THE INFLUENCE OF HUMAN GENETICS AND ENVIRONMENT ON THE SALIVARY MICROBIOTA

by

SIMONE SOPHIE STAHRINGER

Diplom-Agr. Biol., Universität Hohenheim, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Molecular, Cellular and Developmental Biology

2013

This thesis entitled:

THE INFLUENCE OF HUMAN GENETICS AND ENVIRONMENT

ON THE SALIVARY MICROBIOTA

written by Simone Sophie Stahringer

has been approved for the Department of

Molecular, Cellular and Developmental Biology

_____

Norman Pace (Chair)

_____

Kenneth Krauter

Date_____

The final copy of this thesis has been examined by the signatories, and we

Find that both the content and the form meet acceptable presentation standards

Of scholarly work in the above mentioned discipline.

IRB protocol 0399.11

Stahringer, Simone S (Ph.D., Molecular, Cellular and Developmental Biology)

The influence of human genetics and environment on the salivary microbiota

Thesis directed by Professor Kenneth S. Krauter

Oral bacterial communities have an influence on human oral and systemic health. Variation in the presence and abundance of these communities between individuals has been demonstrated, but it is unclear what factors drive these variations. This dissertation analyzed the influence of host genetics, temporal changes, and environmental variables such as cohabitation and substance use, on the oral microbiota via a culture-independent approach.

A portion of the bacterial small ribosomal subunit (16S rRNA genes) were PCR amplified and subjected to 454 pyrosequencing from banked saliva samples derived from various studies of the Institute of Behavioral Genetics at the University of Colorado. Sequences were analyzed with QIIME, a software package for microbial analysis, in order to assess taxonomy and diversity differences. A commonly shared group of eight oral genera was identified. A longitudinal twin study design (age 12, 17 and 22) of 264 saliva samples obtained from 107 individuals revealed no differences between monozygotic (n=27) and dizygotic (n=18) twin pairs, which suggests a low influence of heritability. Intra-individual stability over two five year spans during adolescence and young adulthood was observed and twins were not significantly different from their co-twin during this time period. Cohabitation was a driving factor of microbiota similarity as individuals at age 22, following likely departure from shared environment, showed increased diversity compared to earlier time points when they lived together. Furthermore, several bacterial families changed significantly in abundance with age during adolescence. In addition, a second study of 210 individuals (aged 12 to 65) demonstrated an increasing dissimilarity based on age and smoking. The presence and abundance of a number of bacterial genera decreased with age and smoking, while potential oral pathogens increased.

Acknowledgment

First and foremost, I want to express my sincere gratitude to my adviser, Dr. Kenneth Krauter, for the opportunity to work on this project and his continuous support and guidance throughout my studies. Your encouragement, the numerous opportunities you provided, and the freedom you granted me helped me excel in my professional development in multiple ways.

Second, I would like to thank my thesis committee, Dr. Norman Pace, Dr. Robin Dowell, Dr. Matthew McQueen, Dr. Matthew Keller, and Dr. Rob Knight for their helpful suggestions and for serving on my dissertation committee. Furthermore, I would like to thank all individuals involved in my research at the Institute of Behavioral Genetics, Biochemistry, and MCDB.

I would also like to thank my fellow lab members, past and present, for their assistance and support over the years as well as the wonderful working atmosphere: especially Helen Marshall, Paula Roberts, Justin Eagles-Soukup, and Brittany Demmitt for their active help with my research and their cheerful attitude. Thank you for being my "American family".

Great thanks goes to all my friends inside and outside MCDB for sharing lunches, happiness and disappointments. I would also like to thank my family for their support in various way, believing in me, and allowing me to always pursue my dreams, even if that meant moving to another continent.

My deepest gratitude goes to my husband Jochen for his love and support. This year was quite eventful and I love you more than ever. Finally, I would like to thank my daughter Marie for being a constant source of motivation and joy. Without you, I would not have finished as quickly.

Contents

CHAPTER I 1
INTRODUCTION AND BACKGROUND 1
1.1 Introduction 1
1.2 Culture independent analyses of microbial communities 2
1.2.1 Historical aspects and theory 2
1.2.2 Additional analytical methods 5
1.2.3 16S/18S rRNA gene based microbial analysis 8
1.2.4 Technical aspects 9
1.2.5 Sequencing technologies and post-processing 13
1.2.6 Ecological diversity measures 15
1.2.7 Applications 17
1.3 Oral microbiology 18
1.3.1 Introduction 18
1.3.2 Biofilm 21
1.3.3 The oral microbiota in health and disease 32
1.3.4 Gingivitis and periodontitis 34
1.3.5 Caries 37
1.3.6 Oral and non-oral diseases associated with oral microbes 39
1.3.7 Comparison of saliva to other oral microbial habitats 42
1.4 Purpose of the study 45
CHAPTER II 47
SALIVARY MICROBIAL COMMUNITIES IN ADOLESCENT AND YOUNG ADULT TWINS 47
2.1 Introduction 47
2.1.1 The oral microbiota in children and young adults and routes of transmission 47
2.1.2 The effect of the host genotype on microbial communities 50
2.2 Sample description 53

2.2.1 Saliva collection and sample selection 53
2.2.2 Characterization of the dataset 54
2.3 Results 56
2.3.1 Sequencing results and taxonomic overview 56
2.3.2 A core microbiota in human saliva 57
2.3.3 Monozygotic twins are statistically not more similar to each other than dizygotic twins 60
2.3.4 Twins become more dissimilar when moving apart 61
2.3.5 Temporal intra-individual similarity during two 5 year spans 63
2.3.6 Age effects on the oral microbiota 65
2.3.7 The effects of obesity on the oral microbiota 69
CHAPTER III 70
SALIVARY MICROBIAL COMMUNITIES DIFFER BY AGE AND SMOKING BEHAVIOR 70
3.1 Introduction 70
3.1.1 The effect of environmental variables on human microbial communities 70
3.1.2 The effect of smoking 72
3.1.3 The effect of alcohol consumption 92
3.1.4 The effect of stimulants using the example of methamphetamine 100
3.2 Sample description 107
3.2.1 Overview 107
3.2.2 Phenotype definitions 107
3.2.3 Characterization of the dataset 112
3.3 Results 117
3.3.1 Sequencing results and taxonomic overview 117
3.3.2 Bacterial organisms of interest 119
3.3.3 Alpha diversity 121
3.3.4 Age increases the beta diversity of the oral microbiota 121
3.3.5 Smoking increases the beta diversity of the oral microbiota 125
3.3.6 Multiple linear regression reveals independent age and smoking effects 128

3.3.7 Changes in the core microbiota based on age and smoking status 130
3.3.8 Specific bacterial organisms respond to age and smoking status 132
3.3.9 Effect of alcohol and stimulants 137
CHAPTER IV 138
DISCUSSION 138
4.1 General remarks about the studies 138
4.2 The effect of host genetics and cohabitation on the oral microbiota 139
4.3 The effect of age on the oral microbiota 143
4.4 The effect of environmental variables on the oral microbiota on the example of drug consumption 145
CHAPTER V 150
OUTLOOK AND SUMMARY 150
5.1 Future of oral microbiology and microbiology in general 150
5.1.1 Understanding interactions 150
5.1.2 Diagnostics 154
5.1.3 Therapeutics 155
5.2 Significance of my findings 157
Bibliography 158
Appendix A: Universal methods 180
A.1 Molecular methods and 454 sequencing 180
A.2 Quality filtering, clustering and taxonomy assignment 181
A.3 Data analysis and statistical methods 183
A.3.1 Alpha and beta diversity calculations 183
A.3.2 Wilcoxon test 184
A.3.3 ANOVA 184
A.3.4 Pearson product-moment correlation 184
A.3.5 Bonferroni correction 185
A.3.6 Multiple linear regression 185
A.3.7 Two-part test 185

A.3.8 UPGMA - hierarchical clustering 186
Appendix B: The effect of sampling and extraction methods on the oral microbiota 187
B.1 Introduction 187
B.2 Sample description 188
B.2.1 Traditional DNA extraction protocol 190
B.3 Results 190
B.4 Discussion 198
Appendix C: Supplemental information 200
C.1 Questionnaire wording 200

List of Tables

Table 2.1: Demographic data of the twin dataset 56

Table 2.2: OTUs correlated with age during adolescence and young adulthood 67

Table 3.1: A selection of oral microbiology studies based on smoking status 88

Table 3.2: Demographic characterization of the drug dataset 114

Table 3.3: Results of multiple linear regression for PC1, PC2 and PC3 according to the model PC = age + gender + smoking + race + ε for the full dataset and the two stratified age groups individually (< 30 years, >30 years) 129

Table 3.4: Presence of bacterial genera based on occurrence in the full dataset and stratified into subgroups based on age and smoking status 131

Table 3.5: Genera influenced by smoking or age group. The two part test 133

List of Figures

Figure 1.1: Leeuwenhoek's figures of bacteria from the human mouth 3

Figure 1.2: Conservation and entropy along the bacterial 16S rRNA gene 4

Figure 1.3: Anatomical depictions of human oral sites 20

Figure 1.4: Maturation of an oral biofilm 24

Figure 1.5: Interactions in the oral bacterial biofilm 28

Figure 1.6: Confocal micrograph of an artificial biofilm consisting of three species 29

Figure 1.7: Dissection of salivary glands on the right side 44

Figure 2.1: Age distribution in the twin dataset 54

Figure 2.2: Phylum level variation between individual saliva samples 57

Figure 2.3: The "core" - Presence of bacterial genera based on occurrence in samples 58

Figure 2.4: Genetic effect on the salivary microbiome at the ages of 12, 17, and 22 years 60

Figure 2.5: Similarities based on age of sampling (ages 12, 17, and 22 years) of twins and the sample population 62

Figure 2.6: Time progression of individuals, their twins, and the sample population at ages 12, 17, and 22 64

Figure 2.7: A. Relative abundance of bacterial families Veillonellaceae and Actinomycetaceae and their negative and positive correlation with age from adolescence to early adulthood 66

Figure 2.8: Close clustering of children age 8-11 (blue) compared to adolescents age 15-17 (red) using PCoA of the unweighted UniFrac distance matrix 68

Figure 3.1: Average number of deaths attributable to cigarette smoking in the US (2000-2004) 75

Figure 3.2: Dental status of a 55 year old male smoker 77

Figure 3.3: Factors that increase risk of periodontitis with emphasize on smokers 91

Figure 3.4: Metabolism of ethanol and its implication in cancer 93

Figure 3.5: Structure, street names and medical uses of amphetamine and its analogs 100

Figure 3.6: Reduced brain activity after short-term and long-term methamphetamine abstinence 103

Figure 3.7: Photographs of the dental status of a 32-year old methamphetamine user 103

Figure 3.8: Reduced alpha diversity in samples prior to 2001 113

Figure 3.9: Smoking habits, alcohol consumption, and stimulant administration distribution 115

Figure 3.10: Overlapping drug consumption 117

Figure 3.11: Average phyla and genera distribution in full dataset, grouped by age, and smoking status 118

Figure 3.12: Heatmap of sequences which may correspond to the three members of the red complex involved in severe periodontitis (*Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*) 120

Figure 3.13: Uneven distribution of ages using PCoA of unweighted UniFrac distance matrix 122

Figure 3.14: Differences in beta diversity based on bimodal age distribution 123

Figure 3.15: Distribution of ages in subgroups under and over 30 years based on PCoA of unweighted UniFrac distance matrices 124

Figure 3.16: Nonsmokers and irregular smokers are more similar to themselves than regular smokers 125

Figure 3.17: Smoking effect is demonstrated in both age groups 127

Figure 3.18: Presence and non-zero abundances of the genera *Filifactor* and *Neisseria* change significantly with smoking in all age groups 134

Abbreviations

Bacterial ecology
SSU – small ribosomal subunit (16S and 18S rRNA)
OTU – organizational taxonomic unit
HMP – human microbiome project
TGGE – temperature gradient gel electrophoresis
DGGE – denaturing gradient gel electrophoresis
T-RFLP – terminal restriction fragment length polymorphisms

Human genetics
MZ – monozygotic (identical twins)
DZ – dizygotic (fraternal twins)
ID – identity
QTL- quantitative trait locus

Gene names, Immunology and Bacteriology
TLR – toll like receptor
PAMP – pathogen associated molecular pattern
LPS – lipopolysaccharide
IL – interleukin
ADH – alcohol dehydrogenase
ALDH2 – aldehyde dehydrogenase 2
CYP2E1 – cytochrome P450 2E1

Statistics
PCoA – principal coordinate analysis
PC – principal coordinate
SEM – standard error of the mean

CHAPTER I
INTRODUCTION AND BACKGROUND

1.1 Introduction

"By the means of Telescopes, there is nothing so far distant but may be represented to our view; and by the help of Microscopes, there is nothing so small as to escape our inquiry; hence there is a new visible World discovered to the understanding. By this means the Heavens are open'd and a vast number of new Stars and new Motions, and new Productions appear in them, to which all the ancient Astronomers were utterly strangers. By this the Earth it self, which lyes so neer to us, under our feet, shews quite a new thing to us, and in every little particle of its matter, we now behold almost as great a variety of Creatures, as we were able before to reckon up in the whole Universe itself."

Robert Hooke, 1665 (in the Preface of Micrographia) from Gest (2004, pg. 8)

Despite being more than 340 years old, this quote by Robert Hooke remains valuable today. With the availability of mass DNA sequencing technologies and therefore decreasing sequencing cost, we are starting to discover new worlds of microorganisms virtually everywhere on earth, often invisible to the unaided eye, but vast in numbers and biomass. The total cell number of bacteria and archaea is approximated at $5x\ 10^{30}$ cells and their carbon mass is estimated to be 60-100% of the global carbon in plants (Whitman et al. 1998). Their habitats range from extreme to familiar, classified from a human standpoint. This includes not only deep sea (Sogin et al. 2006) or extreme terrestrial environments, such as hypersaline microbial mats (Ley et al. 2006a), but also human associated environments closer to us such as soil (Lauber et al. 2009) or leaf surfaces (Redford et al. 2010). Going further, we discover new worlds within our own homes. Indoor environments such as showerheads (Feazel et al. 2009) and air (Robertson et al. 2013a) do not cease to surprise us. Ultimately, our human bodies are covered by an invisible microbial biofilm inside and out (Fierer et al. 2008; Costello et al. 2009; Frank et al. 2010; Qin et

al. 2010; Stahringer et al. 2012; Yatsunenko et al. 2012). We are starting to see a glimpse of the extensive diversity of bacteria, archaea, and eukaryotes that are currently not culturable, most only identified by parts of their 16S rRNA gene sequence (Rappe and Giovannoni 2003).

This dissertation, which consists of five chapters, focuses on the influence of human genetics and environment on the salivary human bacterial communities. The first chapter reviews culture independent analysis of microbial communities and oral microbiology. The second chapter presents salivary bacterial data obtained from a longitudinal twin study (Stahringer et al. 2012), while the third chapter will discuss results from a salivary bacterial study on individuals with and without consumption of tobacco smoke, alcohol, and stimulants. Chapter 4 discusses the results and Chapter 5 concludes the dissertation research with an outlook for extending the field.

1.2 Culture independent analyses of microbial communities

1.2.1 Historical aspects and theory

Microorganisms were observed for the first time in the late 17th century by Robert Hooke and Antoni van Leeuwenhoek (Gest 2004). Hooke discovered microscopic fungi in 1665, while Leeuwenhoek spotted bacteria for the first time in 1676 (Gest 2004).

One of the first questions to ask is what members comprise a particular microbial community. Leeuwenhoek described his observations of oral bacteria in a letter to the Royal Society and included drawings of oral bacteria he saw; see Figure 1.1 for a reprint of his original drawings of bacterial species in the human mouth.

