leq 4$); hyposmics ($5 \leq SIS \leq 11$); normosmics ($SIS \geq 12$). We then grouped anosmics and hyposmic due to insufficient anosmics cases (N = 4). We retained participants with available data on sex, age, smoking status, number of teeth, and antibiotics usage.

Afterwards, we calculated beta diversity with the Bray-Curtis dissimilarity on the filtered microbiota data (10 counts in at least 1% samples) and used it to plot a PCoA coloring both by SIS and AHN scores.

Next, we tested the statistical association between salivary microbiota variability and smell scores using *adonis2* in the “vegan” package, using 2,000 permutations and the by = “margin” parameter, which tests the proportion of variance attributed to each variable in the model against the remaining variables. In addition to SIS and AHN, we included sex, age, smoking status, number of teeth, and antibiotics usage.

Finally, we performed regression and differential abundance analysis using LinDA to test whether, respectively, SIS or AHN scores were associated with abundance variation of each ASV abundance. We applied the “adaptive = TRUE ”, and “winsor.quan = 0.97” parameters. ASVs with Benjamini-Hochberg adjusted Q-values below 0.1 were considered significant.

5.1.3 Results

After excluding samples with missing data on age, sex, smoking, number of teeth, antibiotics usage, and SIS, we retained 1,610 participants. We observed that SIS had a negative skewness, with a median of 14 pens correctly identified, which resulted in 230 hyposmics/anosmics, and 1,380 normosmics. The salivary microbiota composition did not vary in relation to SIS (PERMANOVA R^2 = 0.1%, P-value = 0.087) an minorly in relation to AHN (PERMANOVA R^2 = 0.1%, P-value = 0.013) (**Figure 5-1** B, C). We then performed both microbiota regression in response to SIS and differential abundance analysis comparing hyposmics against normosmics, adjusting for age, sex, smoking, number of teeth, antibiotics usage. Respectively, the regression model did not highlight taxa correlated with SIS, while the differential abundance showed *Prevotella nanceinensis* ($\log_2$ Fold Change = -1.18, Q-value = 0.07), and *Haemophilus parainfluenzae* ($\log_2$ Fold Change = -0.38, Q-value = 0.06) to be significantly less abundant in hyposmics, while *Prevotella marshii* ($\log_2$ Fold Change = 0.30, Q-value = 0.07), was slightly more abundant (BH Q-value < 0.1) (**Figure 5-1** D).

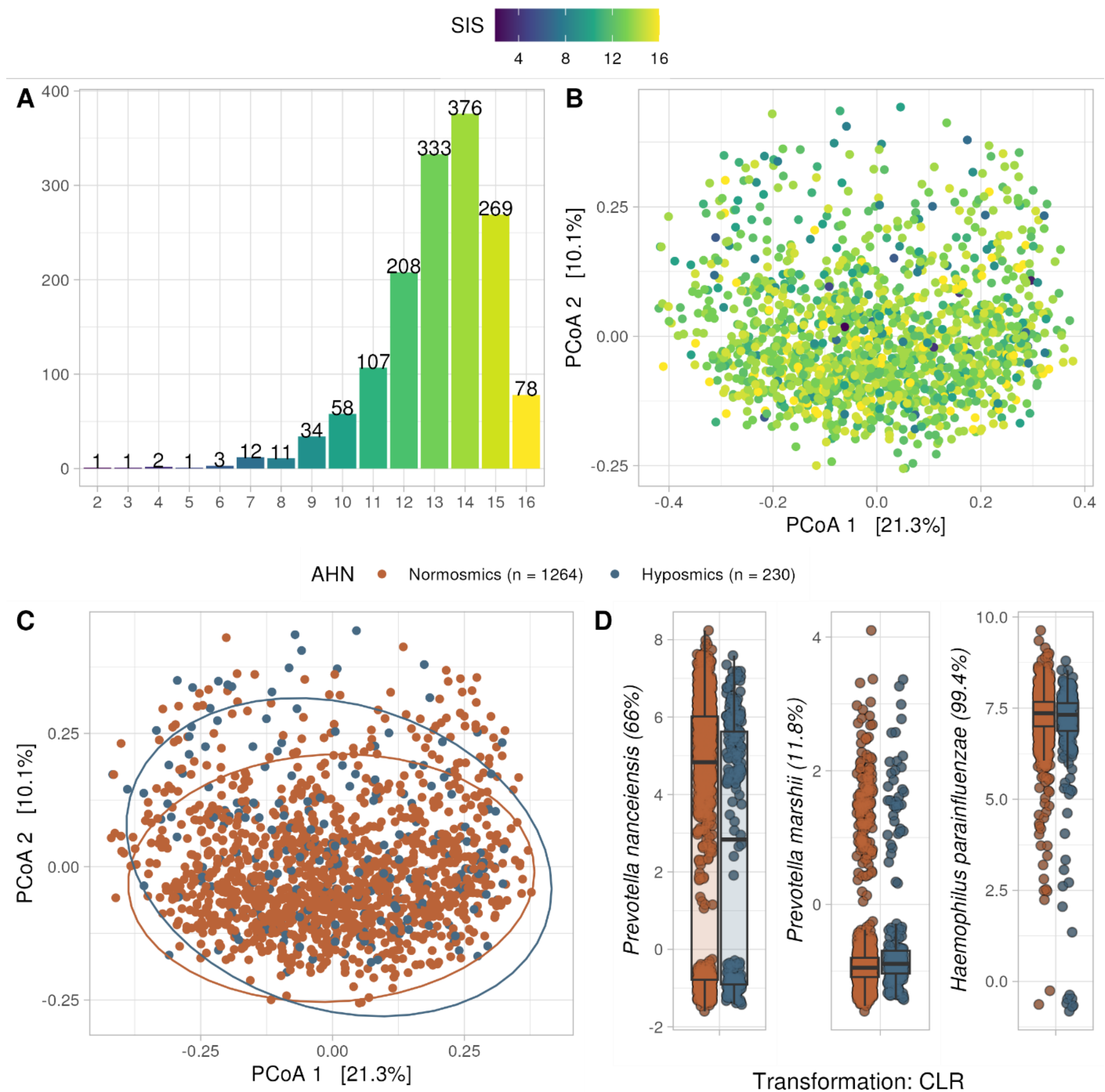


Figure 5-1. Relationship between the salivary microbiota composition and the cumulative and categorized smell identification scores. (A) Histogram of the number of smell pens correctly recognized by CHRISMB participants, defined as smell identification score (SIS). **(B)** Principal Coordinate Analysis on the Bray-Curtis dissimilarity at ASV level. confidence areas (95%) were drawn as ellipses. Axes x and y were chosen as the principal components which explained most of the overall microbiota variability, which is shown in square brackets. Samples were colored by SIS as seen in panel A. **(C)** Same representation as B, coloring by Anosmic/Hyposmic/Normosmic (AHN) groups: Hyposmic ($0 \leq SIS \leq 11$), Normosmics ($SIS \geq 12$). **(D)** Center-log ratio abundances of the differentially abundant (Benjamini-Hochberg Q-values < 0.1) ASVs comparing hyposmics against normosmics, adjusting for age, sex, smoking, number of teeth, and antibiotics usage with the LinDA [228] framework. ASV prevalence was shown as a percentage on the y axis.

5.1.4 Conclusions

In this analysis we explored whether salivary microbiota composition associated with odor identification capabilities, with the hypothesis that salivary microbiota catabolites might alter the perception of some smells. To our knowledge, this is the largest data set (N = 1,610) that investigates the association between odor identification performance and the salivary microbiota composition in a healthy adult population. We found weak associations between smell identification performance and the salivary microbiota composition with this smell assessment setup.

5.2 SALIVARY MICROBIOTA EXPOSOME SCAN

5.2.1 Background

The CHRIS study data collection includes more than 2328 phenotypes, ranging from bloodwork biochemical traits to lifestyle and neuropsychiatric questionnaires. To study the association between individual phenotypes and the salivary microbiota, it was necessary to develop a protocol to test as many variables as possible with the same statistical framework. This approach could confirm known associations, reveal previously overlooked ones, and open the field to future research directions.

5.2.2 Methods

We filtered the 2328 phenotypic variables present in CHRISMB according to two main technical criteria: (1) Less than 20% missing data in each variable, to ensure a sufficiently large sample size; (2) Factor variables should have between 2 and 15 levels to allow inclusion of municipality and occupation variables.

We filtered microbiota data for a minimum of 10 counts in at least 1% of samples and transformed it to relative abundance. We used the filtered microbiota data to generate a Bray-Curtis dissimilarity matrix with *phyloseq::distance*, which we modeled with PERMANOVA [173,225] with *vegan::adons2*. We set the “by” and “permutations” parameters to NULL and 2,000, respectively. The NULL method tests the overall significance of the variables in the model. We performed one PERMANOVA per trait, adjusting P-values with *stats::p.adjust* using the BH method (FDR 5%).

We considered variables with Q-value < 0.1 and $R^2 \geq 0.5\%$ as significant, removing variables correlated variables, for instance age, rounded age, and birth date.

5.2.3 Results

After filtering the phenotypic variables (see Methods), we retained 1,454 out of 2,328 variables. Of those, the PERMANOVA resulted in 62 significant variables with BH Q-value < 0.1 and at least 0.5% microbiota variance explained (R^2) (**Figure 5-2**). We observed that the variables with a considerable proportion of variance explained (R^2 > 1%) were related to smoking, sample processing batch, number of teeth, and age, all covariates which we considered in previous analyses. Variables with R^2 between 0.5 and 1% belonged to questionnaire sections “Personal Information” (level of education, employment status), “Anthropometrics” (BMI, blood pressure), municipality of residence, and 4 oral health questionnaire items (self-reported teeth health and gums health, difficulty biting food and frequency of dry mouth). Moreover, we found associations between beta diversity and food frequency questionnaire items related to foods high in simple sugars (sugar, soda, sweet beverages, sweetened juice, dried plums), followed by alcoholic drinks (spirits, wine, beer), fats (oils and butter), acidic foods and drinks (sodas, tomatoes, coffee), and fermented foods (beer, wine, bread, yoghurt).