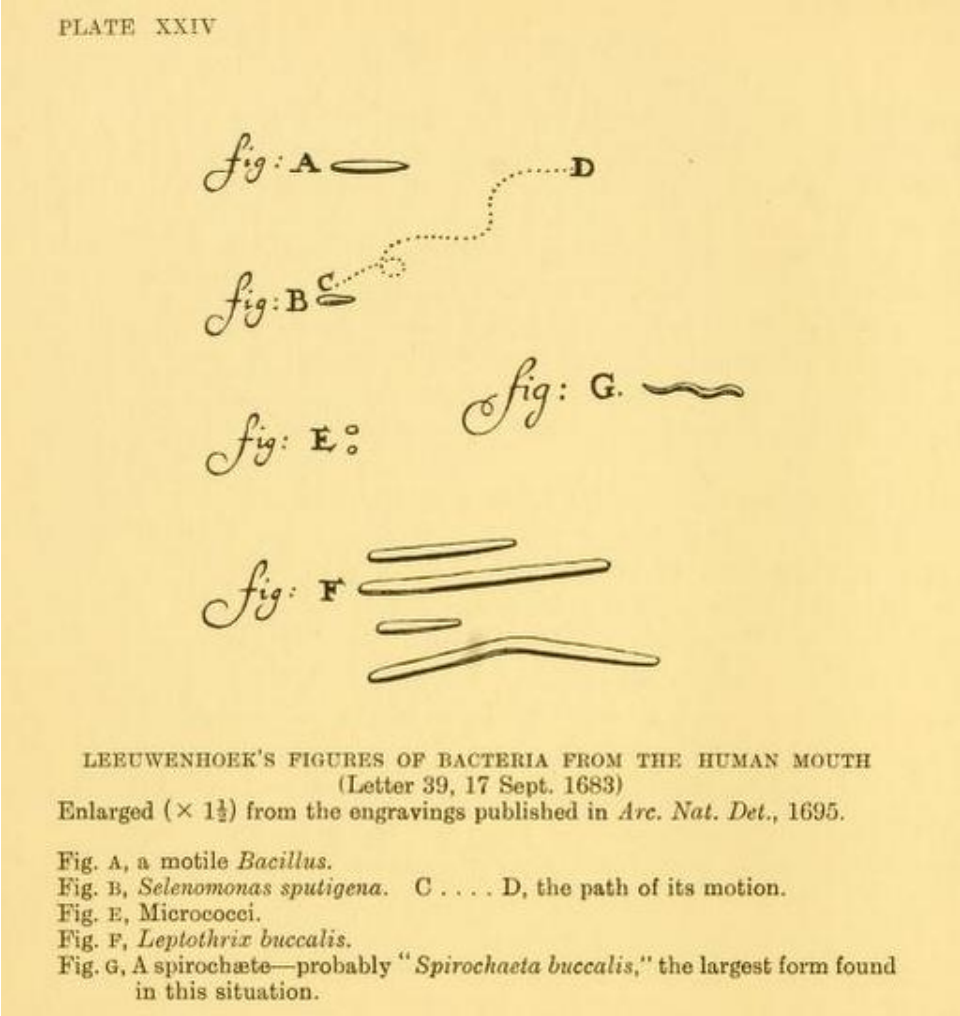


Figure 1.1: Leeuwenhoek's figures of bacteria from the human mouth from Dobell and Leeuwenhoek (1932, pg. 239).

From the mid 1800's until the second half of the 20th century, the only method for the study of microorganisms was isolation and culture where necessary and possible, and identification based on morphology, biochemistry, or both. However, the majority of microbes are not readily culturable.

A major discovery in phylogenetics came from Carl Woese. He discovered that archaea, which morphologically resemble bacteria, have indeed evolved independently from bacteria and eukaryotes by analyzing the sequence of the small ribosomal subunit (SSU) (Woese and Fox 1977). Due to the essential and direct function in protein synthesis as a functional RNA, the SSU

ribosomal RNA has properties that make it exceptionally useful for microbial taxonomy identification and evolutionary reconstruction. This molecule, which is part of all biological self-replicating systems, was chosen because it is highly conserved across all life, readily isolated, and evolves slowly (Woese and Fox 1977; Tringe and Hugenholtz 2008; Pace 2009). The small ribosomal subunit has conserved regions as well as variable regions as Figure 1.2 A and B depicts.

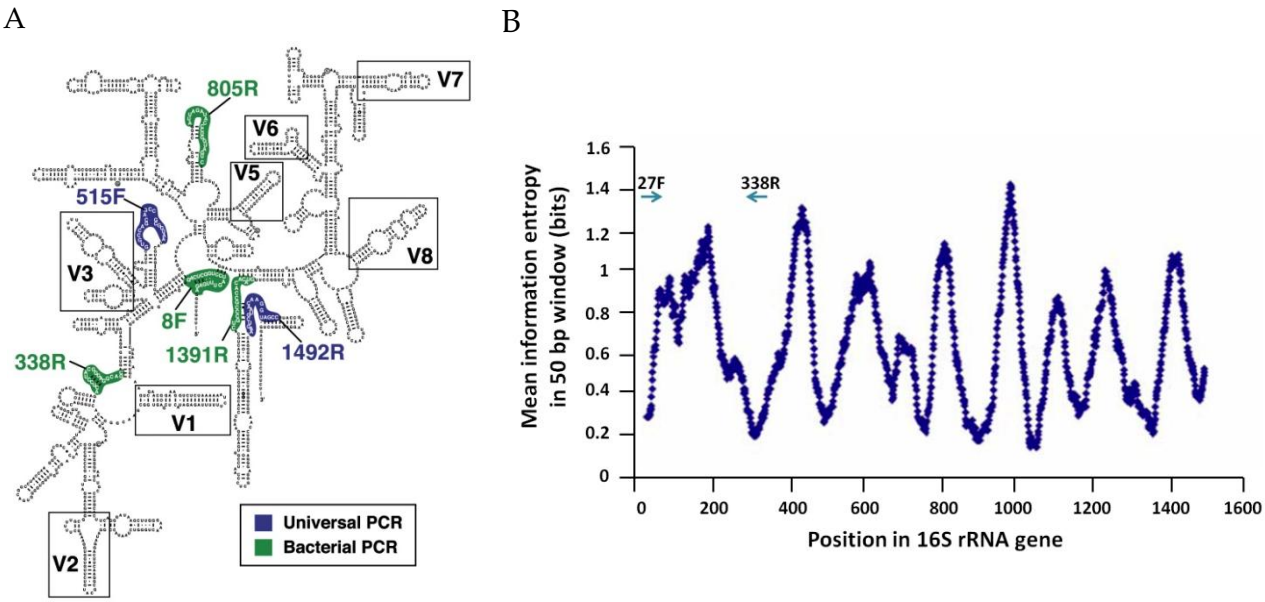


Figure 1.2: Conservation and entropy along the bacterial 16S rRNA gene. A) General secondary structure of the bacterial 16S rRNA gene. Numbers depict popular primer sites and correspond to the position in the E. coli SSU. A number of variable regions are boxed, from Peterson (2008). B) Linear representation of the variability within the 16S rRNA gene with primer sites of the primers used in my study (27F and 338R). The positions are approximate, modified from Andersson et al. (2008).

Gene sequence variations give us the opportunity to measure biological diversity directly and infer the structure of the tree of life (Pace 2009). While there are other possible genes which are useful for discrimination of closely related organisms, the SSU with its highly conserved sequences and structure due to its essential functional constraints, has been the gold

standard for phylogenetic tree reconstruction of microbial evolution and ecological community surveys (Tringe and Hugenholtz 2008; Pace 2009).

Fundamental developments in the field of molecular biology in the late 1970s and 1980s, Sanger sequencing (Sanger and Coulson 1975; Sanger et al. 1977) and PCR (Saiki et al. 1985), have proven to be essential for the field of microbial phylogeny and ecology. Besides direct sequencing of the SSU rRNA gene, which will be covered in Chapter 1.2.3, other techniques are used for culture dependent and independent analysis of microbial communities. I will list the most commonly used in oral microbiology; for more specific descriptions and applications please refer to the literature (Ogunseitan 2005, pg. 61 ff.; Asikainen and Karched 2008; Paster and Dewhirst 2009).

1.2.2 Additional analytical methods

Culture methods: Culture is an invaluable tool for studying properties of microorganisms. It can also aid in identification of bacterial colonies via selective media and colony properties such as texture, color, and growth rate. However, technical difficulties exist. The current biological knowledge is insufficient to provide the necessary culture conditions (i.e. media, nutrition, pH, air composition) for all organisms found in a particular environment (Stewart 2012). Furthermore, it is likely that many organisms require the metabolites or direct presence of other microorganisms for successful growth. In addition, pure cultures often behave very differently than the same organism in the context of a biofilm (see also Chapter 1.3.2). While a vast body of research in oral microbiology is based on culture, only an estimated 50% of oral bacteria can be cultured (Paster et al. 2001). In many other environmental samples, "the great plate count anomaly" (Staley and Konopka 1985) can be several orders of magnitude (Stewart 2012).

Microscopically assessing microbial diversity: Light microscopy allows for a direct assessment of the microbial gross morphology, aided by cell wall specific or other dyes if

necessary. To assess microbial density directly, the total number of microorganisms in a sample is counted. Staining with a DNA intercalating dye facilitates this process. Light microscopy also aids in estimating diversity based on bacterial morphology; for example, for identification of organisms with pathogenic potential in dental plaque communities. Viability is assessed with vital stains, a distinction which is impossible with most of the other methods presented below. Organism-specific fluorescently labeled oligonucleotides or antibodies allow for discrimination of known members in a community (Wilson 2005, pg. 29f.). In addition, electron microscopy is used to determine microbial community composition and arrangement. However, due to sample preparation it distorts the indigenous spatial organization.

Molecular and genomic methods: Besides sequencing the SSU genes, there is a vast array of other molecular and genomic methods for microbial identification and community comparison available. PCR utilizing taxonomic specific primers is able to detect and identify known microbes, whether through traditional PCR to test for presence or quantitative PCR to assess abundance. While these techniques are very powerful and specific, prior knowledge of the organisms of interest is necessary for specific primer design and does not allow for a survey approach without subsequent DNA sequencing.

Before inexpensive next generation sequencing technologies became available, the sequencing of 16S rRNA gene composition in an oral sample was a comparatively expensive approach to characterize microbial communities. Therefore, less costly methods have been developed to access oral microbial composition, many of which are based on PCR, in spite of its challenges. Examples of PCR based techniques include *Denaturing or temperature gradient gel electrophoresis (D/TGGE),* where an increasing chemical or temperature gradient is applied to denature double stranded DNA strands based on sequence along electrophoresis. It can detect small sequence differences within similar genes (Muyzer et al. 1993). *Terminal restriction fragment length polymorphism (T-RFLP)* utilizes restriction enzyme digestion display of 5' fluorescently labeled 16S rRNA gene amplicon fragments through electrophoretic resolution

(Liu et al. 1997). *PCR-RFLP* is similar to T-RFLP, but analyzes the size distribution of restriction digestion products of specific PCR products and allows for a finer resolution in less complex communities. *Ribosomal intergenic spacer analysis* allows distinguishing closely related organisms. Instead of amplifying and analyzing the SSU gene directly, the internal transcribed spacer (ITS) is examined which can differ in length and sequence (Baldwin 1992; Borneman and Triplett 1997). Further possibilities to distinguish closely related organisms include *multilocus sequence typing (MLST),* which compares the sequence of a number of housekeeping genes (Maiden et al. 1998).

Other techniques do not require PCR, but prior sequence information of the microorganisms of interest. For example, single stranded *DNA oligonucleotide probes* that hybridize to a target sequence, including *fluorescence in-situ hybridization (FISH),* can be used in microscopy as discussed above (DeLong et al. 1989). This technique has the advantage of the ability to detect species in their spatial organization in an intact biofilm. First developed in 1994 by Socransky and colleagues (Socransky et al. 1994), *whole-genomic checkerboard DNA-DNA hybridization* does not require PCR on the environmental sample, but cultured representatives of the organisms of interest. Whole-genomic bacterial DNA is labeled and hybridized to DNA of known and culturable species, which are bound to a nylon membrane. This technique has been used extensively in oral microbiology research by the Forsyth institute (Haffajee and Socransky 2001; Mager et al. 2003a; Socransky et al. 2004; Mager et al. 2005). Another twist on hybridization is *reverse-capture oligonucleotide hybridization* (also known as *reverse-capture checkerboard hybridization*), where instead of hybridizing with whole-genomic probes, the 16S rDNA of samples of interest is amplified via PCR first and then hybridized to species-specific immobilized 16S rRNA oligonuclotide probes (Paster et al. 1998). This technique has also been used in recent studies (Corby et al. 2005; Corby et al. 2007; Aas et al. 2008).

Microarrays with oral bacterial specific probes have been developed; for example, the HOMIM (Human Oral Microbe Identification Microarray), targeting 200 bacterial species

(Paster et al. 2006; Colombo et al. 2009; Colombo et al. 2012). A version of a microarray measuring gene expression in common oral bacteria (NIDCR Oral Microbial Microarray Initiative, NOMMI) had been used (Kolenbrander et al. 2010) but is no longer manufactured at the time of writing (in September 2013, http://www.nidcr.nih.gov/Research/DER/IntegrativeBiologyAndInfectiousDiseases/NOMMI.htm).

1.2.3 16S/18S rRNA gene based microbial analysis

Today, it is possible to examine all microbes present in the oral ecosystem based on their SSU (Jenkinson 2011). This development away from targeted analysis of microbial habitat is necessary due to various reasons including the inability of most microbes to grow in culture as discussed above and a shift in pathogenicity definitions in human medicine, especially dentistry. The one pathogen/one disease hypothesis in oral microbiology is no longer valid. Periodontitis, a gum disease, for example is a community disease and no one single organisms is consistently found in all patients with periodontitis (Bizzarro et al. 2013). The same observation has been made for tooth decay as *Streptococcus mutans* is not found on all caries lesions (Kanasi et al. 2010). An open approach such as 16S rRNA gene sequencing is more suitable to understand oral diseases than targeted approaches such as culture or qPCR for identification of a selected number of organisms. Bizzaro et al. even describes: "identification of a few targeted species for diagnostic purposes to be an out-of-date procedure" (Bizzarro et al. 2013).

In this dissertation, PCR with bacterial 16S rRNA specific primers amplifying variable regions 1 and 2 (27F/338R) was performed on human saliva samples and the PCR products were pyrosequenced with 454 FLX titanium.

1.2.4 Technical aspects

The sequence determination of the SSU genes in a sample allows the identification of microbial taxonomies and the estimation of a taxonomic distribution and therefore analyzing microbial diversity for virtually any habitat.

The standard steps are as following:

1. Whole community DNA is extracted.
2. A PCR with SSU specific primers is performed and the amplicons are prepared for sequencing. Depending on the sequence of the primers, the PCR can be amplifying representatives of all members of the tree of life including bacteria, archaea, and eukaryotes with so-called "universal" primers or it can be specific for taxa at various levels. The sequencing libraries are constructed, which includes cloning in the case of Sanger sequencing or adding sequencing adapters for many next generation sequencing technologies including 454 or Illumina. To facilitate the sample preparation, the sequencing adapters are often times part of the 16S primer sequence.
3. The DNA sequence of the amplicons is determined with one or more sequencing technologies (Sanger, 454, Illumina or others).
4. Taxonomy is assigned to the sequences and phylogenetic trees are built when necessary.

Advantages

As discussed above, 16S rRNA gene based analysis has many advantages.

To name a few:

- a culture independent, open ended approach to study most organisms or bacteria in an environment
- availability of extensive reference data bases (Pace 2009)
- small sample volume requirement, which allows the analysis of minute environmental samples or samples with low bacterial densities such as air (von Wintzingerode et al. 1997).

Limitations:

There are a number of potential problems and difficulties in 16S rRNA gene based studies, as well as conceptual limitations for interpretation of data obtained by this method:

1. Sample collection: Storage of specimen after sampling for a prolonged time could cause lysis of sensitive organisms, overgrowth of others or both (von Wintzingerode et al. 1997). However, experiments of human feces, human skin and soil samples also suggested that storage at various temperatures including room temperature for up 14 days did not influence the samples significantly (Lauber et al. 2010). This limitation is inherent to all microbiology research.