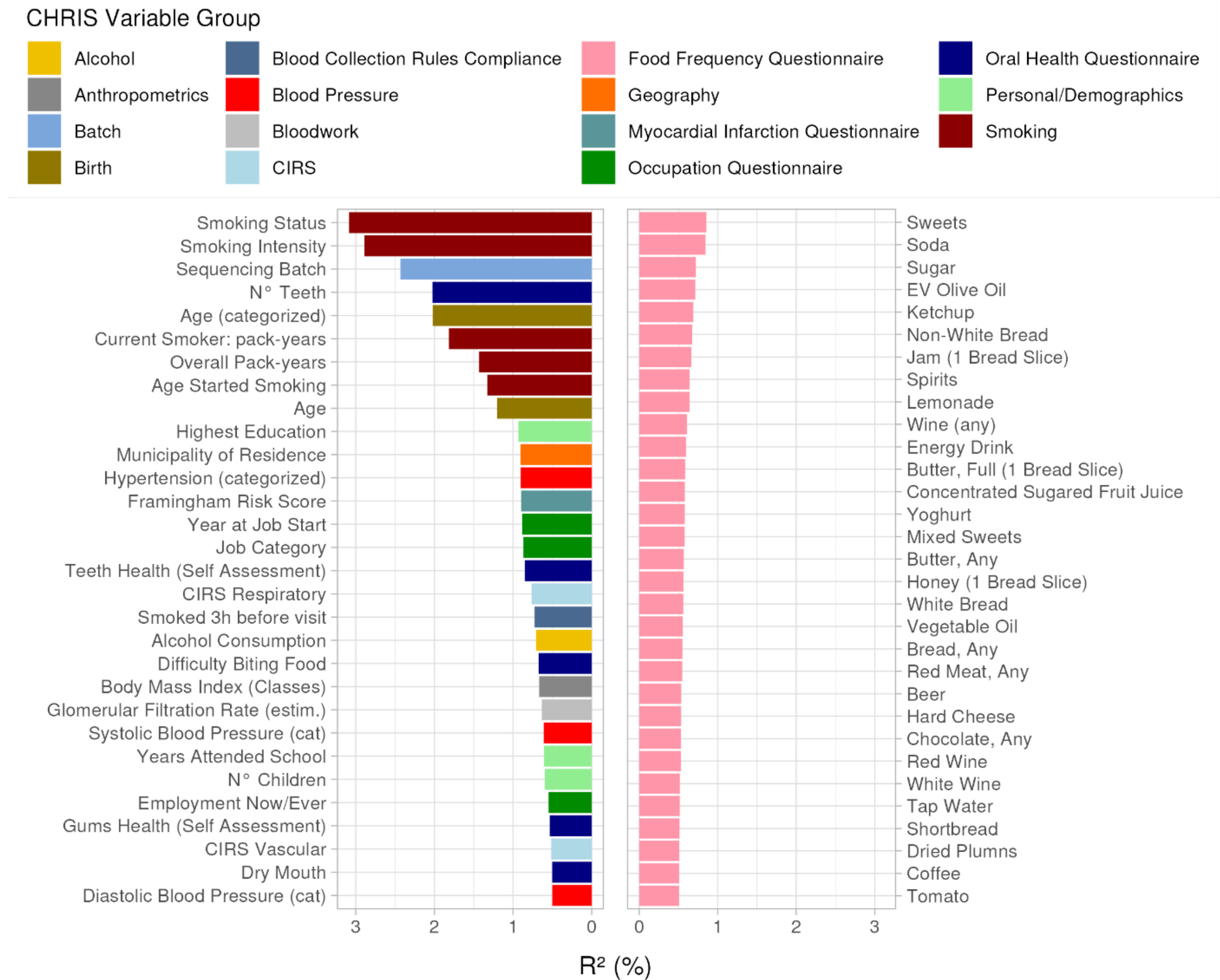


Figure 5-2. The CHRISMB exposome analysis revealed several lifestyle and demographic factors associated with salivary microbiota variability, and minor involvement of the diet. Variables were tested one by one with the PERMANOVA framework modeling the Bray-Curtis dissimilarity matrix as information of microbiota variability. The variables were split into food frequency questionnaire-related (right panel) and the rest (left panel). Significantly associated variables were considered with minimum R^2 (0.5%), and Benjamini-Hochberg adjusted Q-value less than 0.1.

5.2.4 Conclusions

We found several lifestyle and dietary factors influencing the salivary microbiota composition in the largest exposome scan for the salivary microbiota. Several other questionnaires – such as disease and neuropsychiatric questionnaires, and plasma biochemical and immunity markers were not associated. These results could facilitate researchers in testing new hypotheses and considering which variables influence the microbiota when modeling or designing experiments.

6 GENERAL CONCLUSIONS

In my thesis I characterized the salivary microbiota compositional diversity in CHRISMB, a sample of healthy Alpine Italian adults in Vinschgau/Val Venosta. With 1,923 samples, it is among the three largest single-cohort salivary microbiota studies in the world, between the Genome Center Cohort of the SPARK collection from New York, USA (N = 1929) and the 4D-SZ Shenzen cohort, China (N = 1915). Using this resource, I was able to study the impact of several lifestyle, environmental and genetic characteristics.

First, I explored the association between the salivary microbiota and smoking not only with a qualitative assessment (Never, Former, Current) but also with quantitative information regarding current smokers' cigarettes smoked per day and years since former smokers had quit. I found that the genera *Neisseria, Lautropia, Atopobium* and *Megaspaera* were significantly associated with smoking in CHRISMB and in studies of similar setup, showing a generalized effect of smoking. Additionally, I extended previous findings regarding the tobacco consumption per day, showing that heavier smokers had more marked compositional changes in both their compositional and predicted metabolic profile. The same could be seen in former smokers: several markers were slightly varying with the years since quitting. I hypothesized a mechanism of mediation of the oral microbiota in smoking-associated risk of periodontal and cardiovascular diseases, which could be explored in the future.

Second, I investigated the association of geographic distribution, relatedness, and genetics with the salivary microbiota composition. We show that cohabitation strongly associated with a higher microbiota similarity, while municipality and altitude showed weak associations. Fourteen genera and 16 ASVs had a significant heritability component ($17.4\% \leq h^2 \leq 30.9\%$), while 14 genera and 37 ASVs showed a significant household component ($9.9\% \leq c^2 \leq 30.8\%$). Using mbGWAS, we found no study-wide significant (SWS) Genus-variant associations. Conversely, we found four SWS ASV-variant associations after adjusting for age, sex, smoking, number of teeth, antibiotics usage, and 10 genetic principal components: *Selenomonas* ASV 113 – rs9511156; *Aggregatibacter* ASV 90 – rs916234787 and rs535001228; *Prevotella* ASV 178 – rs72809470. Gene level analysis on rare variants (minor allele frequency < 1%) further identified *SRFBP1,* and *LOX* associated with *Selenomonas noxia* (P-value < 3.05×10^{-8}).

Additionally, I showed that several other factors could play a role in the salivary microbiota composition. CHRISMB was mostly shaped by life habits (smoking, and cohabitation), age and oral health (number of teeth), and to a lower extent by diet. Sample processing batch was also significantly impactful, which stressed the importance of carefully designing and reporting technical choices to facilitate reproducibility. Furthermore, sex and antibiotics usage within 3 months prior to the visit were statistically significant but with 0.2% of variance associated, further pointing to a careful consideration of which factor should be considered when conducting salivary microbiota research.

CHRISMB data may be further explored designing epidemiological analyses based on the statistical associations observed in the exposome scan. Causality of genetic variants associated with the microbiota could be investigated with mendelian randomization, provided that the robustness of the association we reported is confirmed in other mbGWAS experiments. Genetic variants linked with microbiota composition could be further investigated with mendelian randomization to define causality of such variants. The microbiota, including the oral microbiota, has been acclaimed as a milestone in the direction of personalized health. To reach this goal, it would be advisable to perform further longitudinal research. For instance, the microbiota varies at different ages, which reflects several physiological and behavioral changes. However, understanding the chain of causality is pivotal to predict the impact of an alteration to the human body system on the microbiota, and how that reflects on the system. In the long run, the objective is to develop strategies for early detection and balancing of diseases via lifestyle and microbiota targeted approaches. Microbiota investigations using DNA sequencing methods (16S, and Shotgun metagenomics) have been pivotal, but they provide information only under one perspective. As stated by Whipps et al. [8], the microbiome is the "theater of activity", meaning that *what they are doing* is more informative than *who they are*. To achieve this, longitudinal sampling should be coupled with meta-transcriptomics, proteomics, and metabolic profiling, to study the community's metabolic dynamics in response to stimuli.

In summary, CHRISMB highlights that several behavioral, environmental and health factors are associated with variability of the salivary microbiome. My main contributions to the field are in the direction of a mechanistic explanation of increased periodontal and cardiovascular risk in heavy smokers, the first salivary microbiota heritability scan and second microbiome GWAS, which produced novel associations. I additionally performed an exposome scan on more than 1,400 variables showing associations with several phenotypes which could be further investigated in the future.

7 Final Acknowledgements /Ringraziamenti Finali

My first thank you goes, deservedly, to **Christian** and **Nicola**, my patient supervisors. During these years, you have supported me with patience and kindness through thick and thin. I even got married during my PhD! It wouldn't have been possible without your support. So thank you Christian, thank you Nicola, it was a really meaningful journey to me. Thank you also to professors Olivier Jousson and Alessandro Romanel, who helped me adjust my work at the end of each year. Special thanks to Isabella, our PhD secretary for always being there with patient reminders and answers regarding the bureaucratic duties of a PhD student.

There are many **colleagues** I should thank, and the beautiful thing is that they are too many! So, thank you all from the Computational Metagenomics group at CIBIO and the Computational Genomics group – as well as the whole Institute for Biomedicine – at Eurac. A special thanks goes to the Betsy Foxman Lab for generating the microbiome data I analyzed and the fruitful scientific discussions.

I would like to thank my **friends** Francesco T., Francesco M., Sebastiano, Alex, Giovanni, Niccolò, Filips (aka Filipp)., Maciek, Elisa, Elena, e Riccardo. Together we accumulated uncountable hours of coffees, *aperitivi*, evenings, walks and phone calls. Sharing doubts, difficulties, successes, suggestions, or simply time, is always refreshing and helpful.