2. Errors from extraction of DNA: Cell lysis and the subsequent release of microbial DNA in aqueous solution is a critical step in the protocol. A complete lysis of cells is required without denaturing or mechanically shearing DNA. Insufficient or predominant lysis can distort the view of bacterial diversity in a given habitat. While insufficient lysis can cause an underrepresentation of affected phylotypes, overprocessing can cause misrepresentation through shearing of DNA or it could cause an excessive amount of chimeras (von Wintzingerode et al. 1997; Wilson 2005, pg. 33; Asikainen and Karched 2008, pg. 4).

3. Primer design and primer selection: A large number of possible primer pairs for SSU amplification of the whole or parts of the molecule are available. However, there is no continuous stretch of conserved bases long enough to allow the design of a truly universal bacterial primer with 100% match, let alone, for the whole tree of life. In addition, annealing temperatures and primer-primer complimentary need to be considered in primer design (Baker et al. 2003). PCR specific for the SSU genes in a community sample is by definition a PCR on a complex mixture of homologous templates. As a result, primers have a variable affinity for different taxonomic groups, which can lead to differential PCR amplification and alter the quantitative abundances (von Wintzingerode et al. 1997). Even very popular primers such as the primer pair 27F/338R, used in this study, do not cover bacterial phyla equally well and

cause an incomplete representation of most phyla as depicted in supplemental Figure 2 of Walters et al. (2011). Primer design software (PrimerProspector) developed at the University of Colorado after the initiation of this project recommends primer 515F and 806R which is almost universal to archaea and bacteria (Walters et al. 2011). A different approach to capture true diversity is to use a number of different primers on a sample and pool the results (Baker et al. 2003).

Therefore, abundances of specific phyla or higher taxonomic groups in a study including this are not necessarily a true representation of total distribution in the sample, but are relative to other samples of the same study. Comparisons of studies based on different primer pairs should be interpreted with caution.

4. PCR amplification: In addition to differences in primer binding, PCR can be influenced by PCR inhibitors. Many environmental samples contain *PCR inhibitors*; in oral samples blood is the most common PCR inhibitor (Abu Al-Soud and Radstrom 1998). Also, the number of rRNA gene copies varies by organisms and therefore, the number of observed sequences from a particular organism does not necessarily linearly scale to the number of cells of this organism, but to the number of cells multiplied by the number of gene copies (Klappenbach et al. 2001). Furthermore, random priming in early rounds of PCR can distort the abundance of individual taxa. Therefore, it is advised to limit the number of PCR rounds and pool three independent PCR reactions for analysis.

5. Chimeras: Most microbial diversity is known exclusively from observing ribosomal RNA sequences and not by culture (Rappe and Giovannoni 2003). Ashelford and colleagues estimated in 2005 that approximately 5% of sequences in databases are chimeras or have other substantial errors (Ashelford et al. 2005). These erroneous sequences can falsely inflate diversity. Chimeras are caused by in vitro recombination of homologous molecules (Haas et al. 2011). By definition, 16S rRNA sequences are homologous and therefore encourage chimera formation. Chimera formation decreases with increase in elongation time and reduction in cycle numbers

as chimeras are predominately caused by premature termination of the PCR product (Wang and Wang 1997; Ahn et al. 2012b). Multiple chimera detection programs have been developed such as Chimera Slayer (Haas et al. 2011) or UCHIME (Edgar et al. 2011). Chimeras and further artifact formation is discussed in a review by von Wintzingerode and colleagues (1997).

6. Contaminations: Additional problems with PCR are contaminations in buffers or the environment that will be amplified together with the actual sample, even from dead bacteria (von Wintzingerode et al. 1997). Clean working conditions, equipment, and buffers are essential. In addition, allochthonous (transient) species such as contaminations with food, e.g. Rhizobium from legumes, drinks, e.g. *Saccharomyces calsbergensis* in beer, or air can be detected in saliva samples along with the autochthonous (naturally occurring) species (Kolenbrander et al. 2010).

7. 16S relatedness does not necessarily equal functional relatedness: While the 16S rRNA gene is the gold standard of genomic phylogeny (Robinson et al. 2010), it is not necessarily reflective of the overall genomic similarity between microorganisms. For example, three *E. coli* strains (uropathogenic CFT073, enterohemorrhagic EDL933, and laboratory strain MG1655, derivative of K12), which are 99% identical based on 16S rRNA share only 39.2% of their proteins (Welch et al. 2002). This can be explained by horizontal gene transfer, which among others commonly transfers antibiotic resistance genes.

8. Other limitations: Other limitations which are inherent to the field are described in Wilson (2005, pg. 4) and include technical problems in obtaining representative samples, small sample numbers which limits reliability, different analysis methods used and secondary variation (such as sex, age, diet, hygiene, health, or occupation), and lastly, inconsistencies in bacterial nomenclature, which makes comparing studies difficult.

1.2.5 Sequencing technologies and post-processing

Sanger sequencing, also referred to as first generation sequencing, was the only sequencing technique for several decades. The up to 1000bp long reads are able to cover the majority of the 1.5kb 16S rRNA molecule giving the maximum of taxonomic information possible. It was not until 2004, with the introduction of 454 pyrosequencing (Margulies et al. 2005) that the whole sequencing field started to explode. One of the main advantages of next generation sequencing techniques is mere throughput which is several orders of magnitude higher than Sanger sequencing (Logares et al. 2012). This allows for identification of rare organisms that would stay undetected with low throughput methods such as Sanger. In addition, the increase in output allows for analyzing a higher number of individual samples, which could be longitudinal, spatial or by individual. This allows to increase statistical power to detect minor differences apart from random sampling variation (Siqueira et al. 2012).

Most next generation sequencing methods including 454 are sequence-by-synthesis techniques. The technical details have been exhaustively reviewed (for example Metzker 2010; Siqueira et al. 2012; Stranneheim and Lundeberg 2012). Read length of next generation sequencing technologies are typically shorter than Sanger is, but with technological progress, this gap is closing. As read length, sequences per run, and therefore data output is constantly improving, I refer to Glenn (2011) for an excellent summary on key characteristics in 2011 and manufacturers' websites for current updates.

At the time of writing this dissertation, three major sequencing technologies are routinely applied to 16S rRNA gene amplicons: Sanger sequencing, 454 pyrosequencing, and Illumina sequencing.

454 was chosen as the sequencing method of choice for this project. It provided a compromise of sequencing length and sequence reads per sample compared to less sequence reads and more taxonomic information with Sanger or more sequence reads and less taxonomic information with Illumina (then 50-100bp). At the time of data generation, approximately

500,000 sequencing reads of up to 400bp could be expected from one full 454 run for approximately $10,000. This reduces the taxonomic resolution compared to Sanger, but increases the number of sequences per sample and sample numbers tremendously. While a low sample number is sufficient to discriminate very different microbial habitats such as different body sites (Kuczynski et al. 2010), it might not be sufficient for other environments when differences between samples are minimal (Lemos et al. 2011).

Briefly, 454 pyrosequencing is based on a water-in-oil emulsion PCR and pyrosequencing. Nucleotide incorporation is measured by the amount of pyrophosphate generated through a reaction of pyrophosphate ATP sulfurylase, which is subsequently coupled to light emission via luciferase. Besides limitations in read length and other problems, which are common for all culture independent methods as discussed above, one drawback specific of 454 specifically is the inability to discriminate the length of stretches of homopolymers (Stranneheim and Lundeberg 2012). A study on human saliva has directly compared 454 sequencing to traditional Sanger sequencing and while there are differences, the overall distribution agrees with data derived from both methods with a significant correlation with an R2 of 0.7 within individuals (Nasidze et al. 2009b).

Post-sequencing processing methods

Binning (97% identity). Sequencing and PCR can introduce errors in the DNA sequence. Quality filtering and binning is essential with 454 data, but likely also for data generated with other sequencing technologies, as it can tremendously overestimate diversity in a dataset (Sogin et al. 2006; Huse et al. 2010). Kunin et al. have shown that sequencing parts of the 16S of one single *E. coli* strain can be interpreted as several hundred different "species" when no proper quality filtering and binning is applied (Kunin et al. 2010). "Species" in bacteria cannot be understood in an animal sense and its definition is highly debated (Lozupone and Knight 2008; Robinson et al. 2010); instead they comprise groups of taxonomic related

organisms. One should refer to *operational taxonomic units* (OTUs) rather than bacterial "species" (Lozupone and Knight 2008). While no definite recommendation of binning sequences into OTUs exists, many researchers have settled for a sequence identity of a minimum of 97% identity of the 16S rRNA gene which corresponds often to a degree of DNA-DNA hybridization of a minimum of 70% of the genome (Gevers et al. 2005). This threshold was also recommended based on the experiment mentioned above by Kunin (Kunin et al. 2010). However, there are distinct differences between lineages, so while this cutoff is often used, it is highly arbitrary and for example unsuitable to distinguish viridans streptococci (Kawamura et al. 1995). Members of organisms within an OTU typically show a shared core genome of 40% or more of their genes (Pace 2009). As next generation sequencing technologies such as pyrosequencing (454) or Illumina do not allow for full-length sequencing of the 16S rRNA gene to date (September 2013), the 16S rRNA gene is only partially sequenced. The amount of variability differs between different variable regions and therefore, caution should be applied to what cutoff to use (Schloss 2010). A cutoff of 97% of full length SSU rRNA does not necessarily correspond to a 97% cutoff of one or more isolated variable regions (Liu et al. 2007; Schloss 2010).

Taxonomy assignment. Taxonomy is usually assigned with a BLAST search assignment, the ribosomal database project (RDP) classifier or others. Popular reference databases are RDP (Cole et al. 2007), Greengenes (DeSantis et al. 2006; McDonald et al. 2012), SILVA (Quast et al. 2013), or oral specific curated databases are the human oral microbiome database (HOMD) (Chen et al. 2010) or CORE (Griffen et al. 2011). A summary of current (2012) bioinformatic tools can be found from Logares et al. (2012).

1.2.6 Ecological diversity measures

Diversity measures the variability of types; for example, microbes in a sample based on evenness (the distribution of various types) and richness (number of types). Key concepts are **alpha diversity** (also called local diversity) and **beta diversity**. Alpha diversity measures the

diversity for one particular environment, while beta diversity measures the change in diversity between different environments (Robinson et al. 2010). A number of different alpha and beta diversity metrics have been developed; I will briefly mention a selection including the ones used in this dissertation. The simplest alpha diversity metric is to rarefy the observed species (S_{obs}) in a particular environment at different sequencing depths and compare the curves for different environments. Other examples of classical alpha diversity metrics are Chao1 (Chao 1984) as a representative of a qualitative species based measure (based only on richness) and Shannon's index (Shannon and Weaver 1949) as a representative of a quantitative species based metric (based on richness and evenness). Beta diversity can also be measured qualitatively (e.g. Sörensen index) or quantitatively (e.g. Morisita-Horn index).

While these metrics do not account for taxonomic relatedness, divergence based metrics such as Phylogenetic Distance (PD) for alpha diversity and UniFrac for beta diversity consider the position of individual members of the community on a shared phylogenetic tree. PD measures the length of all branches in a shared phylogenetic tree that all members of one environment cover. UniFrac measures the difference between two environments by summing the **uni**que **frac**tion of branch length leading to the observed members of these two samples on a shared phylogenetic tree. UniFrac either can be unweighted (qualitative) or weighted, the latter accounting for quantitative differences. UniFrac values range from 0 (distance between identical communities) to 1 (distance between communities that do not share any branches and are therefore maximally different).

For further reading, see Magurran for a classical introduction to ecological diversity (Magurran 1988) and recent reviews on ecological diversity measures specifically for microbial studies (Lozupone and Knight 2008; Robinson et al. 2010).

1.2.7 Applications

As mentioned at the beginning, 16S rRNA gene based studies have been applied to a multitude of environments. The human microbiota have been found to be essential for animal function and development, including development of immune system, digestion, vitamin synthesis, resistance against pathogens, and more (Turnbaugh et al. 2007; Clemente et al. 2012). New evidence for the importance of our microbial cohabitants is constantly added, including microbial influence on behavior (Bercik et al. 2012; Cryan and Dinan 2012; Ezenwa et al. 2012), on obesity and malnutrition (Ley et al. 2005; Turnbaugh et al. 2006; Turnbaugh et al. 2009; Ridaura et al. 2013), and on type 1 and 2 diabetes (Giongo et al. 2011; Atkinson and Chervonsky 2012; Tremaroli and Backhed 2012). To assess the diversity and normal variation of the microbial communities on the human body, two enormous collaborative consortiums were established recently. In 2012, the 5 year human microbiome project (HMP) was completed. HMP was funded by the NIH to assess the diversity of 242 healthy humans across a number of body sites including oral cavity, skin, feces, and vagina for females. Nine specimens from the oral cavity and oropharynx were collected, which were saliva, keratinized gingiva (gums), palate, buccal mucosa (cheek), tonsils, throat, tongue soft tissues, as well as supra-and subgingival plaque (see Figure 1.3). A similar project funded by the European Union, is the Metagenomics of the Human Intestinal Tract (MetaHIT) project, with at least 124 individuals that exhibit variations in GI health and weight status (Qin et al. 2010).

After characterizing normal healthy variation in various human body habitats (Ursell et al. 2012) including the skin (Fierer et al. 2008; Grice et al. 2009), nose (Frank et al. 2010), vagina (Ravel et al. 2011), and gut (Qin et al. 2010) as well as studies covering multiple body habitats of the same individuals (Costello et al. 2009; Caporaso et al. 2011; Human Microbiome Project Consortium 2012), normal temporal variation needs to be studied (Caporaso et al. 2011; Koenig et al. 2011; Stahringer et al. 2012; Yatsunenko et al. 2012). The next logical step is to analyze changes of the microbial communities in reaction to changes in health or environment.

Examples are the changes of the vaginal microbiome in response to pregnancy (Aagaard et al. 2012) or the changes in the gut microbiome in response to dietary alterations (Wu et al. 2011). Also, correlations of disease states with an altered microbiota is important which can ultimately lead to a better understanding of the disease, development of diagnostics, preventions or even therapeutics. We need to understand our "second genome" (Zhu et al. 2010) better, as we are an amalgamation of our human self and our microbial communities.

The research in this dissertation focuses on the oral microbiome. I analyzed the bacterial composition of several hundred saliva samples. The oral cavity is a major point of entry to the human body and densely colonized with bacteria itself. Oral health is essential for overall health and quality of life (Gift and Redford 1992). The World Health Organization states that "oral health is integral to general health and is essential for well-being" (Petersen 2003). The next chapter introduces and reviews current knowledge on oral microbiology.

1.3 Oral microbiology

1.3.1 Introduction

Significance. The oral cavity harbors a diverse set of bacteria, some of which are responsible for diseases in the oral cavity and elsewhere in the human body. As mentioned before, oral health is a critical component for overall health and quality of life, a statement that has been recognized by the world health organization (Gift and Redford 1992; Petersen 2003). Pain, missing teeth, and xerostomia (dry mouth) are common problems associated with oral disease (Gift and Redford 1992). Furthermore, oral disease has an important socioeconomic impact. It can negatively influence visual appearance, lead to problems with speaking and eating. Food choices can be impacted by the inability to chew certain foods. In US children, dental disease and dental visits cause the loss of more than 51 million school hours annually. In employed adults, 164 million hours of work are lost in the US each year (U.S. Department of Health and Human Services 2000,pg. 2f.).

Oral diseases affect almost everybody at some point in their lifetime. Caries (tooth decay) is the most common chronic childhood disease and 85% of individuals 18 years or older are affected in the US (U.S. Department of Health and Human Services 2000, pg. 37, 63). Periodontitis prevalence increases with age, where 50% of middle aged individuals (age 55-64) experience medium attachment loss of 4mm or more and 19% have severe attachment loss of 6 mm or more (U.S. Department of Health and Human Services 2000, pg. 65). It has been suggested that modern lifestyle makes teeth more vulnerable as an increase in dental cavities, overcrowding of teeth, overbite, and gum disease shows compared to fossil and historic skulls and teeth (Gibbons 2012).

History. No description of the history of oral microbiology, or microbiology in general, would be complete without mentioning the fundamental work of Anthoni van Leeuwenhoek in the 17th century. Van Leeuwenhoek discovered bacteria around 1676. But he wasn't the first to observe microorganisms; Robert Hooke described microscopic fungi in 1665 (Gest 2004). Van Leeuwenhoek, who was not a formally trained scientist, described oral bacteria which he obtained from his teeth in a letter to the Royal Society in 1683. This has often been regarded as the beginning of oral microbiology (Bardell 1983; Gest 2004; He and Shi 2009). His drawings of "animalcules" (see Figure 1.1 in Chapter 1.2.1) found in dental plaque show common oral bacteria such as the spherical streptococci, rod shaped *Actinomyces* or *Fusobacterium* as well as the spiral formed Spirochetes in surprisingly accurate detail (Jenkinson 2011).

In a way, this thesis is the continuation of van Leeuwenhoek's work. He described in his letter that he performed anecdotal studies on the effect of gender, age, and even regular alcohol consumption and smoking; questions which this dissertation also attempts to address. In a biography about his work, one learns: "He also included an old man who smoked and frequently drank alcoholic beverages. [V]an Leeuwenhoek was quite curious as to whether or not little animals could live in a mouth which was frequently used to take in alcoholic drinks",

(Bardell 1983). Van Leeuwenhoek observed that a certain kind of spiral shaped organisms was missing from the daily alcohol consumer (Bardell 1983). Another milestone in oral microbiology history was reached in the 1890s, when Miller suspected a connection between tooth decay (caries) and oral bacteria, but it was not proven until the 1960s. The bacterial etiology for gum disease was not universally accepted until fairly recently in the mid 1980s (Jenkinson and Lamont 2005; He and Shi 2009). This and further landmark findings in oral microbiology are described in Chapters 1.3.4 on periodontitis and 1.3.5 on caries.

Anatomy. The oral cavity is a unique bacterial habitat in the human body. It unites soft shedding mucosal surfaces with the hard non-shedding surface of teeth. Mucosal surfaces include the gingiva (gums), buccal mucosa (inner cheeks), hard and soft palate, tongue, tonsils and throat (see Figure 1.3 A and B).

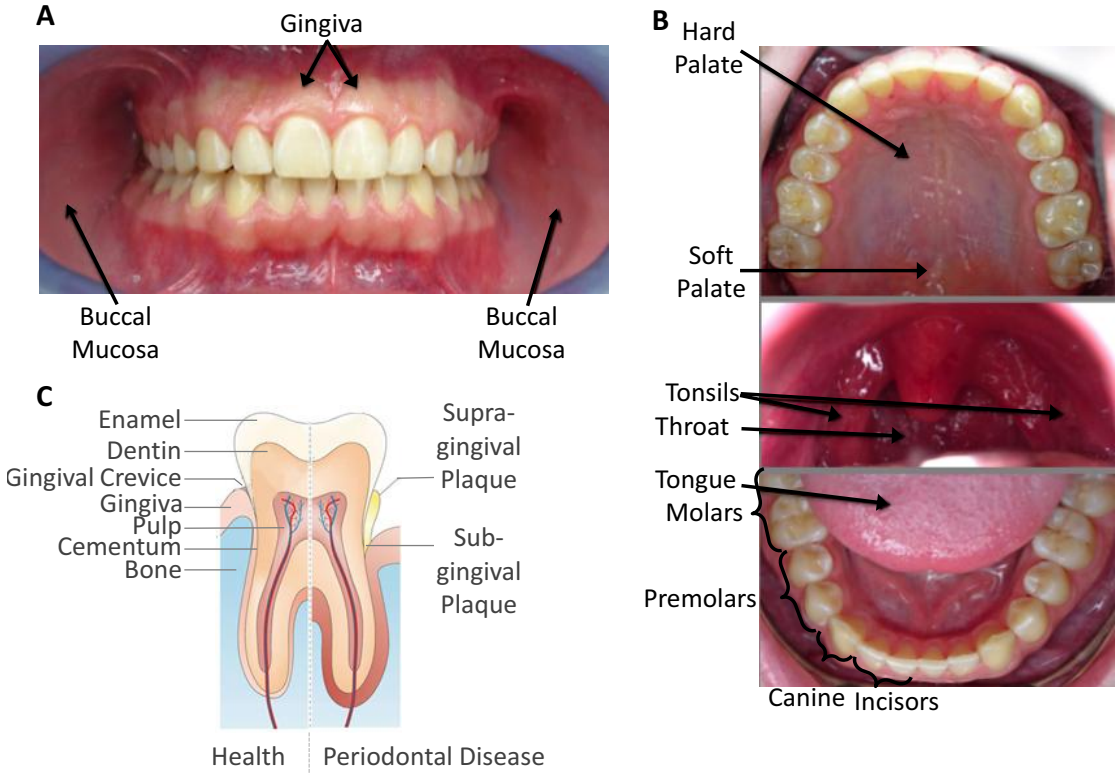


Figure 1.3: Anatomical depictions of human oral sites. A and B: Mucosal sites and dentition in the human mouth, modified from Segata et al. (2012). A: Front view with closed teeth. B: Top, straight and bottom view with open teeth. C: Cross-section of a human tooth surrounded by gingiva and bones. Left: Healthy, right: periodontal disease, modified from Darveau (2010).

The location between tooth and gingiva, known as gingival crevice, harbors plaque (see Figure 1.3 A and C). Plaque can be divided into subgingival plaque, which is below the gumline including periodontal pockets, and supragingival plaque, which is found above the gumline on the exposed enamel (Kolenbrander et al. 2010). The full dentition of an adult encompasses 32 permanent teeth. There are 16 teeth in the upper and lower jaw, with two incisors, one canine, two premolars, and three molars on each side (Figure 1.3 B). A tooth consists of four major tissues (from the outside to the inside): enamel, dentin, cementum, and pulp (Figure 1.3 C). Enamel is the outermost layer and consists mainly of hydroxylapatite, a crystalline calcium phosphate. It is the hardest substance in the human body. Dentin and cementum are softer structures and therefore more prone to tooth decay. The innermost tissue, pulp, contains blood vessels and nerves (Lamont and Jenkinson 2010, pg. 34f.).

1.3.2 Biofilm

As mentioned in the previous section, the oral cavity provides a unique microbial habitat compared to other human ecological niches as it is composed of hard, non-shedding surfaces of teeth as well as soft, shedding surfaces of various mucosal tissues. Furthermore, substantial mechanical forces are applied through chewing. Additional factors that distinguish the oral cavity from other human habitats are succession, regular natural disturbances (oral hygiene), and various tooth shapes. Comparable to other habitats such as the gut or vagina, interaction with host tissues and secretions are observed. The oral cavity provides a diverse, species rich ecological niche and polymicrobial populations are the norm, not the exception (Kolenbrander et al. 2010). The microbial communities are not homogenous within the oral cavity; instead, there are a large number of microhabitats on the various mucosal surfaces and different teeth surfaces. Saliva is the vehicle that transports bacteria within the oral cavity and constantly flushes the oral cavity to keep microbial loads low (Boonanantansarn and Gill 2011, pg. 23).

The mucosa acts as a protective layer to separate microbes and the body through a direct barrier function, shedding surfaces (Wilson 2005, pg. 39), and secreting mucins (Slomiany et al. 1996). While shedding, exposed surfaces such as cheek mucosa only allows the growth of microcolonies and are relatively sparingly colonized through constant desquamation (cell shedding), non-shedding surfaces (teeth) and mucosal surfaces with appropriate architecture such as the crypts on the tongue and tonsils allow for the formation of biofilm (Wilson 2005, pg. 7).

In 2005, it was estimated that about 50% of all oral bacteria are culturable (Aas 2005) and culture techniques are constantly developed which is important for metabolic and biochemical characterization. Oral microbiological research has identified a small number of potential pathogens implicated in oral diseases such as *Streptococcus mutans* in tooth decay, *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* in chronic and aggressive juvenile gingivitis. However, dental plaque and other oral biofilms are very complex (Kolenbrander et al. 2010). The "reductionism" to selected members of this complex community has gained a lot of insight in oral microbiology, but it is not able to explain the whole system and a "holism" approach is needed to understand the interactions between the different organisms within a healthy biofilm and polymicrobial diseases (He and Shi 2009). Approximately 700 different oral species have been described (Aas et al. 2005). The number of OTUs that have been found in various 16S rRNA gene surveys easily exceeds this number, for example up to 8000 in Lazarevic et al. (2009). However, the number of OTUs depends heavily on the amplified region within the 16S rRNA gene, the sequencing technology, the quality filtering settings, as well as the sequencing depth. Almost 700 oral species are described in the curated Human Oral Microbiome Database (Chen et al. 2010; Dewhirst et al. 2010).

Plaque is an excellent model for biofilm development and the best understood example of natural biofilm formation across disciplines (Kolenbrander et al. 2010; Kolenbrander 2011b, pg. 4). This description of colonization succession will focus on plaque, but it is likely that most

biofilms follow a similar, specific colonization pattern as has been shown in other habitats such as freshwater (Min and Rickard 2009). The oral biofilm is not a random accumulation of organisms, but a repeatable process of events to create a structured biofilm in which inter-individual variation exists (Diaz et al. 2006). Within minutes of dental hygiene, which has the purpose to reduce the load of bacterial organisms temporarily, the biofilm starts to reform. Saliva transports suspended bacterial cells to the freshly cleaned surfaces. Generally, earlier colonizers are more beneficial to oral health than later colonizers, which often include potential pathogenic species (Kolenbrander 2011b, pg. 3f.; Kolenbrander 2011a).

The properties of an organism growing in a biofilm are different from cells in planktonic phase. Bacteria in a biofilm exhibit an increased resilience and resistance against antimicrobial agents and host defense factors. A biofilm creates a new habitat with chemical and physiochemical gradients within and provides a well-structured, organized network of matrix with open water channels as well as fluid-filled voids (Wilson 2005, pg. 8; Marsh and Percival 2006; He and Shi 2009).

Biofilm formation can be divided into four steps:

1. Initial colonization
2. Early colonization
3. Middle colonization
4. Late colonization

While the affiliation of different genera and species with early, middle, and late colonizers is ambiguous, the successive model itself stays valid (see Figure 1.4).

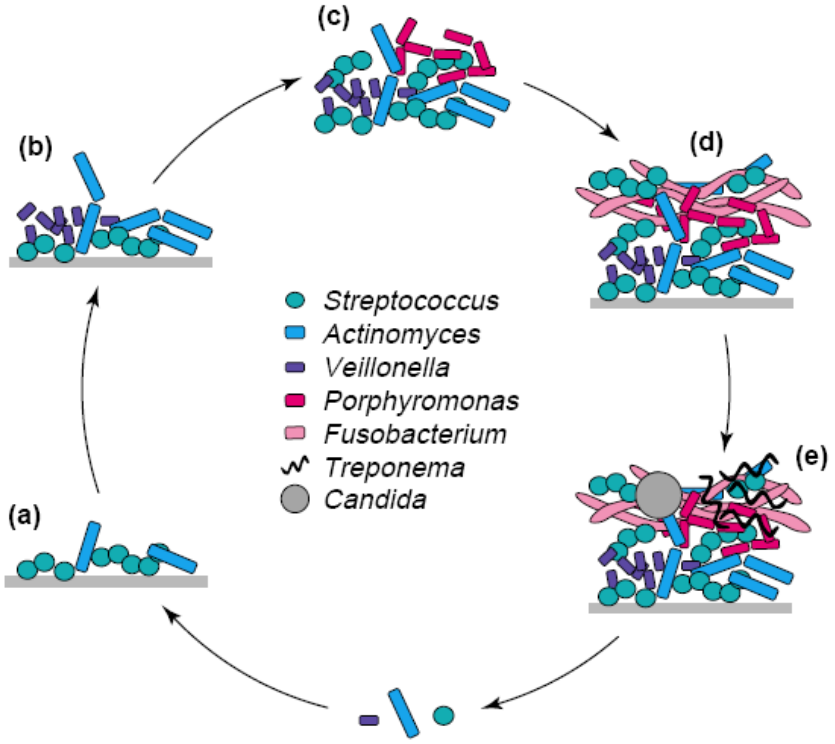


Figure 1.4: Maturation of an oral biofilm. a) Initial colonizers such as *Streptococcus* and *Actinomyces* bind to a freshly cleaned tooth surface coated with salivary pellicle. b) Early colonizers such as *Actinomyces* and *Veillonella* bind to the initial colonizers and form metabolic relationships with streptococci. c) *Porphyromonas* is able to colonize the growing biofilm relatively early. d) Anaerobic conditions develop and allow middle colonizers such as *Fusobacterium* to bind. e) Late colonizers including potential pathogens such as *Treponema* or even fungi such as *Candida* are able to bind to the mature biofilm. Graphics from Jenkinson and Lamont (2005).

1. Initial colonization. Planktonic oral bacteria require a mechanism for adhesion to avoid dislodgment through the constant flow of saliva. The salivary pellicle, which coats the tooth surfaces, provides a substrate for initial colonization. It consists of bacterial fragments and host-derived proteins, such as proline-rich proteins, salivary α-amylase, mucins, and salivary agglutinin that bind to the enamel of teeth. Pioneering organisms, primarily streptococci and *Actinomyces*, are able to bind to receptors of salivary proteins in the pellicle (Kolenbrander et al. 2010; Kolenbrander 2011b, pg. 5f.).

Oral bacteria start to accumulate as single cells or small clusters on the cleaned enamel. Specific species found were predominantly *Streptococcus oralis*/*S. mitis*, but also *S. sanguinis*, *S. gordonii* and *Actinomyces*. Various oral streptococci species have been shown to aggregate intragenerically (within its genus), which facilitates the repeatable process of initial colonization. Streptococci have versatile adherence properties, which allow them to bind to various oral sites including salivary pellicle-coated tooth surfaces, epithelium, dentures and implants, tonsils and the tongue, but there is species specific site specificity. For example *Streptococcus salivarius* predominantly colonizes the tongue (Aas et al. 2005). Only a small number of other organisms, for example *Actinomyces oris* and *Actinomyces naeslundii*, have been shown to coaggregate with these oral streptococci. *Actinomyces oris* type 2 fimbriae and *Streptococcus oralis* receptor polysaccharides have been shown to play a role in the coaggregation of these two distinct oral species (Jenkinson 2011; Kolenbrander 2011b, pg. 5f.). Actinomyces have been found to bind to proline-rich proteins and statherin in the salivary pellicle independent of streptococci (Kolenbrander et al. 2010; Jenkinson 2011).

2. Early colonizers. With the initial colonizers providing a base for other organisms to bind, coadherence of early colonizers can occur individually or in multispecies clusters. According to Kolenbrander, genera of early colonizers that bind to the growing biofilm of the initial colonizers include: *Eikenella*, *Neisseria*, *Porphyromonas*, *Prevotella*, *Rothia,* and *Veillonella* (Kolenbrander 2011b, pg. 6).

Veillonella spp. is known to coaggregate readily with the initial colonizers streptococci and *Actinomyces*. This close spatial connection can be explained by metabolic dependencies. Unlike streptococci and *Actinomyces*, which can utilize sugars for growth, *Veillonella* spp. are dependent on organic acids such as lactic acid (Kolenbrander 2011b, pg. 6,7). Lactic acid is produced by streptococci and *Actinomyces* and therefore *Veillonella* spp. thrive in close proximity to these initial colonizers (Jenkinson 2011; Kolenbrander 2011b, pg. 7). *Porphyromonas gingivalis*

has also been shown to be able to grow in the presence of the initial colonizer *Streptococcus gordonii* and has been isolated from developing biofilm as early as 4 hours past initiation. This is a surprising finding given that *Porphyromonas gingivalis* is an anaerobe and one of the main organisms associated with periodontitis (Kolenbrander 2011b, pg. 7). *Porphyromonas gingivalis* has been shown to interact with early, middle, and late colonizers (Periasamy and Kolenbrander 2009).