Un meritato grazie va agli amici delle "**Gite della domenica**" (Ig: @in_solitedomeniche): Martino, Erica, Stefano, Laura, Giordano, Alessia, Mariano, Anna, Elisabetta. Insieme abbiamo camminato e mangiato tanto in questi anni, rigorosamente in rapporto 1:8. In voi ho trovato persone di cuore grande e tanti gesti: Sono tanti a dire "dovrei" o "potrei", sono pochi quelli che si mettono realmente in gioco.

Natalia, è anche grazie a te se adesso sono arrivato qui. Con immensa pazienza hai ascoltato le mie crisi e mi hai aiutato a ridimensionare tante paranoie! Con la tua tenacia e doti di pianificatrice seriale siamo anche riusciti a sposarci nel mio ultimo anno di dottorato! Sono proprio fortunato. Grazie anche a **Martino, Giordano, Emilia, Catia e Maurizio**: Stando con voi sto scoprendo lati di me che avrei altrimenti ignorato, aiutandomi ad essere più onesto con me stesso.

Mamma Ornella, papà Mauro, e fratello Guglielmo. Difficile riassumere in due frasi quanto grato sia di avervi vicino. Le esperienze che ho fatto grazie a voi hanno contribuito a portarmi fino a qui. Spero di poterne fare ancora tante, insieme! Insieme a loro vorrei ringraziare **zii, zie, e cugini** che hanno creduto in me e mi hanno sempre sostenuto, da vicino o da lontano.

Grazie **nonna Terry**, che hai assecondato la mia passione per i camion delle immondizie, le tisane, le prugne secche, ma soprattutto gli integratori. Sono grato che tu possa essere qui a festeggiare con me anche questo traguardo. Un grazie anche ai **nonni** che non ci sono più. Mi piace pensare che siate qui a gioire con me!

One last thank you goes to **me**. During these years I have change a lot. Acknowledging my limits has done more for me than boosting my ego. Especially if admitting them was followed by a "let's work together about it" instead of "I am not good enough". Someone else could have probably done a better scientific job, but I am sure I did my best, and this makes it of the utmost value to me.

BIBLIOGRAPHY

1. Tan SY, Berman E. Robert Koch (1843-1910): father of microbiology and Nobel laureate. Singapore Med J. 2008;49: 854–855. Available: https://pubmed.ncbi.nlm.nih.gov/19037548/

2. Uzal FA, Freedman JC, Shrestha A, Theoret JR, Garcia J, Awad MM, et al. Towards an understanding of the role of *Clostridium perfringens* toxins in human and animal disease. Future Microbiol. 2014;9: 361–377. doi:10.2217/fmb.13.168

3. Piscitelli A, Cacciatore S, Ambrosio F, Ragozzino R, Pasquini FM, Incordino F, et al. Clostridium tetani infection in a geriatric patient: Do not let your guard off! Ann Geriatr Med Res. 2023;27: 269–273. doi:10.4235/agmr.23.0067

4. Hambleton P. Clostridium botulinum toxins: a general review of involvement in disease, structure, mode of action and preparation for clinical use. J Neurol. 1992;239: 16–20. doi:10.1007/bf00839205

5. Nagano Y, Itoh K, Honda K. The induction of Treg cells by gut-indigenous Clostridium. Curr Opin Immunol. 2012;24: 392–397. doi:10.1016/j.coi.2012.05.007

6. Guo P, Zhang K, Ma X, He P. Clostridium species as probiotics: potentials and challenges. J Anim Sci Biotechnol. 2020;11. doi:10.1186/s40104-019-0402-1

7. Wassenaar TM. Insights from 100 years of research with probiotic *E. coli*. Eur J Microbiol Immunol . 2016;6: 147–161. doi:10.1556/1886.2016.00029

8. J. M. Whipps, K. Lewis, R. C. Cooke. Mycoparasitism and plant disease control. Fungi in Biological Control Systems. Manchester University Press; 1988. pp. 161–187.

9. Deng Y-J, Wang SY. Synergistic growth in bacteria depends on substrate complexity. J Microbiol. 2016;54: 23–30. doi:10.1007/s12275-016-5461-9

10. Lukeš J, Stensvold CR, Jirků-Pomajbíková K, Wegener Parfrey L. Are human intestinal eukaryotes beneficial or commensals? PLoS Pathog. 2015;11: e1005039. doi:10.1371/journal.ppat.1005039

11. Richard ML, Sokol H. The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. Nat Rev Gastroenterol Hepatol. 2019;16. doi:10.1038/s41575-019-0121-2

12. Lurie-Weinberger MN, Gophna U. Archaea in and on the human body: Health implications and future directions. PLoS Pathog. 2015;11: e1004833. doi:10.1371/journal.ppat.1004833

13. Mirzaei MK, Maurice CF. Ménage à trois in the human gut: interactions between host, bacteria and phages. Nat Rev Microbiol. 2017;15: 397–408. doi:10.1038/nrmicro.2017.30

14. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 2016;14: e1002533. doi:10.1371/journal.pbio.1002533

15. Baquero F, Nombela C. The microbiome as a human organ. Clin Microbiol Infect. 2012;18: 2–4. doi:10.1111/j.1469-0691.2012.03916.x

16. Simon J-C, Marchesi JR, Mougel C, Selosse M-A. Host-microbiota interactions: from holobiont theory to analysis. Microbiome. 2019;7. doi:10.1186/s40168-019-0619-4

17. Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev. 2008;32: 723–735. doi:10.1111/j.1574-6976.2008.00123.x

18. Montenegro J, Armet AM, Willing BP, Deehan EC, Fassini PG, Mota JF, et al. Exploring the influence of gut microbiome on energy metabolism in humans. Adv Nutr. 2023;14: 840–857. doi:10.1016/j.advnut.2023.03.015

19. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun. 2019;10: 1–10. doi:10.1038/s41467-019-12476-z

20. Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. Cell. 2016;165: 1332–1345. doi:10.1016/j.cell.2016.05.041

21. Portincasa P, Bonfrate L, Vacca M, De Angelis M, Farella I, Lanza E, et al. Gut Microbiota and short chain fatty acids: Implications in glucose homeostasis. Int J Mol Sci. 2022;23: 1105. doi:10.3390/ijms23031105

22. Li Y, Zhu M, Liu Y, Luo B, Cui J, Huang L, et al. The oral microbiota and cardiometabolic health: A comprehensive review and emerging insights. Front Immunol. 2022;13. doi:10.3389/fimmu.2022.1010368

23. O'Mahony SM, Clarke G, Borre YE, Dinan TG, Cryan JF. Serotonin, tryptophan metabolism and the brain-gut-microbiome axis. Behav Brain Res. 2015;277: 32–48. doi:10.1016/j.bbr.2014.07.027

24. Kaur H, Bose C, Mande SS. Tryptophan metabolism by gut microbiome and gut-brain-axis: An in silico analysis. Front Neurosci. 2019;13. doi:10.3389/fnins.2019.01365

25. Steinert RE, Lee Y-K, Sybesma W. Vitamins for the gut microbiome. Trends Mol Med. 2020;26: 137–140. doi:10.1016/j.molmed.2019.11.005

26. Swer NM, Venkidesh BS, Murali TS, Mumbrekar KD. Gut microbiota-derived metabolites and their importance in neurological disorders. Mol Biol Rep. 2023;50: 1663–1675. doi:10.1007/s11033-022-08038-0

27. Arakawa K, Kawai Y, Iioka H, Tanioka M, Nishimura J, Kitazawa H, et al. Microbial community analysis of food-spoilage bacteria in commercial custard creams using culture-dependent and independent methods. J Dairy Sci. 2008;91: 2938–2946. doi:10.3168/jds.2007-0677

28. Hatti-Kaul R, Chen L, Dishisha T, Enshasy HE. Lactic acid bacteria: from starter cultures to producers of chemicals. FEMS Microbiol Lett. 2018;365. doi:10.1093/femsle/fny213

29. Kapinusova G, Lopez Marin MA, Uhlik O. Reaching unreachables: Obstacles and successes of microbial cultivation and their reasons. Front Microbiol. 2023;14. doi:10.3389/fmicb.2023.1089630

30. Scherlach K, Hertweck C. Mediators of mutualistic microbe–microbe interactions. Nat Prod Rep. 2018;35: 303–308. doi:10.1039/c7np00035a

31. Rosero-Chasoy G, Rodríguez-Jasso RM, Aguilar CN, Buitrón G, Chairez I, Ruiz HA. Microbial co-culturing strategies for the production high value compounds, a reliable framework towards sustainable biorefinery implementation – an overview. Bioresour Technol. 2021;321: 124458. doi:10.1016/j.biortech.2020.124458

32. Mitra M, Nguyen KM-A-K, Box TW, Gilpin JS, Hamby SR, Berry TL, et al. Isolation and characterization of a novel bacterial strain from a Tris-Acetate-Phosphate agar medium plate of the green micro-alga Chlamydomonas reinhardtii that can utilize common environmental pollutants as a carbon source. F1000Res. 2020;9: 656. doi:10.12688/f1000research.24680.1

33. Bandounas L, Wierckx NJP, de Winde JH, Ruijssenaars HJ. Isolation and characterization of novel bacterial strains exhibiting ligninolytic potential. BMC Biotechnol. 2011;11. doi:10.1186/1472-6750-11-94

34. Timm CM, Loomis K, Stone W, Mehoke T, Brensinger B, Pellicore M, et al. Isolation and characterization of diverse microbial representatives from the human skin microbiome. Microbiome. 2020;8. doi:10.1186/s40168-020-00831-y

35. Limoli DH, Whitfield GB, Kitao T, Ivey ML, Davis MR Jr, Grahl N, et al. *Pseudomonas aeruginosa* alginate overproduction promotes coexistence with *Staphylococcus aureus* in a model of cystic fibrosis respiratory infection. MBio. 2017;8. doi:10.1128/mbio.00186-17

36. Dubois T, Tremblay YDN, Hamiot A, Martin-Verstraete I, Deschamps J, Monot M, et al. A microbiota-generated bile salt induces biofilm formation in Clostridium difficile. NPJ Biofilms Microbiomes. 2019;5: 14. doi:10.1038/s41522-019-0087-4

37. Mackie AM, Hassan KA, Paulsen IT, Tetu SG. Biolog phenotype MicroArrays for phenotypic characterization of microbial cells. Methods in Molecular Biology. Totowa, NJ: Humana Press; 2014. pp. 123–130. doi:10.1007/978-1-62703-712-9_10