3. Middle colonizers. Middle colonizers such as *Fusobacterium nucleatum*, *Streptococcus mutans*, and *Capnocytophaga gingivalis* can be isolated from the developing biofilm at approximately 4 to 8 hours post-cleaning (Kolenbrander 2011b, pg. 8, color plate 1). During the biofilm development, the initial and early colonizers do not stop multiplying; instead, the addition of new species as well as the growths of the existing organisms contributes to the increase in biomass. *Fusobacterium nucleatum* is able to coaggregate with all the initial and early colonizers (Kolenbrander 2011b, pg. 8) as well as many late colonizers (Kolenbrander et al. 2010). Interestingly, *Fusobacterium* rarely aggregates within its genus, but readily with intergenetic members of the biofilm. This genus interacts with the largest number of different taxa (see Figure 1.5, long orange bacterium). *Fusobacterium* is therefore regarded a "bridging organism". *Fusobacterium nucleatum* itself is not pathogenic, but it allows late colonizing pathogens to adhere and usually appears before periodontitis develops and is clinically recognized. The stage of middle colonization has also been called "the crossroad between health and disease". An in vitro colonization experiment showed *Fusobacterium nucleatum*'s coaggregation to obligate anaerobes such as *Porphyromonas gingivalis* or *Prevotella nigrescens* allowed the latter to survive in oxygenated environments.

Another middle colonizer is *Streptococcus mutans*, which is one of the causative organisms in dental caries. Intriguingly, this pathogenic *Streptococcus* species does not appear

with the other streptococci in the initial and early colonizers (He and Shi 2009; Kolenbrander et al. 2010; Kolenbrander 2011b, pg. 8; Kolenbrander 2011a).

4. Late colonizers. Many late colonizers are associated with periodontal disease including *Tannerella forsythia*, *Treponema denticola,* and *Porphyromonas gingivalis* (Kolenbrander 2011b, pg. 9). Although, *Porphyromonas gingivalis* can interact with early colonizers as previously discussed. These three organisms comprise the so called "red complex" of organisms associated with severe periodontitis (Socransky et al. 1998). Another organism which is associated with an aggressive form of periodontitis and usually does not appear until later in the biofilm is *Aggregatibacter actinomycetemcomitans* (Kolenbrander 2011b, pg. 9). Periodontitis and the milder form of gum disease, gingivitis, will be discussed in Chapter 1.3.4. Overall the interactions of middle and late colonizers are not as well understood; the main reasons being that there is a larger number of organisms involved which vary with host and site (Jenkinson 2011).

The following section is going to discuss examples of interspecies dependencies and interactions. There are many interspecies regulatory and inhibitory mechanisms, which alter the properties of the mixed biofilm (Kolenbrander 2011b, pg. 10-11).

Synergistic and antagonistic interactions. Many oral organisms exhibit metabolic dependencies. One example is the lactate metabolism between streptococci and *Veillonella* spp. as mentioned earlier. In addition, *Aggregatibacter actinomycetemcomitans* also utilizes lactate from streptococci. However, many beneficial combinations have been observed but their metabolic basis is not known. Figure 1.5 shows known interactions of oral bacteria in a dental plaque biofilm.

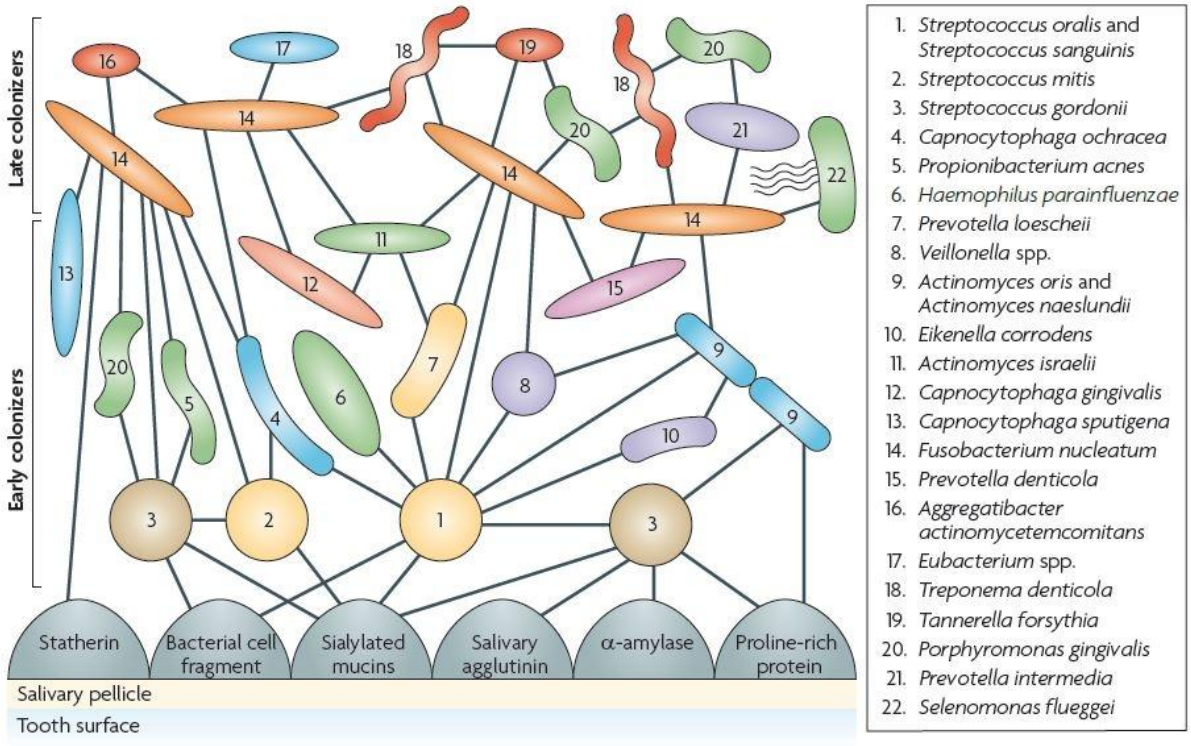


Figure 1.5: Interactions in the oral bacterial biofilm. Interaction of a selected number of oral bacteria is shown. Primary colonizers, *Streptococcus oralis*, *S. sanguinis*, *S. mitis*, and *S. gordonii* bind to host and bacterial proteins in the salivary pellicle bound the tooth surface. Secondary interactions from these initial organizers allow the formation of a mature biofilm. Not all interactions are shown. Graphic from Kolenbrander et al. (2010).

Kolenbrander and colleagues performed a series of experiment of single and combination growth experiments in flow cell to mimic the saliva flow in the oral cavity with saliva as the only nutrient source. As expected, most individual species could not grow on saliva alone, but were able to propagate in specific combinations, but not in others. These experiments showed that the sequential colonization is a highly specific and regulated process, but also very complex (Kolenbrander 2011b).

For example: *Streptococcus gordonii* grows in monoculture in suspension as well as a biofilm on a solid surface. In contrast, *Actinomyces naeslundii* and *Streptococcus oralis* cannot grow individually in monoculture, but are able to grow if cocultured (Palmer et al. 2001). *Veillonella parvula*, *Aggregatibacter actinomycetemcomitans*, and *Fusobacterium nucleatum* cannot

grow isolated, but in pairs as well as in a triple combination as Figure 1.6 shows. It is noticeable that the red *Fusobacterium* is always in contact with at least one of the other species (Kolenbrander 2011a).

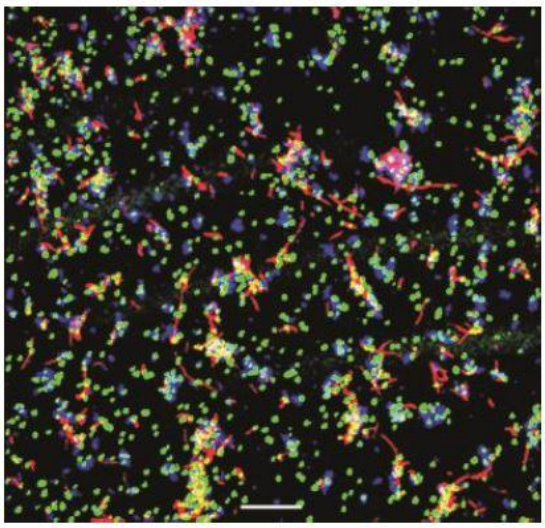


Figure 1.6: Confocal micrograph of an artificial biofilm consisting of three species after 18 h in a flow cell with saliva as the only nutrient source. The bacteria are labeled with specific antibodies with different fluorophores: *Veillonella parvula* (blue), *Aggregatibacter actinomycetemcomitans* (green), and *Fusobacterium nucleatum* (red). Bar = 50um. Image from Kolenbrander (2011a).

This highly specific coaggregation process, especially between early and middle colonizers, can be partly reversed by the addition of proteases or simple sugars such as lactose, which suggests that there is a nutrient dependency network. This has been shown for some initial colonizing species streptococci and *Actinomyces*, but also between two streptococci and the early colonizer *Veillonella* spp. (Kolenbrander et al. 2010; Kolenbrander 2011a). Another example is the metabolic utilization of mucins and other glycosylated proteins. The combination of biochemical properties of a community of bacteria compared to a single group of organism, allows for a more efficient breakdown of the substrate. A community of four oral species (*Streptococcus mitis*, *S. gordonii*, *S. cristatus* and *Actinomyces naeslundii*) together is able to degrade mucin MUC5B, but the individual species alone fail to (Wickstrom et al. 2009).

Some interaction combinations inhibit growth. For example, any pairwise combination of *Actinomyces oris*, *Porphyromonas gingivalis*, and *Veillonella parvula* exhibit increased growth, although, no growth occurs, when all three are cocultured. Initial colonizers, including *Streptococcus oralis*, *S. gordonii*, *Actinmyces oris*, interact mutualistically with early colonizers such as *Veillonella parvula* and *Porphyromonas gingivitis*. However, the initial colonizers usually do not grow well together. They seem to be more mutualistic exclusive while the early colonizers can interact with a larger number of initial colonizers as well as middle (*Fusobacterium nucleatum*) and late colonizers (*Aggregatibacter actinomycetemcomitans*).

Inhibition and competition between members of the biofilm can be partly explained by defense molecules and colonization resistance. Defense molecules in the repertoire of oral bacteria include hydrogen peroxide, lactic acid and bacteriocins (Jenkinson 2011). Many streptococci produce hydrogen peroxide, which gives them a competitive advantage in the growing biofilm. It can cross cell membranes and cause oxidative damage in cells. Some oral bacteria such as *Actinomyces oris* produce catalase to neutralize hydrogen peroxide. Interestingly, *Streptococcus gordonii* kills itself in monoculture after it enters stationary phase, but a mixed biofilm with *Actinomyces oris* can rescue it (Kolenbrander et al. 2010).

Other antimicrobial compounds have an effect on biofilm structure. Mutacins are antibiotic peptides produced by a number of oral organisms including *Streptococcus salivarius*, *Streptococcus mutans*, and *Lactococcus lactis*. *Streptococcus salivarius* produces mutacins, which are effective against a wide range of streptococci including *Streptococcus pyogenes*. While a confocal laser scanning microscopy image of monocultures of both shows a confluent and compact biofilm, the coculture reveals that *Streptococcus salivarius* dominates the biofilm, the biofilm includes protrusions and *Streptococcus pyogenes* forms pillar-like compact structures as if to evade the *Streptococcus salivarius* by surrounding itself with a layer of less-sensitive cell. Another example is the *Streptococcus mutans* mutacins that are effective against almost all streptococci species. This property could allow *Streptococcus mutants* to overcome the initial colonization

with other streptococci and establish itself in a more mature, cariogenic biofilm. *Veillonella* and *Bifidobacteria* which are often in association with *Streptococcus mutans* are likely to be more resistant to these mutacins. The production of mutacins is coupled to the production of autoinducer 2 (AI2), which signals the maturation of the biofilm (Jenkinson 2011). Interspecies communication through quorum sensing molecules such as AI2 is another essential part in biofilm formation. Increasing AI2 concentration through a denser biofilm could be the signal to progress in biofilm maturation. Initial colonizers only thrive under relatively low AI2 concentrations and are inhibited by higher concentrations. *Fusobacteria* require a higher AI2 concentration for colonization. This explains why *Fusobacteria* are usually not seen before 8h of biofilm formation, despite their relatively high concentration in saliva. Oral hygiene resets the biofilm to the initial stage and the biofilm formation can restart again (Kolenbrander et al. 2010).

The simple colonization with commensal bacteria can lead to a resistance against pathogens. When binding sites, for example in the salivary pellicle, are occupied by commensals, colonization with pathogens is hindered. This phenomena is called colonization resistance and can be disrupted by antibiotic treatment (Wade 2013) . Other modulatory effects are competition for nutrients (effective degradation of complex molecules), creation of unfavorable condition for other bacteria (changes in pH, oxygen availability, redox potential), or production of antimicrobial substances (mutacins, H_20_2, acids, etc.) (Marsh and Percival 2006). In addition, the human host contributes antimicrobial molecules such as secretory IgA, lysozyme, and other antimicrobial proteins to modulate the biofilm (Boonanantansarn and Gill 2011, pg. 28).

The biofilm formation in dental plaque and likely in other oral communities is a highly regulated and complex succession, especially in a mature biofilm. Even though or because the community structure in dental biofilm is very complex, it withstands perturbations surprisingly well and is relatively stable. The high variability in oral communities between healthy

individuals that has been observed in open survey experiments emphasizes this finding. However, the functional properties (gene content) are likely to be very similar in healthy oral bacterial communities even if they are comprised of different taxa. This observation has been made in the oral cavity and also in other human habitats. While the taxonomic distribution at different sites of the healthy oral cavities and the gut varies substantially, the functional distributions of broad enzymatic functions remains very similar (Human Microbiome Project Consortium 2012).

1.3.3 The oral microbiota in health and disease

In a healthy state, there is an ecological balance between the host and the affiliated microbiome. The healthy microbiome supplies essential natural host defense mechanisms as mentioned in the last section (Jenkinson and Lamont 2005; Marsh and Percival 2006; Filoche et al. 2010; Boonanantansarn and Gill 2011, pg. 28). However, internal and external factors such as maturation of the biofilm due to neglect of oral hygiene, a diet high in carbohydrates, reduced host immune defense, or smoking can predispose an individual to oral disease and systemic disease caused by oral bacteria (Marsh and Percival 2006; Filoche et al. 2010).

A number of studies have attempted to characterize the bacterial composition of a healthy oral cavity devoid of disease (Aas et al. 2005; Keijser et al. 2008; Costello et al. 2009; Lazarevic et al. 2009; Zaura et al. 2009; Bik et al. 2010; Lazarevic et al. 2010; Human Microbiome Project Consortium 2012; Ling et al. 2012). Substantial variation between individuals has been found, which did not group by global geography (Nasidze et al. 2009a). This result suggests that broad dietary preferences and cultural habits, which may differ between geographic locations, likely do not define signature oral communities. Bacterial genera associates with a healthy oral microbiome are *Streptrococcus*, *Gemella*, *Abiotrophia*, *Granulicatella*, *Rothia*, *Neisseria*, and *Prevotella* (Boonanantansarn and Gill 2011, pg. 22). However, even "pathogens" such as members of the “red complex” which are associated with severe periodontitis are found in low

abundances in clinically healthy individuals (Bik et al. 2010). These findings suggest that oral diseases are bacterial diseases, but not true infectious diseases. Instead, they are a consequence of environmental factors that cause the microbiome to shift in composition and function. These shifts can eventually lead to an increase in pathogens, which causes disease initiation and progression (Boonanantansarn and Gill 2011, pg. 28; Wade 2013),

Almost 700 bacterial taxa at the species level have been described in the oral cavity in health and disease and summarized in a manually curated database, the human oral microbiome database (HOMD) (Chen et al. 2010; Dewhirst et al. 2010). It includes members in 13 phyla predominantly Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, Spirochaetes, but also Chlamydiae, Chloroflexi, Euryarchaeota, Synergistetes, Tenericutes, as well as the new divisions GN02, SR1, and TM7 (Dewhirst et al. 2010). In addition, a small number of archaea of the methanogens have been found in the oral cavity (Conway de Macario 2012).

Immune defense. The mucosal immune system is a major defense system of the human body. The total surface covered with mucosa is almost 400 square meters, which includes the gastrointestinal, urogenital, and respiratory tract. Most pathogens affect or enter through the mucosa; therefore, a highly functional defense system is in place (Wade 2013). Monocytes and granulozytes invade the affected areas and cause local inflammation with their typical four signs of heat, swelling, redness, and pain. The increase in blood flow recruits further leukocytes and promotes healing (U.S. Department of Health and Human Services 2000). Saliva supplies secretory IgA, lysozyme, lactoferrin, sialoperoxidase, and antimicrobial peptides such as histatins. In addition, it contains proteins and glycoproteins, which act as nutrients for oral microbes and promote attachment points of some bacteria for colonization or aggregation for clearance. Changes in saliva flow rate is very likely to alter microbial composition and have a detrimental effects on oral health (Marsh and Percival 2006).

Because the mucosa is constantly exposed to the indigenous bacteria, it is important for the host immune system to be able to distinguish between normal commensals and pathogens in order to avoid either a constant state of inflammation or an invasion of pathogens. The functional mechanisms are not completely understood. Several mechanisms may play a role in distinguishing commensal from pathogenic bacteria. First, pathogen associated molecular patterns (PAMPs) are recognized by Toll-like receptors (TLRs). Various PAMPs for example lipopolysaccharides (LPS), peptidoglycan, bacteria specific lipoproteins and proteins such as flagellin are recognized by a different set of TLRs, which could explain a specific reaction to pathogens but not commensals. To avoid premature inflammatory response, bacteria might need to reach the basolateral surface to trigger the proinflammatory signaling cascade. Commensal bacteria but not pathogens might elicit immunosuppression by suppression of NF-kappa B response and TNF alpha to prevent Interleukin (IL)-8 release. Antimicrobial peptides and proteins are produced by epithelial cells, which can kill or inhibit bacteria but might also neutralize modulins such as LPS (Wilson 2005, pg. 44).

1.3.4 Gingivitis and periodontitis

Oral microbial communities stay relatively stable over time, exhibiting microbial homeostasis, due to a harmonic balance of host and microbes. When this balance is disturbed (for example due to dietary changes, smoking, medications, loss in saliva flow, or immune suppression), formally minor components of the biofilm can increase in abundance and cause oral disease (Marsh and Percival 2006).

Gum diseases, gingivitis and periodontitis, are very common human burdens. As discussed previously, a successive biofilm starts to form on cleaned surfaces, shortly after oral hygiene procedures. In periodontal disease, inflammation of the gingiva and adjacent tissues occurs due to bacterial triggers (Filoche et al. 2010). Gingivitis originates from the gingival

crevice adjacent to subgingival plaque in contrast to tooth decay which usually originates from supragingival plaque (Loesche 2007). Over 90 % of adults suffer from gingivitis at some point in their life which is a reversible, infectious, inflammatory disease of the gingiva (Coventry et al. 2000). When the amount of biofilm increases to a level incompatible with health, irritation of gingiva causes bleeding and inflammation (Marsh and Percival 2006; Wade 2013). Regular oral hygiene can reset the succession to an healthier initial or early colonizer state (Kolenbrander et al. 2010).-In addition, age, diabetes (Barbour et al. 1997), pregnancy (Coventry et al. 2000) and external factors such as lifestyle play an important role and can shift the balance towards a disease-associated state (Nishida et al. 2008). Smoking, for example, can lead to an increase in periodontitis incidence and a shift towards a decrease of commensal bacteria and an increase in potential pathogens (see Chapter 3.1.2 on the effect of smoking).

If left unattended, gingivitis can progress to periodontitis. In periodontitis, attachment loss between the gingiva and teeth causes the formation of periodontal pockets which are heavily colonized with anaerobic bacteria (Wade 2013), see Figure 1.3 C. Host immune defense is aggravating the disease through protease activity. Bone loss through reabsorption eventually causes tooth loss (Wade 2013).

The first evidence that periodontitis is a bacterial disease came from a study in 1956, where Mitchell and Johnson showed that penicillin inhibits periodontitis in the animal model (Mitchell and Johnson 1956). Keyes and Jordan showed eight years later that periodontitis is transmissible (Keyes and Jordan 1964). In 1965, Howell et al. isolated oral species which were implicated in the development of the disease (Howell et al. 1965). For a long time it was believed that plaque quantity, not composition, is responsible for oral diseases. However, Listgaten showed in 1976 via electron microscopy that the structure between periodontically healthy and diseased plaque differed (Listgarten 1976; He and Shi 2009).

Moving beyond the one pathogen/one disease hypothesis, a landmark paper from Socransky et al. in 1998, defined microbial complexes and their association with clinical parameters of chronic periodontitis (Socransky et al. 1998). Members of the red complex are *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*. Their presence exhibited the strongest correlation with pocket depth and severe periodontitis. Socransky et al. also defined other complexes based on their association with gingival health. The orange complex, which was also associated with periodontal disease, includes the organisms: *Fusobacterium nucleatum*, *Prevotella intermedia*, *Prevotella nigrescens*, *Peptostreptococcus micros*, *Streptococcus constellatus*, *Eubacterium nodatum*, *Campylobacter showae*, *Campylobacter gracilis*, and *Campylobacter rectus*. These definitions are based on a study using checkerboard DNA-DNA hybridization, not an open survey approach. Newer studies found additional genera associated with periodontitis such as *Peptostreptococcus*, *Filifactor*, *Megasphaera*, *Desulfobulbus*, *Campylobacter*, *Selenomonas*, *Deferribacteres*, *Dialister*, *Catonella*, *Tannerella*, *Streptococcus*, *Atopobium*, *Eubacterium*, and *Treponema* (Kumar et al. 2005). In addition, *Aggregatibacter actinomycetemcomitans* is frequently found in aggressive periodontitis (Boonanantansarn and Gill 2011, pg. 24). However, these organisms are not found in all periodontitis cases. Recent culture independent studies (for example Kumar et al. 2003; Kumar et al. 2005; Paster et al. 2006; Liu et al. 2012) could identify more organisms that are associated with periodontitis, however their causative implication on the disease is still unknown (Boonanantansarn and Gill 2011, pg. 24).

While the majority of organisms are symbiotic beneficial or commensals, only a small percentage of organisms are pathogens or obligate pathogens. If the community becomes unbalanced and the abundance of potential or obligate pathogens increases, oral disease including infection and inflammation can occur (Boonanantansarn and Gill 2011, pg. 22). Because periodontitis is a polymicrobial infection and not associated with a singular causative agent, plaque needs to be regarded from a microbial ecological perspective and pathological communities should be analyzed (Filoche et al. 2010). In addition, biofilm function needs to be

analyzed beyond a purely phylogenetic approach to fully understand periodontal disease due to potential functional redundancies (Wade 2013).

1.3.5 Caries

Dental caries or tooth decay, which are used interchangeably in this dissertation, is one of the most prevalent chronic childhood diseases (U.S. Department of Health and Human Services 2000, pg. 63.). Tooth decay develops progressively starting with initial enamel degradation, which can further progress to exposure of dentine eventually leading to a path for microorganisms to the dental pulp (see Figure 1.3 C) (Boonanantansarn and Gill 2011, pg. 25). Even though it has been a human burden for millennia and has been observed even in prehistory, a dramatic increase was observed in Britain around 1850, when cane sugar and refined flour became largely available for the masses (He and Shi 2009).

The perception of a small causative organism for tooth decay has been around since 5000 BC, where Sumerian texts mention "tooth worms". However, the causative nature of caries was not begun to be understood until around 1890, when Miller, a US oral microbiologist who worked with Robert Koch in Berlin, hypothesized that a microorganism would ferment carbohydrates into acid, which would then cause a demineralization of teeth (Burnett and Scherp 1957, pg. 31; He and Shi 2009). Plaque as a habitat for bacteria, as described by Black and Williams, provided another foundation for our modern understanding of caries development (He and Shi 2009). The 20th century finally brought some light onto this disease. In 1924, Clarke was able to isolate a bacterial species from human caries sites, which was able to ferment sugars and produce acidic conditions (pH 4.2) (Clarke 1924). The organism was *Streptococcus mutans*, a well-known causative organism for caries today; Clarke, however, failed to show causality. In the 1960s, Fitzgerald and Keyes used a hamster model to show that caries is transmissible and that the introduction of *Streptococcus mutans* causes the disease (Fitzgerald and Keyes 1960).

The specific plaque hypothesis assumes that a low number of pathogens such as mutans streptococci, *Streptococcus mutans* or *Streptococcus sobrinus*, cause tooth decay. However, newer culture independent research suggests that this is an oversimplification and the pathology should be explained by a nonspecific or ecological plaque hypothesis as only a subset of caries lesions is associated with these potential oral pathogens (Kanasi et al. 2010). In addition, *Streptococcus mutans* colonization is not sufficient for tooth decay, as the organism has been isolated from a small number of healthy oral cavities (Aas et al. 2005; Gross et al. 2010).

The nonspecific plaque hypothesis assumes that numerous bacterial species in dental plaque are causing caries while the ecological plaque hypothesis explains caries with a shift of the microbial community due to changes in the local oral environment (Boonanantansarn and Gill 2011, pg. 25f.). This new "community-and microbial ecology-based pathogenic theory" (He and Shi 2009) serves as a working hypothesis for oral microbiology research as well as disease prevention and treatment approaches.

Disease is usually a result of the interaction of multiple species within the biofilm. One hypothesis of the etiology of tooth decay is that consumption of a diet high in carbohydrates (especially sugar), frequent feeding, and consumption of acidic drinks leads to changes in the local conditions on teeth to allow growth of more acidogenic bacteria. The acidification leads to demineralization of enamel and eventually cavitation (Filoche et al. 2010; Jenkinson 2011). *Streptococcus mutans* and lactobacilli are thought to be especially destructive because they are able to grow at pH levels which are restrictively low for other bacteria (Kolenbrander et al. 2010). In addition, *Streptococcus mutans* as well as others has the ability to ferment a number of carbohydrates to lactic acid. Mutan, a water insoluble glucan with a high proportion of alpha1,3-linkages similar to dextran, aids in adherence of its producer *Streptococcus mutans* to the tooth surface (Russell 2008, pg. 120).

Caries active individuals without *Streptococcus mutans* have multiple distinct microbiological profiles compared to healthy subjects, which suggests that multiple out of order

constellations can cause the disease. The communities on caries lesions also change with caries succession (Aas et al. 2008). In general, alpha diversity declined with caries progression, which suggests that the healthy diverse oral community structure is disturbed (Li et al. 2005; Gross et al. 2010).

Various recent studies associated bacteria other than mutans streptococci with caries including *Lactobacillus* spp., *Bifidobacterium dentium*, *Scardovia wiggsiae*, *Actinomyces*, and *Veillonella*. Organisms that were repeatedly correlated with oral health are *Abiotrophia defectiva*, *Streptococcus sanguinis*, and other non-mutans streptococci (Corby et al. 2005; Aas et al. 2008; Kanasi et al. 2010; Jenkinson 2011). Furthermore, the association of an organism with caries does not necessarily suggest direct involvement in the disease, but the organism might support a biofilm composition that allows the infection (Kanasi et al. 2010). These supporting organisms as well as the actual pathogen could be therapeutic targets.

1.3.6 Oral and non-oral diseases associated with oral microbes

Besides the two most common oral diseases tooth decay and gum disease, which were discussed in the previous two subchapters, a number of other diseases in the oral cavity exist for example oral Candidiasis (infection with the fungus *Candida*) and oral cancer (U.S. Department of Health and Human Services 2000, pg. 42f.; Wade 2013). In addition, a variety of viruses can be detected in saliva or transmitted through the salivary route including rabies, human papilloma viruses and herpes viruses including Epstein-Barr-Virus and Herpes Simplex viruses (Slots and Slots 2011).

Oral cancer. Oral cancer with 300,000 new cases diagnosed each year is the sixth most common cancer worldwide. The prevalence in the US and the world is increasing. In the US alone 30,000 new cases are diagnosed each year and about 8000 deaths will be accredited to this disease with a poor 5-year survival rate of about 50%. The majority of cases derive from

invasive squamous cell carcinomas, which have their origin in the mucosa (Boonanantansarn and Gill 2011, pg. 26). There is epidemiological evidence that even after controlling for alcohol and tobacco consumption, poor dental health is associated with an increase in oral and other cancers of the gastrointestinal tract (Fitzpatrick and Katz 2010). The abundance of three oral bacteria, *Capnocytophaga gingivalis*, *Prevotella melaninogenica*, and *Streptococcus mitis*, was increased in saliva obtained from patients with oral squamous cell carcinoma compared to oral cancer free controls (Mager et al. 2005). An increase in TLR-5 expression in oral cancer cells predicted negative cancer outcome (Kauppila et al. 2013). It is unclear at this point whether microorganisms associated with oral cancer are secondary colonizers or directly involved in the disease (Boonanantansarn and Gill 2011, pg. 26).

General. The significance of oral health expands even further when we regard implications of oral bacteria in systemic diseases. Oral bacteria can gain access to the bloodstream easily and frequently causing transient bacteremia (Wade 2013). A number of diseases display increasing evidence linking them to oral bacterial infections, including cardiovascular disease, pneumonia, and preterm birth. These findings may eventually change prevention and treatment plans (Boonanantansarn and Gill 2011, pg. 27).

Cardiovascular disease. Cardiovascular disease is the leading cause of death in the US (Kochanek et al. 2011). Periodontitis has been associated with an increased risk of cardiovascular events, such as myocardial infarction and stroke. It is believed that the chronic periodontal inflammation that accompanies the very common periodontal disease can cause a significant inflammatory burden on the whole body, which can contribute to cardiovascular disease, including atherosclerosis and cardiovascular events (Boonanantansarn and Gill 2011, pg. 27). Epidemiological meta-analyses showed an increase in cardiovascular disease in periodontitis patients (Meurman et al. 2004; Bahekar et al. 2007) and periodontal intervention was associated with a positive cardiovascular outcome (Tonetti 2009).

Multiple studies have suggested that the oral cavity is a reservoir for bacteria associated with atherosclerotic plaque. Transient bacteremia can be observed in susceptible individuals after mastication (chewing) as well as personal and professional oral hygiene. Among the bacteria observed in bacteremia are several streptococci, *Aggregatibacter actinomycetemcomitans*, *Porphyromonas gingivalis*, and *Fusobacterium nucleatum*. Endocarditis is a common result of bacteremia, but any organ can be affected including the skeleton and central nervous system (Parahitiyawa et al. 2009). Oral bacteria such as *Streptoccocus mutans* and *Aggregatibacter actinomycetemcomitans* have been found in several coronary specimens (Nakano et al. 2009). A study comparing oral, gut, plaque microbiota in patients with atherosclerosis showed some overlap between these body habitats. Several OTUs belonging to *Streptococcus*, *Veillonella*, and *Actinomyces* could be detected in oral and atherosclerosis samples of the same patients (Koren et al. 2011). There is evidence about a shared genetic susceptibility locus of coronary heart disease and aggressive periodontitis (Schaefer et al. 2009). Even if one accounts for the effect of oral disease negatively influenceing the outcome on cardiovascular disease, poor oral health is significantly correlated with increased mortality from other causes (Jansson et al. 2002).