38. Shea A, Wolcott M, Daefler S, Rozak DA. Biolog Phenotype Microarrays. Microbial Systems Biology. Totowa, NJ: Humana Press; 2012. pp. 331–373. doi:10.1007/978-1-61779-827-6_12

39. Lagier J-C, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, et al. Culturing the human microbiota and culturomics. Nat Rev Microbiol. 2018;16: 540–550. doi:10.1038/s41579-018-0041-0

40. Zhang J, Liu Y-X, Guo X, Qin Y, Garrido-Oter R, Schulze-Lefert P, et al. High-throughput cultivation and identification of bacteria from the plant root microbiota. Nat Protoc. 2021;16: 988–1012. doi:10.1038/s41596-020-00444-7

41. Wilson MJ, Weightman AJ, Wade WG. Applications of molecular ecology in the characterization of uncultured microorganisms associated with human disease. Reviews and Research in Medical Microbiology. 1997;8: 91. Available: https://journals.lww.com/revmedmicrobiol/Abstract/1997/04000/Applications_of_molecular_ecology_in_the.5.aspx

42. Clarridge JE III. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17: 840–862. doi:10.1128/cmr.17.4.840-862.2004

43. Jay ZJ, Inskeep WP. The distribution, diversity, and importance of 16S rRNA gene introns in the order Thermoproteales. Biol Direct. 2015;10. doi:10.1186/s13062-015-0065-6

44. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. 1977;74: 5088–5090. doi:10.1073/pnas.74.11.5088

45. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991;173: 697–703. doi:10.1128/jb.173.2.697-703.1991

46. Hiergeist A, Ruelle J, Emler S, Gessner A. Reliability of species detection in 16S microbiome analysis: Comparison of five widely used pipelines and recommendations for a more standardized approach. PLoS One. 2023;18: e0280870. doi:10.1371/journal.pone.0280870

47. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. Proc Natl Acad Sci U S A. 2013;110: 12450–12455. doi:10.1073/pnas.1300833110

48. Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, et al. The human gut virome is highly diverse, stable, and individual specific. Cell Host Microbe. 2019;26: 527-541.e5. doi:10.1016/j.chom.2019.09.009

49. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol. 2005;71: 8966–8969. doi:10.1128/aem.71.12.8966-8969.2005

50. Sze MA, Schloss PD. The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. mSphere. 2019;4. doi:10.1128/msphere.00163-19

51. Tourlousse DM, Narita K, Miura T, Sakamoto M, Ohashi A, Shiina K, et al. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. Microbiome. 2021;9. doi:10.1186/s40168-021-01048-3

52. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016;469: 967–977. doi:10.1016/j.bbrc.2015.12.083

53. Ash C, Farrow JAE, Dorsch M, Stackebrandt E, Collins MD. Comparative Analysis of Bacillus anthracis, Bacillus cereus, and Related Species on the Basis of Reverse Transcriptase Sequencing of 16S rRNA. Int J Syst Bacteriol. 1991;41: 343–346. doi:10.1099/00207713-41-3-343

54. Brenner DJ, Fanning GR, Steigerwalt AG, Ørskov I, Ørskov F. Polynucleotide sequence relatedness among three groups of pathogenic *Escherichia coli* strains. Infect Immun. 1972;6: 308–315. doi:10.1128/iai.6.3.308-315.1972

55. Sonnenborn U. Escherichia coli strain Nissle 1917-from bench to bedside and back: History of a special Escherichia coli strain with probiotic properties. FEMS Microbiol Lett. 2016;363: 1–6. doi:10.1093/femsle/fnw212

56. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Front Microbiol. 2016;7. doi:10.3389/fmicb.2016.00459

57. Leonard SR, Mammel MK, Lacher DW, Elkins CA. Strain-level discrimination of shiga toxin-producing Escherichia coli in spinach using metagenomic sequencing. PLoS One. 2016;11: e0167870. doi:10.1371/journal.pone.0167870

58. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. 2020;38: 685–688. doi:10.1038/s41587-020-0548-6

59. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife. 2021;10: e65088. doi:10.7554/elife.65088

60. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome. 2022;10. doi:10.1186/s40168-021-01213-8

61. Cremers G, Jetten MSM, Op den Camp HJM, Lücker S. Metascan: METabolic analysis, SCreening and ANnotation of metagenomes. Front Bioinform. 2022;2. doi:10.3389/fbinf.2022.861505

62. Deo P, Deshmukh R. Oral microbiome: Unveiling the fundamentals. J Oral Maxillofac Pathol. 2019;23: 122. doi:10.4103/jomfp.jomfp_304_18

63. Li X, Liu Y, Yang X, Li C, Song Z. The oral Microbiota: Community composition, influencing factors, pathogenesis, and interventions. Front Microbiol. 2022;13. doi:10.3389/fmicb.2022.895537

64. Lebeer S, Ahannach S, Gehrmann T, Wittouck S, Eilers T, Oerlemans E, et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. Nat Microbiol. 2023;8: 2183–2195. doi:10.1038/s41564-023-01500-0

65. Reynoso-García J, Miranda-Santiago AE, Meléndez-Vázquez NM, Acosta-Pagán K, Sánchez-Rosado M, Díaz-Rivera J, et al. A complete guide to human microbiomes: Body niches, transmission, development, dysbiosis, and restoration. Front Syst Biol. 2022;2. doi:10.3389/fsysb.2022.951403

66. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr. 2012;96: 544–551. doi:10.3945/ajcn.112.037382

67. Abdill RJ, Adamowicz EM, Blekhman R. Public human microbiome data are dominated by highly developed countries. PLoS Biol. 2022;20: e3001536. doi:10.1371/journal.pbio.3001536

68. Glassner KL, Abraham BP, Quigley EMM. The microbiome and inflammatory bowel disease. J Allergy Clin Immunol. 2020;145: 16–27. doi:10.1016/j.jaci.2019.11.003

69. Porcari S, Benech N, Valles-Colomer M, Segata N, Gasbarrini A, Cammarota G, et al. Key determinants of success in fecal microbiota transplantation: From microbiome to clinic. Cell Host Microbe. 2023;31: 712–733. doi:10.1016/j.chom.2023.03.020

70. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med. 2019;25: 667–678. doi:10.1038/s41591-019-0405-7

71. Chen L, Wang D, Garmaeva S, Kurilshikov A, Vich Vila A, Gacesa R, et al. The long-term genetic stability and individual specificity of the human gut microbiome. Cell. 2021;184: 2302-2315.e12. doi:10.1016/j.cell.2021.03.024

72. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8. doi:10.1186/s13073-016-0307-y

73. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449: 804–810. doi:10.1038/nature06244

74. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604: 732–739. doi:10.1038/s41586-022-04567-7

75. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. American gut: An open platform for citizen science microbiome research. mSystems. 2018;3. doi:10.1128/msystems.00031-18

76. Jie Z, Liang S, Ding Q, Li F, Tang S, Wang D, et al. A transomic cohort as a reference point for promoting a healthy human gut microbiome. Med Microecol. 2021;8: 100039. doi:10.1016/j.medmic.2021.100039

77. Lu J, Zhang L, Zhai Q, Zhao J, Zhang H, Lee Y-K, et al. Chinese gut microbiota and its associations with staple food type, ethnicity, and urbanization. NPJ Biofilms Microbiomes. 2021;7: 1–12. doi:10.1038/s41522-021-00245-0

78. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15: 382–392. doi:10.1016/j.chom.2014.02.005

79. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017;2. doi:10.1038/nmicrobiol.2017.4

80. Keku TO, Dulal S, Deveaux A, Jovov B, Han X. The gastrointestinal microbiota and colorectal cancer. Am J Physiol Gastrointest Liver Physiol. 2015;308: G351–G363. doi:10.1152/ajpgi.00360.2012

81. Liu B-N, Liu X-T, Liang Z-H, Wang J-H. Gut microbiota in obesity. World J Gastroenterol. 2021;27: 3837–3850. doi:10.3748/wjg.v27.i25.3837

82. John GK, Mullin GE. The gut microbiome and obesity. Curr Oncol Rep. 2016;18. doi:10.1007/s11912-016-0528-7

83. Del Chierico F, Rapini N, Deodati A, Matteoli MC, Cianfarani S, Putignani L. Pathophysiology of type 1 diabetes and gut Microbiota role. Int J Mol Sci. 2022;23: 14650. doi:10.3390/ijms232314650

84. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The Microbiota-gut-brain axis. Physiol Rev. 2019;99: 1877–2013. doi:10.1152/physrev.00018.2018

85. Lach G, Schellekens H, Dinan TG, Cryan JF. Anxiety, depression, and the microbiome: A role for gut peptides. Neurotherapeutics. 2018;15: 36–59. doi:10.1007/s13311-017-0585-0

86. Tierney BT, Yang Z, Luber JM, Beaudin M, Wibowo MC, Baek C, et al. The landscape of genetic content in the gut and oral human microbiome. Cell Host Microbe. 2019;26: 283-295.e8. doi:10.1016/j.chom.2019.07.008

87. Yang Y, Cai Q, Zheng W, Steinwandel M, Blot WJ, Shu X-O, et al. Oral microbiome and obesity in a large study of low-income and African-American populations. J Oral Microbiol. 2019;11. doi:10.1080/20002297.2019.1650597

88. Yang Y, Zheng W, Cai QY, Shrubsole MJ, Pei Z, Brucker R, et al. Cigarette smoking and oral microbiota in low-income and African-American populations. J Epidemiol Community Health. 2019;73: 1108–1115. doi:10.1136/jech-2019-212474

89. Sato N, Kakuta M, Uchino E, Hasegawa T, Kojima R, Kobayashi W, et al. The relationship between cigarette smoking and the tongue microbiome in an East Asian population. J Oral Microbiol. 2020;12: 1742527. doi:10.1080/20002297.2020.1742527

90. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol. 2012;13: R42. doi:10.1186/gb-2012-13-6-r42

91. DeGruttola AK, Low D, Mizoguchi A, Mizoguchi E. Current understanding of dysbiosis in disease in human and animal models. Inflamm Bowel Dis. 2016;22: 1137–1150. doi:10.1097/mib.0000000000000750