However, some oral bacteria might also have a positive effect on blood pressure and be protective against gastric ulcers through nitrate reduction and therefore an increase in nitric oxide in mice (Petersson et al. 2009).

Respiratory disease. Respiratory disease is responsible for a large number of deaths in the US each year. Chronic lower respiratory disease is the 3rd most common cause of death, while influenza and pneumonia are grouped together on place 8 (Kochanek et al. 2011). Ten to 20 % of hospitalized patients develop a respiratory tract infection, which often result in pneumonia. Interestingly, pneumonia in patients with ventilator-assisted breathing is more often caused by oral bacteria (*Haemophilus influenzae*, *Escherichia* spp., *Streptococcus pneumoniae*, *Proteus mirabilis*, and *Pseudomonas* spp.) than pneumonia in non-ventilated patients (Boonanantansarn and Gill 2011, pg. 27).

Preterm birth and low birth weight. Epidemiological evidence suggests that periodontitis may be a risk factor for adverse pregnancy outcome including premature birth and low birth weight of the neonate (Bobetsis et al. 2006). In the mouse model, oral bacteria injected into the blood stream to mimic bacteremia could cause intrauterine infections in pregnant mice (Fardini et al. 2010).

1.3.7 Comparison of saliva to other oral microbial habitats

As discussed in the introductory section 1.3.1, the oral cavity consists of several distinct tissues, including hard, nonshedding tooth surfaces and soft, shedding mucosal surfaces.

Most oral microbiology research focuses on the sub- or supragingival plaque because of its involvement in gingivitis and periodontitis. Other mucosal tissues include soft tissues of epithelium of buccal mucosa (cheeks), hard and soft palate, keratinizes gingiva (gums), tonsils, throat, and tongue (see Figure 1.3 A and B). The oral cavity is constantly moisturized and bathed in saliva.

While the number of studies focusing on saliva exclusively is very limited, some studies have compared saliva to other habitats. All oral habitats are more similar among themselves than oral habitats to distinct human body habitats such as the gastrointestinal tract, urogenital tract, or skin (Costello et al. 2009; Human Microbiome Project Consortium 2012). However, the bacterial communities on oral sites within the oral cavity are not uniform. A study by Mager et al. compared the abundance and presence of 40 bacterial species via checkerboard DNA-DNA hybridization on eight oral locations and saliva in 225 individuals (Mager et al. 2003b). The results yielded two distinct clusters. Saliva clustered with the microbiota of the tongue (dorsum and lateral). The remaining locations tested, namely buccal, floor of mouth, attached gingiva, vestibule lip (inner lining of lips), ventral tongue, and hard palate, clustered together. A recent study, part of the HMP, confirmed the finding that the salivary microbiome is most similar to microbiota of the tongue as well as the tonsils and throat (group 2) distinct from the buccal

mucosa, keratinized gingiva, and hard palate (group 1) and from sub- and supragingival plaque (group 3). Group 1 had the highest Firmicute content, while group 2 had a relative increase in Fusobacteria, Actinobacteria, and TM7 compared to group 1 (Segata et al. 2012).

In addition, differences between the mucosal tissues and sub- and supragingival plaque were analyzed. Results of Mager et al. and other groups indicate that all soft mucosal tissues cluster apart from tooth associated communities (Mager et al. 2003b; Keijser et al. 2008; Segata et al. 2012). Sub- and supragingival bacterial communities are distinct despite their close spatial proximity. The main driver of this difference is thought to be oxygen availability which is high in supragingival plaque and low in subgingival (Segata et al. 2012).

The finding that different oral sites harbor a distinct microbial composition has been confirmed by Zaura et al. Furthermore, different teeth and even different sampling sites on a particular tooth (buccally, lingually, or approximal surfaces) harbor different microbial compositions (Haffajee et al. 2009; Zaura et al. 2009). This site-specific colonization can be explained by surface anatomy, including roughness and the resulting differences in plaque accumulation (Quirynen 1994; Boonanantansarn and Gill 2011, pg. 23). Other explanations are differences in nutrient availability, redox potentials, species competition for attachment sites, interspecies interactions, host defense, different receptors on tissues, pH, atmospheric conditions, salinity and water activity in saliva (Boonanantansarn and Gill 2011, pg. 21f.).

Studies of individual sites of the oral cavity suggest that there are very likely hundreds of micro-communities, each functionally optimized to their habitat (Filoche et al. 2010; Jenkinson 2011). Given such a diversity of microhabitats in the oral cavity, comparing studies obtained from different sites is difficult. Saliva is thought to give an overview of bacteria found in the mouth according to Asikainen and Karched (2008).

Saliva. Saliva is the aqueous, slightly viscous fluid produced by three major and hundreds of minor salivary glands distributed in the oral cavity. The three major glands

(submandibular, sublingual, paired parotid, Figure 1.7) produce saliva with different viscosities. The resting flow of saliva is ~ 0.5 ml min^{-1}, but can substantially increase during chewing (U.S. Department of Health and Human Services 2000; Carpenter 2013). The human body produces approximately 750ml saliva per day, with approximately 10^7 bacteria per ml. Bacteria in the mouth are constantly flushed by saliva and swallowed (Curtis et al. 2011).

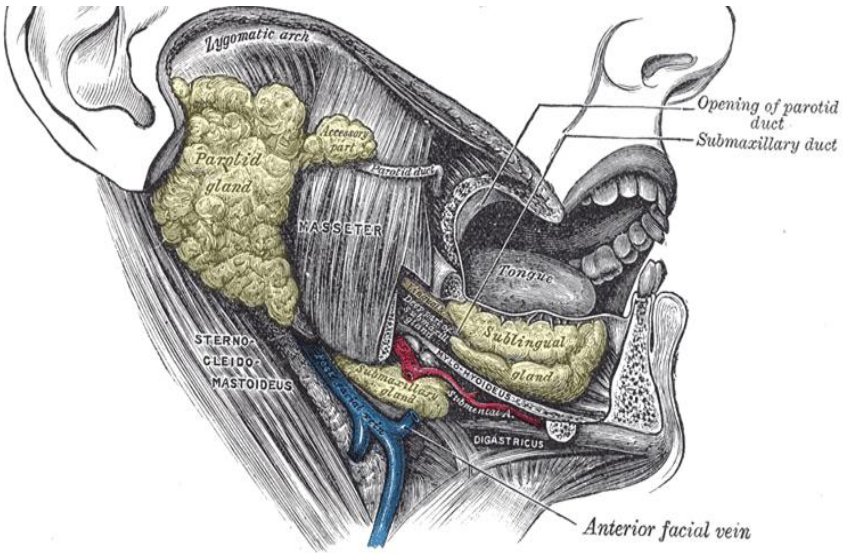


Figure 1.7: Dissection of salivary glands (highlighted in yellow) on the right side, modified from Gray and Lewis (1918).

Saliva production is initiated by a signal of the brain to the acinar cells to secrete sodium and chloride ions into the lumen of the gland. Water from the blood system enters the lumen until isotonic saliva is obtained; the salt is later recovered to yield hypotonic saliva. Saliva consists of 99 % water with several proteins including, secretory IgA, mucins for lubrication, amylase, lingual lipase for fat detection, proline rich proteins and a slightly alkaline electrolyte solution for moisturization of food for easier swallowing. While amylase, the most abundant protein in saliva, aids in starch pre-digestion, its actual function is more likely the clearance of food particles and reduction of microbial substrate than actual digestion of starch containing foods.

The primary purpose of saliva is tissue protection. The food bolus is moisturized, which helps to protect the tissues while swallowing and to enhance digestion. The constant flow of saliva aids to dislodge food, microbes and viruses from the mucosal surfaces and be swallowed. In addition to the physical protection, saliva contains a variety of antimicrobial (antibacterial, antifungal as well as antiviral) components including lysozyme, lactoferrin, peroxidase, and histatins. Lubrication and hydration as well as substantial concentrations of epidermal growth factor (EGF) and vascular endothelial growth factor (VEGF) aid the wound healing in the mouth. Other functions of saliva are enhancement of taste through solubilization of food particles and increasing the direct contact of food with the taste buds on the tongue. Saliva also has pH buffering capacities to neutralize acid foods or drinks as well as metabolites of oral bacteria such as lactic acid from mutans streptocci to protect from caries. Calcium and phosphate in saliva aids in remineralization the hydroxyapatite of teeth (U.S. Department of Health and Human Services 2000, pg. 26f.; Carpenter 2013). In addition, saliva is thought to have a role in recolonizing oral sites and microbial dispersion (Filoche et al. 2010).

1.4 Purpose of the study

Several human microbiome studies have shown that large inter-individual variations exist in the oral bacterial communities, but the driving factors are currently unknown. The objective of this dissertation was to examine factors that influence the composition of the salivary microbiota including host genetics, inter- and intra-individual temporal changes, and environmental variables on the example of cohabitation and substance use.

Parts of the bacterial small ribosomal subunit were PCR amplified and pyrosequenced from approximately 500 banked saliva samples from the Institute of Behavioral Genetics at the University of Colorado. Sequences were analyzed with QIIME in order to asses taxonomic and diversity differences. A longitudinal twin study design allowed for the assessment of host genetics based on monozygotic to dizygotic twin pair differences and an assessment of

temporal changes during adolescence and young adulthood. A cross-sectional study of individuals aged 12 to 65 with various degrees of substance use including smoking, alcohol consumption, and stimulant use, was examined for age and substance effects.

CHAPTER II

SALIVARY MICROBIAL COMMUNITIES IN ADOLESCENT AND YOUNG ADULT TWINS

2.1 Introduction

This chapter discusses the results from a bacterial 16S rRNA gene based study of 264 saliva samples derived from 107 individuals, including 27 monozygotic (MZ) and 28 dizygotic (DZ) twin pairs, sampled up to three times approximately 5 years apart, mainly at age 12-13, 17-18, and 22-24 years of age. Parts of this chapter have been published under the title "Nurture trumps nature in a longitudinal survey of salivary bacterial communities in twins from early adolescence to early adulthood" in the peer-reviewed journal "Genome Research" (Stahringer et al. 2012). The first subchapter discusses the current knowledge of the effect of development on the oral microbiome in humans, the proposed routes of transmission and the effect of host genotype. The result section shows a core microbiome on genus level in this sample set, providing evidence that the overall bacterial composition is mainly influenced by environment, especially cohabitation, not host genetics. In addition, it demonstrates that individuals during adolescence exhibit similarity over two five year spans, but not over 10 years. Lastly, it shows an age-specific abundance profile of certain bacterial taxonomic groups.

2.1.1 The oral microbiota in children and young adults and routes of transmission

The lifelong relationship with microbes begins at birth. During the birth process, sterile neonates are inoculated with their first microbiota, which depends on delivery mode (Dominguez-Bello et al. 2010). Differences can still be detected in the oral cavity in 3 months old infants (Lif Holgerson et al. 2011). Oral bacteria are often transmitted vertically between family members, especially from mother to child (Asikainen et al. 1997; Van Winkelhoff and Boutaga

2005; Li et al. 2007). A recent 16S rRNA based study showed that mothers and their children shared significantly more of their oral microbial communities, even compared to fathers and children (Song et al. 2013). Transmission routes are thought to be mainly through saliva, but human milk also harbors a very diverse set of microbes (Hunt et al. 2011). The oral cavity of the newborn consists exclusively of mucosal surfaces that are predominantly colonized by *Streptococcus oralis*, *S. mitis*, and *S. salivarius*. In the maturing biofilm, *Prevotella*, *Fusobacterium*, and *Veillonella* appear. Teeth eruption and the accompanying formation of new habitats of the tooth surface and the gingival crevice change the colonization patterns. The diversity increases and colonization with *Actinomyces*, *Lactobacillus*, *Rothia*, *Neisseria*, *Capnocytophaga*, and *Prevotella* is observed (Marsh and Percival 2006). The bacterial diversity in the oral cavity steadily increases within the first year of life (Song et al. 2013). Some microorganisms, such as the periodontal pathogen *Aggregatibacter actinomycetemcomitans* remain in the oral cavity once acquired (Saarela et al. 1999) and members of the red and orange complex associated with periodontitis (Socransky et al. 1998) accumulate during childhood with age (Papaioannou et al. 2009). In contrast to the gut which does not structurally change with age and reaches an adult like microbiota by age 3 (Yatsunenko et al. 2012), the oral cavity undergoes changes later in development. The salivary microbial composition in children age 3 to 18 was analyzed with a special emphasis on the development of dentition, from deciduous (milk) teeth via mixed dentition to a full permanent set of teeth. It is apparent that the oral microbiota changes with teeth composition and is still developing even when full permanent dentition is reached (Crielaard et al. 2011). Puberty seems to cause a shift in the oral microbiota in boys from a child like state to a different, but not quite adult like state (Moore et al. 1993). The changes in sex hormones in puberty as well as pregnancy have a negative effect on the gingiva (Guncu et al. 2005) and tooth integrity (Lukacs and Largaespada 2006). At some point in adulthood, which has yet to be clearly defined, the oral bacterial composition remains relatively stable. However, horizontal transfer between unrelated individuals, for example spouses or kindergarteners, has

been reported (Asikainen et al. 1997; Van Winkelhoff and Boutaga 2005; Domejean et al. 2010). In a recent study by Song et al. cohabiting spouses shared a greater proportion of their microbiota on several body habitats including the tongue than random individuals (Song et al. 2013). In the aging mouth, diminished dentition, denture wearing, reduction of saliva flow and changes in immune response have been observed and with it a change in microbial composition. This development may be aggravated by a variety of medication, many of which cause a reduction in salivary flow (U.S. Department of Health and Human Services 2000, pg. 38; Marsh and Percival 2006). The prevalence of periodontitis in older age increases, which could be partly due to a reduction in innate immune function (Hajishengallis 2010).

The oral cavity has been shown to be a human habitat, which is most similar between individuals (low beta diversity), but is very diverse (high alpha diversity) (Human Microbiome Project Consortium 2012). Oral samples taken repeatedly from the same adult individuals clustered together between short time periods of up to three months (Costello et al. 2009; Lazarevic et al. 2010). In the HMP, 131 individuals were sampled twice, 35-404 days apart, and showed that the composition of bacterial OTUs as well as function of the oral cavity of the same individual were on average more similar than measurements between individuals. However, the temporal variation was quite high and did not allow for an identification based on microbial profiles given this time distance (Human Microbiome Project Consortium 2012). Forensic identification of individuals using microbiota is an appealing idea, as has been shown in a study that matched the microbes on fingertips to the computer keyboard (Fierer et al. 2010); also reviewed by Alan and Sarah (2012). However, it does not seem to be feasible in the oral cavity, at least not based on total community metrics. It might be possible in the future to match whole shotgun genome sequences of members of the oral microbial community, but this is mere speculation at this point.

2.1.2 The effect of the host genotype on microbial communities

The composition of microbiota on the human body could be shaped through various community assembly scenarios, for example through exposure to a diverse set of microorganisms as well as environmental selection on the existing set (Costello et al. 2012). The environmental selection forces on the microbial community include the human host environment; for example, host immune system, salivary flow, or tooth shape. It is possible that the human genotype modulates the microbial composition through these or other factors.

Human twins provide a unique study system in which to dissect environmental and genetic contributions towards phenotypic traits and for estimating heritability. Twin studies have been used to estimate heritability for a variety of phenotypes including simple phenotypes such as height (Silventoinen et al. 2003) or body weight (Maes et al. 1997) and complex phenotypes; for example, behavior (Young et al. 2000), cognitive traits and cardiovascular risk (Boomsma et al. 2002), but also tooth size (Boraas et al. 1988). MZ twins share nearly 100% of their alleles while DZ twins share approximately 50% of their alleles. In the traditional twin study it is assumed that resemblance derived from shared environment is the same for MZ and DZ twins and any higher pairwise similarity of MZ twins compared to DZ twins is an indication for heritability.

Heritability can be calculated with the formula:

$$h^2 = 2 \text{ x } (r_{MZ} - r_{DZ})$$

where h^2 = heritability and r = correlation between twins.