92. Tilg H, Cani PD, Mayer EA. Gut microbiome and liver diseases. Gut. 2016;65: 2035–2044. doi:10.1136/gutjnl-2016-312729

93. Vijay A, Valdes AM. Role of the gut microbiome in chronic diseases: a narrative review. Eur J Clin Nutr. 2022;76: 489–501. doi:10.1038/s41430-021-00991-6

94. Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun. 2017;8. doi:10.1038/s41467-017-00900-1

95. Fritz JV, Desai MS, Shah P, Schneider JG, Wilmes P. From meta-omics to causality: experimental models for human microbiome research. Microbiome. 2013;1: 14. doi:10.1186/2049-2618-1-14

96. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444: 1027–1031. doi:10.1038/nature05414

97. Li L, Li X, Zhong W, Yang M, Xu M, Sun Y, et al. Gut microbiota from colorectal cancer patients enhances the progression of intestinal adenoma in Apcmin/+ mice. EBioMedicine. 2019;48: 301–315. doi:10.1016/j.ebiom.2019.09.021

98. Tilg H, Adolph TE, Gerner RR, Moschen AR. The intestinal Microbiota in colorectal cancer. Cancer Cell. 2018;33: 954–964. doi:10.1016/j.ccell.2018.03.004

99. Wu J, Li Q, Fu X. Fusobacterium nucleatum contributes to the carcinogenesis of colorectal cancer by inducing inflammation and suppressing host immunity. Transl Oncol. 2019;12: 846–851. doi:10.1016/j.tranon.2019.03.003

100. Kelly JR, Borre Y, O' Brien C, Patterson E, El Aidy S, Deane J, et al. Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. J Psychiatr Res. 2016;82: 109–118. doi:10.1016/j.jpsychires.2016.07.019

101. Chinna Meyyappan A, Forth E, Wallace CJK, Milev R. Effect of fecal microbiota transplant on symptoms of psychiatric disorders: a systematic review. BMC Psychiatry. 2020;20. doi:10.1186/s12888-020-02654-5

102. de Stefano MC, Mazzanti B, Vespasiano F, Lombardini L, Cardillo M. The regulatory approach for faecal Microbiota transplantation as treatment for Clostridioides difficile infection in Italy. Antibiotics (Basel). 2022;11: 480. doi:10.3390/antibiotics11040480

103. Merrick B, Allen L, Masirah M Zain N, Forbes B, Shawcross DL, Goldenberg SD. Regulation, risk and safety of Faecal Microbiota Transplant. Infect Prev Pract. 2020;2: 100069. doi:10.1016/j.infpip.2020.100069

104. Maki KA, Kazmi N, Barb JJ, Ames N. The oral and gut bacterial microbiomes: Similarities, differences, and connections. Biol Res Nurs. 2021;23: 7–20. doi:10.1177/1099800420941606

105. Zaura E, Keijser BJF, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral microbial communities. BMC Microbiol. 2009;9: 259. doi:10.1186/1471-2180-9-259

106. Xu X, He J, Xue J, Wang Y, Li K, Zhang K, et al. Oral cavity contains distinct niches with dynamic microbial communities. Environ Microbiol. 2015;17: 699–710. doi:10.1111/1462-2920.12502

107. Mark Welch JL, Ramírez-Puebla ST, Borisy GG. Oral microbiome geography: Micron-scale habitat and niche. Cell Host Microbe. 2020;28: 160–168. doi:10.1016/j.chom.2020.07.009

108. Mark Welch JL, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci U S A. 2016;113: E791-800. doi:10.1073/pnas.1522149113

109. Zhang Y, Zhu C, Cao G, Zhan J, Feng X, Chen X. Dynamic alterations of oral Microbiota related to halitosis in preschool children. Front Cell Infect Microbiol. 2021;11. doi:10.3389/fcimb.2021.599467

110. Willis JR, Gabaldón T. The human oral microbiome in health and disease: From sequences to ecosystems. Microorganisms. 2020;8: 308. doi:10.3390/microorganisms8020308

111. Chhibber-Goel J, Singhal V, Bhowmik D, Vivek R, Parakh N, Bhargava B, et al. Linkages between oral commensal bacteria and atherosclerotic plaques in coronary artery disease patients. NPJ Biofilms Microbiomes. 2016;2. doi:10.1038/s41522-016-0009-7

112. Tonelli A, Lumngwena EN, Ntusi NAB. The oral microbiome in the pathophysiology of cardiovascular disease. Nat Rev Cardiol. 2023;20: 386–403. doi:10.1038/s41569-022-00825-3

113. Yuan M, Xu Y, Guo Z. Association of oral microbiome and pancreatic cancer: a systematic review and meta-analysis. Therap Adv Gastroenterol. 2022;15: 175628482211239. doi:10.1177/17562848221123980

114. Mo S, Ru H, Huang M, Cheng L, Mo X, Yan L. Oral-intestinal Microbiota in colorectal cancer: Inflammation and immunosuppression. J Inflamm Res. 2022;15: 747–759. doi:10.2147/jir.s344321

115. Costalonga M, Herzberg MC. The oral microbiome and the immunobiology of periodontal disease and caries. Immunol Lett. 2014;162: 22–38. doi:10.1016/j.imlet.2014.08.017

116. Kim HS, Kim CG, Kim WK, Kim K-A, Yoo J, Min BS, et al. Fusobacterium nucleatum induces a tumor microenvironment with diminished adaptive immunity against colorectal cancers. Front Cell Infect Microbiol. 2023;13: 1101291. doi:10.3389/fcimb.2023.1101291

117. van Houte J. Role of micro-organisms in caries etiology. J Dent Res. 1994;73: 672–681. doi:10.1177/00220345940730031301

118. Socransky SS, Haffajee AD. Periodontal microbial ecology. Periodontol 2000. 2005;38: 135–187. doi:10.1111/j.1600-0757.2005.00107.x

119. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the human oral microbiome during health and disease. MBio. 2014;5. doi:10.1128/mbio.01012-14

120. Blostein F, Bhaumik D, Davis E, Salzman E, Shedden K, Duhaime M, et al. Evaluating the ecological hypothesis: early life salivary microbiome assembly predicts dental caries in a longitudinal case-control study. Microbiome. 2022;10. doi:10.1186/s40168-022-01442-5

121. Walker AW, Hoyles L. Human microbiome myths and misconceptions. Nat Microbiol. 2023;8: 1392–1396. doi:10.1038/s41564-023-01426-7

122. Siddiqui R, Badran Z, Boghossian A, Alharbi AM, Alfahemi H, Khan NA. The increasing importance of the oral microbiome in periodontal health and disease. Future Sci OA. 2023;9. doi:10.2144/fsoa-2023-0062

123. Tian N, Faller L, Leffler DA, Kelly CP, Hansen J, Bosch JA, et al. Salivary gluten degradation and oral microbial profiles in healthy individuals and celiac disease patients. Appl Environ Microbiol. 2017;83. doi:10.1128/aem.03330-16

124. Mortazavi S, Akhlaghi N. Salivary Streptococcus mutans and Lactobacilli levels following probiotic cheese consumption in adults: A double blind randomized clinical trial(*). J Res Med Sci. 2012;17: 57–66. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3523439/

125. İnce G, Gürsoy H, İpçi ŞD, Cakar G, Emekli-Alturfan E, Yılmaz S. Clinical and biochemical evaluation of lozenges containing *Lactobacillus reuteri* as an adjunct to non - surgical periodontal therapy in chronic periodontitis. J Periodontol. 2015;86: 746–754. doi:10.1902/jop.2015.140612

126. Kuru BE, Laleman I, Yalnızoğlu T, Kuru L, Teughels W. The influence of a *Bifidobacterium animalis* probiotic on gingival health: A randomized controlled clinical trial. J Periodontol. 2017;88: 1115–1123. doi:10.1902/jop.2017.170213

127. Kumar S, Madurantakam P. Limited evidence shows short-term benefit of probiotics when used as an adjunct to scaling and root planing in the treatment of chronic periodontitis. Evid Based Dent. 2017;18: 109–110. doi:10.1038/sj.ebd.6401270

128. Xiao H, Fan Y, Li Y, Dong J, Zhang S, Wang B, et al. Oral microbiota transplantation fights against head and neck radiotherapy-induced oral mucositis in mice. Comput Struct Biotechnol J. 2021;19: 5898–5910. doi:10.1016/j.csbj.2021.10.028

129. Pattaro C, Gögele M, Mascalzoni D, Melotti R, Schwienbacher C, De Grandi A, et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J Transl Med. 2015;13: 348. doi:10.1186/s12967-015-0704-9

130. North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. Circ Res. 2012;110: 1097–1108. doi:10.1161/circresaha.111.246876

131. Mozumdar A, Liguori G. Persistent increase of prevalence of metabolic syndrome among U.s. adults: NHANES III to NHANES 1999–2006. Diabetes Care. 2011;34: 216–219. doi:10.2337/dc10-0879

132. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3: 2011–2030. doi:10.1371/journal.pmed.0030442

133. Liu X, Tong X, Zhu J, Tian L, Jie Z, Zou Y, et al. Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. Cell Discov. 2021;7: 117. doi:10.1038/s41421-021-00356-0

134. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect of host genetics on the gut microbiome. Nat Genet. 2016;48: 1407–1412. doi:10.1038/ng.3663

135. Qin Y, Havulinna AS, Liu Y, Jousilahti P, Ritchie SC, Tokolyi A, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat Genet. 2022;54: 134–142. doi:10.1038/s41588-021-00991-z

136. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53: 156–165. doi:10.1038/s41588-020-00763-1

137. Troelsen JT. Adult-type hypolactasia and regulation of lactase expression. Biochim Biophys Acta Gen Subj. 2005;1723: 19–32. doi:10.1016/j.bbagen.2005.02.003

138. González-Rodríguez I, Gaspar P, Sánchez B, Gueimonde M, Margolles A, Neves AR. Catabolism of Glucose and Lactose in Bifidobacterium animalis subsp. lactis, Studied by ^{13} C Nuclear Magnetic Resonance. Appl Environ Microbiol. 2013;79: 7628–7638. doi:10.1128/aem.02529-13

139. Liu X, Tang S, Zhong H, Tong X, Jie Z, Ding Q, et al. A genome-wide association study for gut metagenome in Chinese adults illuminates complex diseases. Cell Discov. 2021;7: 1–15. doi:10.1038/s41421-020-00239-w

140. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491: 56–65. doi:10.1038/nature11632

141. Calafell F, Roubinet F, Ramírez-Soriano A, Saitou N, Bertranpetit J, Blancher A. Evolutionary dynamics of the human ABO gene. Hum Genet. 2008;124: 123–135. doi:10.1007/s00439-008-0530-8

142. Rees JL. Genetics of hair and skin color. Annu Rev Genet. 2003;37: 67–90. doi:10.1146/annurev.genet.37.110801.143233

143. Irvine AD, Christiano AM. Hair on a gene string: recent advances in understanding the molecular genetics of hair loss. Clin Exp Dermatol. 2001;26: 59–71. doi:10.1046/j.1365-2230.2001.00764.x

144. Williams TN, Thein SL. Sickle cell anemia and its phenotypes. Annu Rev Genomics Hum Genet. 2018;19: 113–147. doi:10.1146/annurev-genom-083117-021320

145. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet. 2015;16: 45–56. doi:10.1038/nrg3849

146. Conneally PM. Huntington disease: genetics and epidemiology. Am J Hum Genet. 1984;36: 506–526. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1684448/

147. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. Nature. 2022;610: 704–712. doi:10.1038/s41586-022-05275-y

148. Mackay TFC. The genetic architecture of quantitative traits. Annu Rev Genet. 2001;35: 303–339. doi:10.1146/annurev.genet.35.102401.090633

149. Fisher RA. XV.—the correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1919;52: 399–433. doi:10.1017/s0080456800012163

150. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet. 2008;9: 255–266. doi:10.1038/nrg2322

151. Lush JL. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. J Anim Sci. 1940;1940: 293–301. doi:10.2527/jas1940.19401293x

152. Blangero J, Diego VP, Dyer TD, Almeida M, Peralta J, Kent JW Jr, et al. A Kernel of Truth. Advances in Genetics. Elsevier; 2013. pp. 1–31. doi:10.1016/b978-0-12-407677-8.00001-4

153. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011

154. Vinkhuyzen AAE, Wray NR, Yang J, Goddard ME, Visscher PM. Estimation and partition of heritability in human populations using whole-genome analysis methods. Annu Rev Genet. 2013;47: 75–95. doi:10.1146/annurev-genet-111212-133258

155. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20. doi:10.1038/s41576-019-0127-1

156. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primers. 2021;1: 1–21. doi:10.1038/s43586-021-00056-9

157. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. BMC Res Notes. 2019;12. doi:10.1186/s13104-019-4137-z

158. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376: 44–53. doi:10.1126/science.abj6987

159. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81: 559–575. doi:10.1086/519795

160. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44: 821–824. doi:10.1038/ng.2310

161. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47: D1005–D1012. doi:10.1093/nar/gky1120

162. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32: 381–385. doi:10.1002/gepi.20303

163. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005;437: 1299–1320. doi:10.1038/nature04226

164. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008;32: 227–234. doi:10.1002/gepi.20297

165. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008;32: 361–369. doi:10.1002/gepi.20310

166. Yin R, Kwoh CK, Zheng J. Whole Genome Sequencing Analysis. Encyclopedia of Bioinformatics and Computational Biology. Elsevier; 2019. pp. 176–183. doi:10.1016/b978-0-12-809633-8.20095-2

167. Runheim H, Pettersson M, Hammarsjö A, Nordgren A, Henriksson M, Lindstrand A, et al. The cost-effectiveness of whole genome sequencing in neurodevelopmental disorders. Sci Rep. 2023;13: 1–8. doi:10.1038/s41598-023-33787-8

168. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009;10: 387–406. doi:10.1146/annurev.genom.9.081307.164242

169. Alper CA, Larsen CE. Pedigree-defined haplotypes and their applications to genetic studies. Methods in Molecular Biology. New York, NY: Springer New York; 2017. pp. 113–127. doi:10.1007/978-1-4939-6750-6_6

170. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51: 76–87. doi:10.1038/s41588-018-0286-6

171. König E, Rainer J, Hernandes VV, Paglia G, Del Greco M. F, Bottigliengo D, et al. Whole exome sequencing enhanced imputation identifies 85 metabolite associations in the alpine CHRIS cohort. Metabolites. 2022;12: 604. doi:10.3390/metabo12070604

172. Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome. 2020;8. doi:10.1186/s40168-020-00875-0

173. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online. Wiley; 2017. pp. 1–15. doi:10.1002/9781118445112.stat07841

174. Vineis P. Smoking and impact on health. Eur Respir Rev. 2008;17: 182–186. doi:10.1183/09059180.00011002

175. Müller HP. Rauchen oder parodontale Gesundheit. Das Gesundheitswesen. 2000;62: 400–408. doi:10.1055/s-2000-12585

176. Nociti FH, Casati MZ, Duarte PM. Current perspective of the impact of smoking on the progression and treatment of periodontitis. Periodontol 2000. 2015;67: 187–210. doi:10.1111/prd.12063

177. Ziukaite L, Slot DE, Loos BG, Coucke W, Van der Weijden GA. Family history of periodontal disease and prevalence of smoking status among adult periodontitis patients: a cross-sectional study. Int J Dent Hyg. 2016;15: e28–e34. doi:10.1111/idh.12224

178. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, et al. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res. 1988;48: 3282–3287. doi:10.1007/s10552-007-9026-4

179. Gillison ML, Zhang Q, Jordan R, Xiao W, Westra WH, Trotti A, et al. Tobacco Smoking and Increased Risk of Death and Progression for Patients With p16-Positive and p16-Negative Oropharyngeal Cancer. J Clin Oncol. 2012;30: 2102–2111. doi:10.1200/JCO.2011.38.4099

180. Van Camp G. Cardiovascular disease prevention. Acta Clinica Belgica: International Journal of Clinical and Laboratory Medicine. 2014;69: 407–411. doi:10.1179/2295333714Y.0000000069

181. Lira R, Åkerman S, Klinge B, Boström EA, Gustafsson A. Salivary microbial profiles in relation to age, periodontal, and systemic diseases. PLoS One. 2018;13: 1–14. doi:10.1371/journal.pone.0189374

182. Pereira LC, Nascimento JCR, Rêgo JMC, Canuto KM, Crespo-Lopez ME, Alvarez-Leite JI, et al. Apolipoprotein E, periodontal disease and the risk for atherosclerosis: a review. Arch Oral Biol. 2019;98: 204–212. doi:10.1016/j.archoralbio.2018.11.009

183. Wade WG. The oral microbiome in health and disease. Pharmacol Res. 2013;69: 137–143. doi:10.1016/j.phrs.2012.11.006

184. Mason MR, Preshaw PM, Nagaraja HN, Dabdoub SM, Rahman A, Kumar PS. The subgingival microbiome of clinically healthy current and never smokers. ISME J. 2015;9: 268–272. doi:10.1038/ismej.2014.114

185. Börnigen D, Ren B, Pickard R, Li J, Ozer E, Hartmann EM, et al. Alterations in oral bacterial communities are associated with risk factors for oral and oropharyngeal cancer. Sci Rep. 2017;7: 17686. doi:10.1038/s41598-017-17795-z

186. Huang C, Shi G. Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med. 2019;17: 225. doi:10.1186/s12967-019-1971-7

187. Pietiäinen M, Liljestrand JM, Kopra E, Pussinen PJ. Mediators between oral dysbiosis and cardiovascular diseases. Eur J Oral Sci. 2018;126 Suppl 1: 26–36. doi:10.1111/eos.12423

188. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J. 2016;10: 2435–2446. doi:10.1038/ismej.2016.37

189. Beghini F, Renson A, Zolnik CP, Geistlinger L, Usyk M, Moody TU, et al. Tobacco exposure associated with oral microbiota oxygen utilization in the New York City Health and Nutrition Examination Study. Ann Epidemiol. 2019;34: 18-25.e3. doi:10.1016/j.annepidem.2019.03.005

190. Vallès Y, Inman CK, Peters BA, Ali R, Wareth LA, Abdulle A, et al. Types of tobacco consumption and the oral microbiome in the United Arab Emirates Healthy Future (UAEHFS) Pilot Study. Sci Rep. 2018;8: 11327. doi:10.1038/s41598-018-29730-x

191. Jia Y-J, Liao Y, He Y-Q, Zheng M-Q, Tong X-T, Xue W-Q, et al. Association Between Oral Microbiota and Cigarette Smoking in the Chinese Population. Front Cell Infect Microbiol. 2021;11: 658203. doi:10.3389/fcimb.2021.658203

192. Poulsen CS, Nygaard N, Constancias F, Stankevic E, Kern T, Witte DR, et al. Association of general health and lifestyle factors with the salivary microbiota - Lessons learned from the ADDITION-PRO cohort. Front Cell Infect Microbiol. 2022;12: 1055117. doi:10.3389/fcimb.2022.1055117

193. Raizada MK, Joe B, Bryan NS, Chang EB, Dewhirst FE, Borisy GG, et al. Report of the National Heart, Lung, and Blood Institute Working Group on the Role of Microbiota in Blood Pressure Regulation: Current Status and Future Directions. Hypertension. 2017; 479–485. doi:10.1161/HYPERTENSIONAHA.117.09699

194. Koch CD, Gladwin MT, Freeman BA, Lundberg JO, Weitzberg E, Morris A. Enterosalivary nitrate metabolism and the microbiome: Intersection of microbial metabolism, nitric oxide and diet in cardiac and pulmonary vascular health. Free Radic Biol Med. 2017;105: 48–67. doi:10.1016/j.freeradbiomed.2016.12.015

195. Bryan NS, Tribble G, Angelov N. Oral Microbiome and Nitric Oxide: the Missing Link in the Management of Blood Pressure. Curr Hypertens Rep. 2017;19: 33. doi:10.1007/s11906-017-0725-2

196. Goh CE, Trinh P, Colombo PC, Genkinger JM, Mathema B, Uhlemann AC, et al. Association Between Nitrate-Reducing Oral Bacteria and Cardiometabolic Outcomes: Results From ORIGINS. J Am Heart Assoc. 2019;8: e013324. doi:10.1161/JAHA.119.013324

197. Rosier BT, Moya-Gonzalvez EM, Corell-Escuin P, Mira A. Isolation and characterization of nitrate-reducing bacteria as potential probiotics for oral and systemic health. Front Microbiol. 2020;11: 555465. doi:10.3389/fmicb.2020.555465

198. Burleigh MC, Liddle L, Monaghan C, Muggeridge DJ, Sculthorpe N, Butcher JP, et al. Salivary nitrite production is elevated in individuals with a higher abundance of oral nitrate-reducing bacteria. Free Radic Biol Med. 2018;120: 80–88. doi:10.1016/j.freeradbiomed.2018.03.023

199. Knight TM, Forman D, Al-Dabbagh SA, Doll R. Estimation of dietary intake of nitrate and nitrite in Great Britain. Food Chem Toxicol. 1987;25: 277–285. doi:10.1016/0278-6915(87)90123-2

200. Takeshita T, Kageyama S, Furuta M, Tsuboi H, Takeuchi K, Shibata Y, et al. Bacterial diversity in saliva and oral health-related conditions: The Hisayama Study. Sci Rep. 2016;6: 22164. doi:10.1038/srep22164

201. Pattaro C, Goegele M, Mascalzoni D, Melotti R, Schwienbacher C, De Grandi A, et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J Transl Med. 2015;13: 348. doi:10.1186/s12967-015-0704-9

202. Hyde ER, Andrade F, Vaksman Z, Parthasarathy K, Jiang H, Parthasarathy DK, et al. Metagenomic analysis of nitrate-reducing bacteria in the oral cavity: Implications for nitric oxide homeostasis. PLoS One. 2014;9: e88645. doi:10.1371/journal.pone.0088645

203. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. Genome Biol. 2020;21: 191. doi:10.1186/s13059-020-02104-1

204. Mons U, Müezzinler A, Gellert C, Schöttker B, Abnet CC, Bobak M, et al. Impact of smoking and smoking cessation on cardiovascular events and mortality among older adults: meta-analysis of individual participant data from prospective cohort studies of the CHANCES consortium. BMJ. 2015;350: h1551. doi:10.1136/bmj.h1551

205. Ernster L, Forsmark-Andrée P. Ubiquinol: an endogenous antioxidant in aerobic organisms. Clin Investig. 1993;71: S60-5. doi:10.1007/BF00226842

206. Tribble GD, Angelov N, Weltman R, Wang B-Y, Eswaran SV, Gay IC, et al. Frequency of tongue cleaning impacts the human tongue microbiome composition and enterosalivary circulation of nitrate. Front Cell Infect Microbiol. 2019;9: 434270. doi:10.3389/fcimb.2019.00039

207. Pignatelli P, Fabietti G, Ricci A, Piattelli A, Curia MC. How periodontal disease and presence of nitric oxide reducing oral bacteria can affect blood pressure. Int J Mol Sci. 2020;21: 1–14. doi:10.3390/ijms21207538

208. Preshaw PM. Mouthwash use and risk of diabetes. Br Dent J. 2018;225: 923–926. doi:10.1038/sj.bdj.2018.1020

209. Naseem KM. The role of nitric oxide in cardiovascular diseases. Mol Aspects Med. 2005;26: 33–65. doi:10.1016/j.mam.2004.09.003

210. Taddei S, Virdis A, Ghiadoni L, Salvetti G, Bernini G, Magagna A, et al. Age-related reduction of NO availability and oxidative stress in humans. Hypertension. 2001;38: 274–279. doi:10.1161/01.hyp.38.2.274

211. Ashley C T, Mayank A, Nathan S B. Nitric oxide and geriatrics: Implications in diagnostics and treatment of the elderly. J Geriatr Cardiol. 2012;8: 230–242. doi:10.3724/sp.j.1263.2011.00230

212. Maki KA, Ganesan SM, Meeks B, Farmer N, Kazmi N, Barb JJ, et al. The role of the oral microbiome in smoking-related cardiovascular risk: a review of the literature exploring mechanisms and pathways. J Transl Med. 2022;20: 584. doi:10.1186/s12967-022-03785-x

213. Gallucci G, Tartarone A, Lerose R, Lalinga AV, Capobianco AM. Cardiovascular risk of smoking and benefits of smoking cessation. J Thorac Dis. 2020;12: 3866–3876. doi:10.21037/jtd.2020.02.47

214. Warnakulasuriya S, Dietrich T, Bornstein MM, Casals Peidró E, Preshaw PM, Walter C, et al. Oral health risks of tobacco use and effects of cessation. Int Dent J. 2010;60: 7–30. doi:10.1922/IDJ_2532Warnakulasuriya24

215. Hill AB. The environment and disease: association or causation? 1965. J R Soc Med. 2015;108: 32–37. doi:10.1177/0141076814562718

216. World Health Organization. Oral Health Surveys: Basic Methods. World Health Organization; 2013. Available: https://play.google.com/store/books/details?id=8rEXDAAAQBAJ

217. Janson C, Anto J, Burney P, Chinn S, de Marco R, Heinrich J, et al. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. Eur Respir J. 2001;18: 598–611. doi:10.1183/09031936.01.00205801

218. Murgia F, Melotti R, Foco L, Gögele M, Meraviglia V, Motta B, et al. Effects of smoking status, history and intensity on heart rate variability in the general population: The CHRIS study. Aalto-Setala K, editor. PLoS One. 2019;14: e0215053. doi:10.1371/journal.pone.0215053

219. Chinn S, Jarvis D, Melotti R, Luczynska C, Ackermann-Liebrich U, Antó JM, et al. Smoking cessation, lung function, and weight gain: a follow-up study. Lancet. 2005;365: 1629–35; discussion 1600-1. doi:10.1016/S0140-6736(05)66511-7

220. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl Environ Microbiol. 2013;79: 5112–5120. doi:10.1128/AEM.01043-13

221. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13: 581–583. doi:10.1038/nmeth.3869

222. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2022. Available: https://www.r-project.org/

223. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8. doi:10.1371/journal.pone.0061217

224. Lahti L. SS. microbiome R package. 2012. Available: https://bioconductor.org/packages/release/bioc/html/microbiome.html

225. Anderson MJ. A new method for non - parametric multivariate analysis of variance. Austral Ecol. 2001;26: 32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x

226. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. Nat Commun. 2022;13: 342. doi:10.1038/s41467-022-28034-z

227. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 1–21. doi:10.1186/s13059-014-0550-8

228. Zhou H, He K, Chen J, Zhang X. LinDA: linear models for differential abundance analysis of microbiome compositional data. Genome Biol. 2022;23: 95. doi:10.1186/s13059-022-02655-5

229. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol. 2021;17: e1009442. doi:10.1371/journal.pcbi.1009442

230. Fernandes AD, Reid JNS, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2: 1–13. doi:10.1186/2049-2618-2-15

231. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020;11: 3514. doi:10.1038/s41467-020-17041-7

232. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57: 289–300. Available: http://www.jstor.org/stable/2346101

233. Dzidic M, Collado MC, Abrahamsson T, Artacho A, Stensson M, Jenmalm MC, et al. Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay. ISME J. 2018;12: 2292–2306. doi:10.1038/s41396-018-0204-z

234. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, et al. The human oral microbiome. J Bacteriol. 2010;192: 5002–5017. doi:10.1128/JB.00542-10

235. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. 2016;352: 560–564. doi:10.1126/science.aad3503

236. Valles-Colomer M, Blanco-Míguez A, Manghi P, Asnicar F, Dubois L, Golzato D, et al. The person-to-person transmission landscape of the gut and oral microbiomes. Nature. 2023;614: 125–135. doi:10.1038/s41586-022-05620-1

237. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med. 2018;24: 392–400. doi:10.1038/nm.4517

238. Kilian M, Chapple ILC, Hannig M, Marsh PD, Meuric V, Pedersen AML, et al. The oral microbiome - an update for oral healthcare professionals. Br Dent J. 2016;221: 657–666. doi:10.1038/sj.bdj.2016.865

239. Romanenko YH, Hryhorenko LV, Komskyi MP, Sribnyk PL, Sinkovska OO. Nitric oxide formation in the metabolism of nitrates in the oral cavity. ZAPOROZHYE MEDICAL JOURNAL. 2019;21: 685–690. doi:10.14739/2310-1210.2019.5.179472

240. Hampelska K, Jaworska MM, Babalska ZŁ, Karpiński TM. The role of oral Microbiota in intra-oral halitosis. J Clin Med. 2020;9: 2484. doi:10.3390/jcm9082484

241. Krasse B. The Vipeholm Dental Caries Study: recollections and reflections 50 years later. J Dent Res. 2001;80: 1785–1788. doi:10.1177/00220345010800090201

242. Suzuki N, Yoneda M, Hirofuji T. Mixed red-complex bacterial infection in periodontitis. Int J Dent. 2013;2013: 587279. doi:10.1155/2013/587279

243. Wilkinson JE, Franzosa EA, Everett C, Li C, HCMPH researchers and trainees, HCMPH investigators, et al. A framework for microbiome science in public health. Nat Med. 2021;27: 766–774. doi:10.1038/s41591-021-01258-0

244. Belstrøm D, Eiberg JM, Enevold C, Grande MA, Jensen CAJ, Skov L, et al. Salivary microbiota and inflammation‐related proteins in patients with psoriasis. Oral Dis. 2020;26: 677–687. doi:10.1111/odi.13277

245. Oliveira SG, Nishiyama RR, Trigo CAC, Mattos-Guaraldi AL, Dávila AMR, Jardim R, et al. Core of the saliva microbiome: an analysis of the MG-RAST data. BMC Oral Health. 2021;21. doi:10.1186/s12903-021-01719-5

246. Ogbanga N, Nelson A, Ghignone S, Voyron S, Lovisolo F, Sguazzi G, et al. The oral microbiome for geographic origin: An Italian study. Forensic Sci Int Genet. 2023;64: 102841. doi:10.1016/j.fsigen.2023.102841

247. Parati G, Agostoni P, Basnyat B, Bilo G, Brugger H, Coca A, et al. Clinical recommendations for high altitude exposure of individuals with pre-existing cardiovascular conditions: A joint statement by the European Society of Cardiology, the Council on Hypertension of the European Society of Cardiology, the European Society of Hypertension, the International Society of Mountain Medicine, the Italian Society of Hypertension and the Italian Society of Mountain Medicine. Eur Heart J. 2018;39: 1546–1554. doi:10.1093/eurheartj/ehx720

248. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998;62: 1198–1211. doi:10.1086/301844

249. Schönherr S, Schachtl-Riess J, Maio SD, Filosi M, Mark M, Lamina C, et al. Performing highly parallelized and reproducible GWAS analysis on biobank-scale data. bioRxiv. 2023. doi:10.1101/2023.08.08.552417

250. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021;53: 1097–1103. doi:10.1038/s41588-021-00870-7

251. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17. doi:10.1186/s13059-016-0974-4

252. Brito IL, Gurry T, Zhao S, Huang K, Young SK, Shea TP, et al. Transmission of human-associated microbiota along family and social networks. Nat Microbiol. 2019;4: 964–971. doi:10.1038/s41564-019-0409-6

253. Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. Microbiome. 2014;2. doi:10.1186/2049-2618-2-41

254. Raulo A, Allen BE, Troitsky T, Husby A, Firth JA, Coulson T, et al. Social networks strongly predict the gut microbiota of wild mice. ISME J. 2021;15: 2601–2613. doi:10.1038/s41396-021-00949-3

255. Pullman J, Beghini F, Alexander M, Vishnempet Shridhar S, Prinster D, Brito IL, et al. Detailed social network interactions and gut microbiome strain-sharing within isolated Honduras villages. bioRxiv. 2023. doi:10.1101/2023.04.06.535875

256. Müller F, Naharro M, Carlsson GE. What are the prevalence and incidence of tooth loss in the adult and elderly population in Europe? Clin Oral Implants Res. 2007;18: 2–14. doi:10.1111/j.1600-0501.2007.01459.x

257. Natto ZS, Aladmawy M, Alasqah M, Papas A. Factors contributing to tooth loss among the elderly: A cross sectional study. Singapore Dent J. 2014;35: 17–22. doi:10.1016/j.sdj.2014.11.002

258. Muhammad T, Srivastava S. Tooth loss and associated self-rated health and psychological and subjective wellbeing among community-dwelling older adults: A cross-sectional study in India. BMC Public Health. 2022;22. doi:10.1186/s12889-021-12457-2

259. Schwartz JL, Peña N, Kawar N, Zhang A, Callahan N, Robles SJ, et al. Old age and other factors associated with salivary microbiome variation. BMC Oral Health. 2021;21. doi:10.1186/s12903-021-01828-1

260. Gazdeck RK, Fruscione SR, Adami GR, Zhou Y, Cooper LF, Schwartz JL. Diversity of the oral microbiome between dentate and edentulous individuals. Oral Dis. 2019;25: 911–918. doi:10.1111/odi.13039

261. Shaw L, Ribeiro ALR, Levine AP, Pontikos N, Balloux F, Segal AW, et al. The human salivary microbiome is shaped by shared environment rather than genetics: Evidence from a large family of closely related individuals. MBio. 2017;8: 1–13. doi:10.1128/mBio.01237-17

262. Ruan X, Luo J, Zhang P, Howell K. The salivary microbiome shows a high prevalence of core bacterial members yet variability across human populations. NPJ Biofilms Microbiomes. 2022;8: 1–14. doi:10.1038/s41522-022-00343-7

263. Kapatral V, Ivanova N, Anderson I, Reznik G, Bhattacharyya A, Gardner WL, et al. Genome Analysis of *F. nucleatum sub spp vincentii* and Its Comparison With the Genome of *F. nucleatum* ATCC 25586. Genome Res. 2003;13: 1180–1189. doi:10.1101/gr.566003

264. Karpathy SE, Qin X, Gioia J, Jiang H, Liu Y, Petrosino JF, et al. Genome Sequence of Fusobacterium nucleatum Subspecies Polymorphum — a Genetically Tractable Fusobacterium. PLoS One. 2007;2: e659. doi:10.1371/journal.pone.0000659

265. AlShahrani I, Hosmani J, AlShahrani A, Togoo RA, Syed S, Yassin SM, et al. High altitude as a possible factor for dysbiosis of salivary microbiome in orthodontic patients. Arch Oral Biol. 2020;119: 104917. doi:10.1016/j.archoralbio.2020.104917

266. Liu F, Liang T, Zhang Z, Liu L, Li J, Dong W, et al. Effects of altitude on human oral microbes. AMB Express. 2021;11. doi:10.1186/s13568-021-01200-0

267. Corby PM, Bretz WA, Hart TC, Schork NJ, Wessel J, Lyons-Weiler J, et al. Heritability of oral microbial species in caries-active and caries-free twins. Twin Res Hum Genet. 2007;10: 821–828. doi:10.1375/twin.10.6.821

268. Demmitt BA, Corley RP, Huibregtse BM, Keller MC, Hewitt JK, McQueen MB, et al. Genetic influences on the human oral microbiome. BMC Genomics. 2017;18: 1–15. doi:10.1186/s12864-017-4008-8

269. Esberg A, Haworth S, Kuja-Halkola R, Magnusson PKE, Johansson I. Heritability of oral microbiota and immune responses to oral bacteria. Microorganisms. 2020;8: 1–20. doi:10.3390/microorganisms8081126

270. Gomez A, Espinoza JL, Harkins DM, Leong P, Saffery R, Bockmann M, et al. Host Genetic Control of the Oral Microbiome in Health and Disease. Cell Host Microbe. 2017;22: 269-278.e3. doi:10.1016/j.chom.2017.08.013

271. Rice TK, Sarzynski MA, Sung YJ, Argyropoulos G, Stütz AM, Teran-Garcia M, et al. Fine mapping of a QTL on chromosome 13 for submaximal exercise capacity training response: the HERITAGE Family Study. Eur J Appl Physiol. 2012;112: 2969–2978. doi:10.1007/s00421-011-2274-8

272. Hardin M, Cho MH, McDonald M-L, Wan E, Lomas DA, Coxson HO, et al. A genome-wide analysis of the response to inhaled β2-agonists in chronic obstructive pulmonary disease. Pharmacogenomics J. 2016;16: 326–335. doi:10.1038/tpj.2015.65

273. Gabellini N, Bortoluzzi S, Danieli GA, Carafoli E. The human SLC8A3 gene and the tissue-specific Na+/Ca2+ exchanger 3 isoforms. Gene. 2002;298: 1–7. doi:10.1016/s0378-1119(02)00982-4

274. Kanakkanthara A, Huntoon CJ, Hou X, Zhang M, Heinzen EP, O'Brien DR, et al. ZC3H18 specifically binds and activates the BRCA1 promoter to facilitate homologous recombination in ovarian cancer. Nat Commun. 2019;10. doi:10.1038/s41467-019-12610-x

275. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics. 2014;13: 397–406. doi:10.1074/mcp.m113.035600

276. Lee VS, Halabi CM, Hoffman EP, Carmichael N, Leshchiner I, Lian CG, et al. Loss of function mutation in LOX causes thoracic aortic aneurysm and dissection in humans. Proc Natl Acad Sci U S A. 2016;113: 8759–8764. doi:10.1073/pnas.1601442113

277. Schulz S, Reichert S, Grollmitz J, Friebe L, Kohnert M, Hofmann B, et al. The role of Saccharibacteria (TM7) in the subginival microbiome as a predictor for secondary cardiovascular events. Int J Cardiol. 2021;331: 255–261. doi:10.1016/j.ijcard.2021.01.054

278. Sanna S, Kurilshikov A, van der Graaf A, Fu J, Zhernakova A. Challenges and future directions for studying effects of host genetics on the gut microbiome. Nat Genet. 2022;54: 100–106. doi:10.1038/s41588-021-00983-z

279. Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR, et al. Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. Nat Genet. 2015;47: 921–925. doi:10.1038/ng.3340

280. Liu X, Tong X, Jie Z, Zhu J, Tian L, Sun Q, et al. Sex differences in the oral microbiome, host traits, and their causal relationships. iScience. 2023;26: 105839. doi:10.1016/j.isci.2022.105839

281. Bruijning M, Ayroles JF, Henry LP, Koskella B, Meyer KM, Metcalf CJE. Relative abundance data can misrepresent heritability of the microbiome. Microbiome. 2023;11. doi:10.1186/s40168-023-01669-w

282. Antonello G, Blostein F, Bhaumik D, Davis E, Gögele M, Melotti R, et al. Smoking and salivary microbiota: a cross-sectional analysis of an Italian alpine population. Sci Rep. 2023;13: 18904. doi:10.1038/s41598-023-42474-7

283. Blangero J, Williams JT, Almasy L. 12 Variance component methods for detecting complex trait loci. Advances in Genetics. Elsevier; 2001. pp. 151–181. doi:10.1016/s0065-2660(01)42021-9

284. Zaura E, Brandt BW, Teixeira de Mattos MJ, Buijs MJ, Caspers MPM, Rashid M-U, et al. Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. MBio. 2015;6: e01693-15. doi:10.1128/mBio.01693-15

285. Menon RK, Gomez A, Brandt BW, Leung YY, Gopinath D, Watt RM, et al. Long-term impact of oral surgery with or without amoxicillin on the oral microbiome-A prospective cohort study. Sci Rep. 2019;9. doi:10.1038/s41598-019-55056-3

286. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26: 2336–2337. doi:10.1093/bioinformatics/btq419

287. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011;43: 969–976. doi:10.1038/ng.940

288. Liu Y, Xie J. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. J Am Stat Assoc. 2020;115: 393–402. doi:10.1080/01621459.2018.1554485

289. Feeney E, O'Brien S, Scannell A, Markey A, Gibney ER. Genetic variation in taste perception: does it have a role in healthy eating? Proc Nutr Soc. 2011;70: 135–143. doi:10.1017/s0029665110003976

290. Gardner A, So PW, Carpenter GH. Intraoral microbial metabolism and association with host taste perception. J Dent Res. 2020;99: 739–745. doi:10.1177/0022034520917142

291. Logan DW. Do you smell what I smell? Genetic variation in olfactory perception. Biochem Soc Trans. 2014;42: 861–865. doi:10.1042/bst20140052

292. Koskinen K, Reichert JL, Hoier S, Schachenreiter J, Duller S, Moissl-Eichinger C, et al. The nasal microbiome mirrors and potentially shapes olfactory function. Sci Rep. 2018;8: 1296. doi:10.1038/s41598-018-19438-3