Twin studies have been used for estimating the heritability in microbiome studies of the gut (Stewart et al. 2005; Turnbaugh et al. 2009; Hansen et al. 2011) and the oral cavity (Corby et

al. 2007). Turnbaugh et al. (2009) have demonstrated that the composition of microbial communities in the adult gut is more similar between twin pairs than it is between unrelated individuals and even mother to child (31 MZ pairs, 25 DZ pairs, age 21-32). These results suggest that early exposure is critical for the formation of the gut microbiome. Mother and twin shared significantly more than the unrelated persons, but less than the twins. This remains true even if the twin pair lived separately at the time of sampling. An older study of the gut, based on coarse community comparisons (TGGE) showed heritability in twins between 4 months and 10 years of age based on 13 MZ pairs and 7 DZ pairs (Stewart et al. 2005). However, newer 16S rRNA gene based studies with more statistical power due to increases in sample sizes failed to show a significant difference in beta diversity between MZ and DZ twins in the US and Malawi across a wide age range (Turnbaugh et al. 2009; Yatsunenko et al. 2012). While the study by Turnbaugh (2009) analyzed US adult twins age 21-32 (31 MZ pairs, 25 DZ pairs), the study by Yatsunenko et al (2012) examined twins from age 1 month to 17 years (US cohort: 36 MZ pairs, 44 DZ pairs; Malawi cohort: 15 MZ pairs, 6 DZ pairs). Genetically related individuals share more of their microbiota than unrelated, but MZ have not been found to be more similar than DZ twins, which suggests that shared environment has a greater effect than genetic disposition and the overall heritability of the gut microbiota is low.

Few studies have utilized twin studies for oral microbiology studies. Evidence suggests that a subset of oral organisms detected by reverse capture checkerboard may be attributed to genetic components in children around age 4 in 40 MZ pairs and 62 DZ pairs (Corby et al. 2007).

Possible mechanisms of heritable factors on the microbiota

Since the origin of cells, evolutionary development had to be influenced by the "microbial universe" connected to each organism. It is therefore likely that heritable forces shape the microbial composition in animals, including humans. Furthermore, one could speculate that the oral and axillary microbiota was under intense evolutionary selection due to its odor

production and association with mating preferences in hominids (Cho and Blaser 2012). Any two humans share an average of 99.9% of their DNA (Jorde and Wooding 2004). Some of the remaining 0.01% polymorphisms could be responsible for the composition differences of the human microbiota in various body habitats. Intuitively, immune system genes are excellent candidate genes. The majority of research on this topic has been done in the gut, which has been recently reviewed (Spor et al. 2011). The next section will give examples of mouse and human studies that support this hypothesis of a genetic effect.

TLR. As mentioned in Chapter 1.3.3, TLRs which are part of the innate immune system play an important role in recognizing bacterial cell components, PAMPs, and likely in distinguishing commensals from pathogens. TLR-5 recognizes bacterial flagellin, a conserved component of the bacterial flagellum. TLR-5 knockout mice exhibit a drastic change in beta diversity of the gut microbiota and suffer from metabolic syndrome. Transplanting the gut microbiota in wild type germ-free mice causes similar symptoms (Vijay-Kumar et al. 2010). Furthermore, a downstream signaling protein of TLRs in mice, MyD88, is necessary for *Aggregatibacter actinomycetemcomitans* induced tooth supporting bone loss (Madeira et al. 2013).

MEFV. However, knockout phenotypes are often more severe than naturally occurring genetic polymorphisms. Furthermore, deleterious gene mutations observed in mice may not transfer to humans. A rare example of the effect of a gene variant occurring in the human population that has an impact on bacterial composition is found in the C terminus of the gene MEFV. MEFV encodes the protein pyrin, which is one of the regulators of immunity through regulation of inflammation. It locates to the cytoskeleton, and is thought to regulate caspase-1, but the mechanism is unknown. MEFV mutations cause familial Mediterranean fever due to increases in IL-1 beta levels, which is characterized by short episodes of fever and localized inflammation with weeks or months of remission between attacks. Consequences of the mutation include increases in IL-1beta levels causing inflammation (Chae et al. 2009).

Associated with the disease episodes are large shifts of the composition of gut commensal bacteria (Khachatryan et al. 2008).

QTLs show a direct effect on the bacterial abundance in the murine gut. A quantitative trait locus (QTL) study in mice showed 13 significant and 5 suggestive QTLs out of 530 tested SNPs to co-segregate with the abundance of specific microbial taxa in the mouse feces. Each QTL accounted for 1.6–9.0% of phenotypic variation. A QTL which was associated with the abundance of the family Coriobacteriaceae and the genus *Lactococcus* included amongst others, a cluster of genes involved in immune function, namely Irak3 (encoding IL-1 receptor-associated kinase involved in the MYD88-dependent TLR pathway), interferon-gamma, Interleukin-22, and Lyz1 and Lyz2 (murine lysozymes) (Benson et al. 2010).

Example of periodontitis – Interleukin 1 and 10. Periodontitis is thought to be about 50% heritable based on twin studies (Michalowicz et al. 2000). The influence of host genetic variations of the IL-1 gene and IL-10 on severity of periodontal disease is suspected. IL-1 alpha is a pro-inflammatory cytokine, whereas IL-10 is an anti-inflammatory cytokine. Meta-analyses demonstrated an association of IL-1alpha as well as IL-10 variants with periodontitis (Karimbux et al. 2012; Zhong et al. 2012).

2.2 Sample description

2.2.1 Saliva collection and sample selection

The aim of this study was to compare the bacterial communities of MZ and DZ twins longitudinally. The subjects are part of the Colorado Twin Registry (Rhea et al. 2006; Rhea et al. 2013c) and Colorado Adoption Project (Rhea et al. 2013a; Rhea et al. 2013b), Institute for Behavioral Genetics, University of Colorado from 1997 to 2009 and were selected based on the availability of multiple saliva samples for the twins. The birth years of the sample subjects included are between 1976 and 1989. Informed consent was obtained from all individuals, and the anonymity of all subjects is ensured by separation of all personal information from the

sample and replacing it with a numerical identifier. Written informed consent was obtained and approved by the University of Colorado Human Research Committee (protocol 0399.11). Samples were collected after 2h of abstinence from eating. Ten milliliters of Scope mouthwash was vigorously swished in the mouth for a minimum of 30sec and released into a 50mL Falcon tube and stored at 4°C until extraction. The total DNA of the sample was isolated using the Puregene (QIAGEN) extraction kit, dissolved in TE buffer and stored at −80°C until needed.

2.2.2 Characterization of the dataset

The variability in the microbiome of 264 individual saliva samples derived from 107 individuals between the ages of 8 and 26 (Average age 16.3, 4.6 years standard deviation) was studied. Saliva samples were obtained from 27 MZ twin pairs, 18 DZ twin pairs, eight unrelated sibling pairs of adopted children, and one unrelated individual from the same cohort. Eighty-two individuals were sampled more than once approximately in 5-year intervals at up to three time-points (12/13, 17/18, and 22/23/24 years of age, labeled as 12, 17, and 22 years of age), see Figure 2.1.

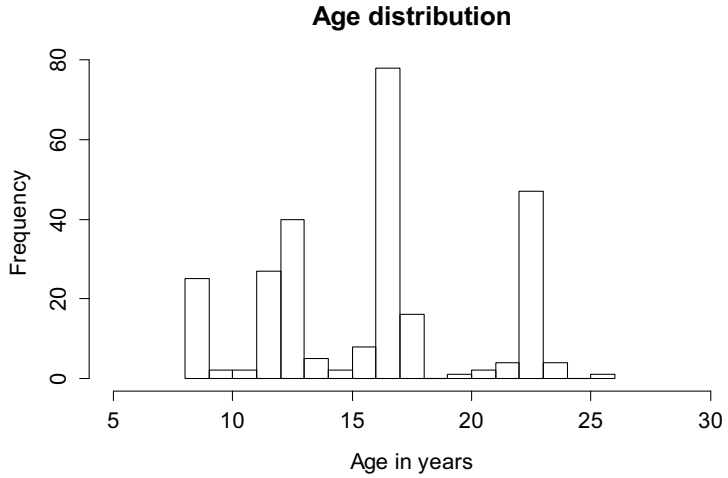


Figure 2.1: Age distribution in the twin dataset. In this histogram the three peaks of follow up years (age 12, age 17, age 22) are clearly visible.

The majority (93.5%) of the individuals identified themselves as Non-Hispanic Whites. Oral health information was not available, but the subjects rated their general health on a scale of 1-5 (excellent health, very good health, good health, fair health, poor health). Ninety-four percent (50 out of 52) reported their general health as good, very good, or excellent at age 22-24. Therefore, this study is based on a mainly white, healthy population. This profile is due to compliance rate in study enrollment, not ethnical selection. In the supplemental information of Rhea et al. (2006), Table 3B shows that the average of all births in the Colorado Front Range between the years of 1984 and 1989 were 77% Non-Hispanic White and 16% Hispanic White. In the Colorado twin cohort, 96% Non-Hispanic Whites and 3% Hispanic Whites were enrolled.

Body mass index measurements (BMI = (height in meters)^2/weight in kg) were available for 258 of the 264 (97.7%) data points. BMI was used to classify each individual into one of four weight groups, namely underweight, normal, overweight, or obese. During child development, the target range of healthy BMIs constantly changes. Therefore, the individual BMI with the corresponding ages were used to identify the target BMI using the United States CDC growth charts for boys or girls, respectively (Kuczmarski et al. 2000). Subjects that maintained a BMI below the 5th percentile were classified as being underweight. Those between the 5th and 85th percentile were considered normal and a BMI between the 85th and 95th percentile and obese was classified as overweight. Lastly, obese subjects were classified as such if their BMI was above the 95th percentile. For individuals older than 18 years of age the categories were as follows: a BMI less than 18.5 was considered underweight, between BMI 18.5-24.9 was considered normal, overweight was a BMI between 25-29.9, and a BMI over 30 was considered obese. The majority of individuals were in the normal weight range (68.6%), see Table 2.1.

Table 2.1: Demographic data of the twin dataset

Variable	Coding	n	%
Gender	Male	50	46.7
(107 individuals)	Female	57	53.3
Weight status	Normal	181	68.6
(264 samples)	Overweight	50	18.9
	Obese	18	6.8
	Underweight	7	2.7
	NA	8	3.0
Health status	1 (excellent health)	18	34.0
(at age 22-24, n= 53)	2 (very good health)	22	41.5
	3 (good health)	10	18.9
	4 (fair health)	1	1.9
	5 (poor health)	1	1.9
	NA	1	1.9
Race	Non-Hispanic Whites	99	92.5
(107 individuals)	Hispanic	6	5.6
	American Indian	1	0.9
	Multi Ethnic	1	0.9

2.3 Results

2.3.1 Sequencing results and taxonomic overview

PCR amplification and subsequent 454 pyrosequencing of the 16S rDNA gene hypervariable regions V1 and V2 of 264 samples resulted in 593,220 reads, which were quality-filtered and processed as described in Appendix A – Universal Methods. Of the barcoded reads, 427,189 were used for analysis after filtering. Samples with fewer than 698 sequence reads and internal controls were not included in the analysis. The average number of reads/sample were 1616.9 with a range between 698 and 3021 reads/sample. The average length of the sequence reads was 367 bp (range 200–513 bp). The dataset was grouped into a total of 678 OTUs, which belong to 10 phyla and 61 genera or higher taxonomic groups.

It has been previously described (Bik et al. 2010; Contreras et al. 2010; Lazarevic et al. 2010) that the main bacterial phyla in saliva were Firmicutes (with predominant genera *Streptococcus*,*Veillonella*, *Gemella*, and *Granulicatella*), Proteobacteria (*Neisseria*), Bacteroidetes (*Prevotella*), and Actinobacteria (*Rothia*). Less abundant phyla include Fusobacteria, TM7,

Cyanobacteria, SR1, Spirochaetes, and Tenericutes. The relative abundance of each phylum is highly variable between samples as Figure 2.2 shows.

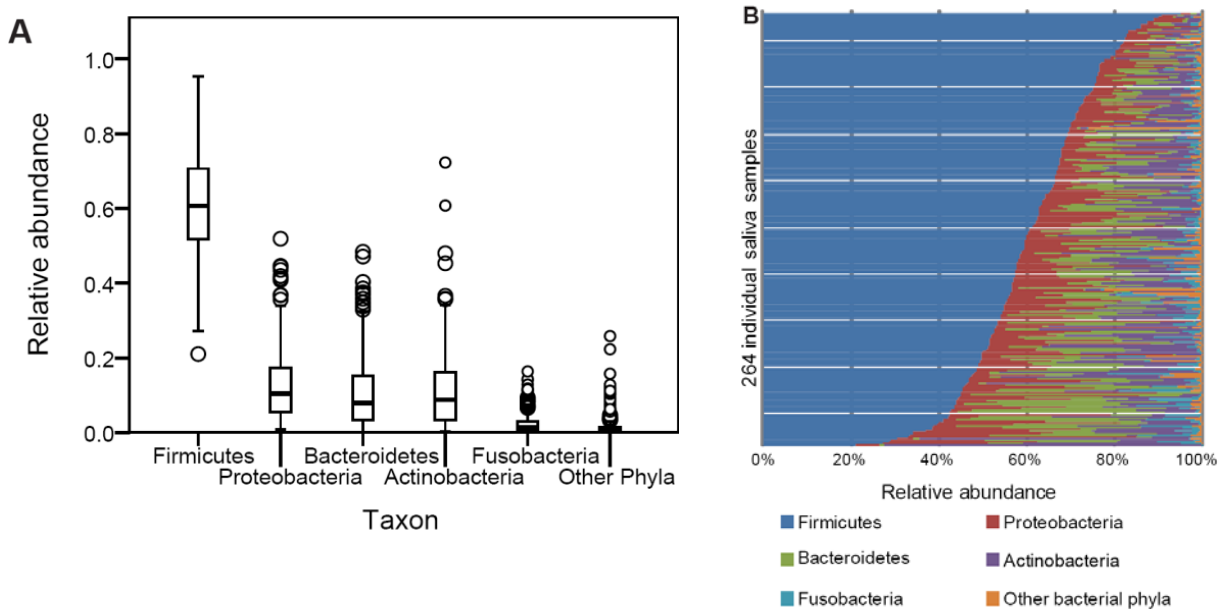


Figure 2.2: Phylum level variation between individual saliva samples. A) Box-and-whiskers plot of the five major bacterial phyla of the 264 human saliva samples. The top of the box represents the 75th percentile, the bottom of the box points to the 25th percentile, and the black line in the middle shows the median. The whiskers represent the highest and lowest values up to 1.5 times the interquartile range; extreme values and outliers are represented by empty circles. B) Relative abundance of the five major bacterial phyla of all individual saliva samples, sorted by decreasing Firmicutes content.

2.3.2 A core microbiota in human saliva

One of the questions, which this study attempted to answer, is whether a core oral microbiome exists. The oral cavity is a diverse environment with many micro-niches due to morphology, tissue types, pH, and saliva flow (Jenkinson 2011). Small studies of 3 to 10 individuals raised the possibility of a core, but the number of samples was too low to define a universal core oral microbiome (Zaura et al. 2009; Bik et al. 2010). This study with 264 samples derived from 107 individuals provides a better statistical base. At the time of completion, the sample set was the largest salivary microbiota study derived from next generation 16S rRNA

genes sequences. No universal definition of a microbiota core exists. I defined a core salivary microbiome at the genus-level based on the percentage of samples in which each genus was found. Some rare genera might have been missed due to incomplete sampling (Sogin et al. 2006). This is in contrast to other human body habitats, such as the gut where no core on the genus level exists (Turnbaugh et al. 2009). Eight genera were observed in >95% of all samples (*Streptococcus*, *Veillonella*, *Gemella*, *Granulicatella*, *Neisseria*, *Prevotella*, *Rothia*, *Fusobacterium*); an additional 13 in >50% (Fig. 2.3).

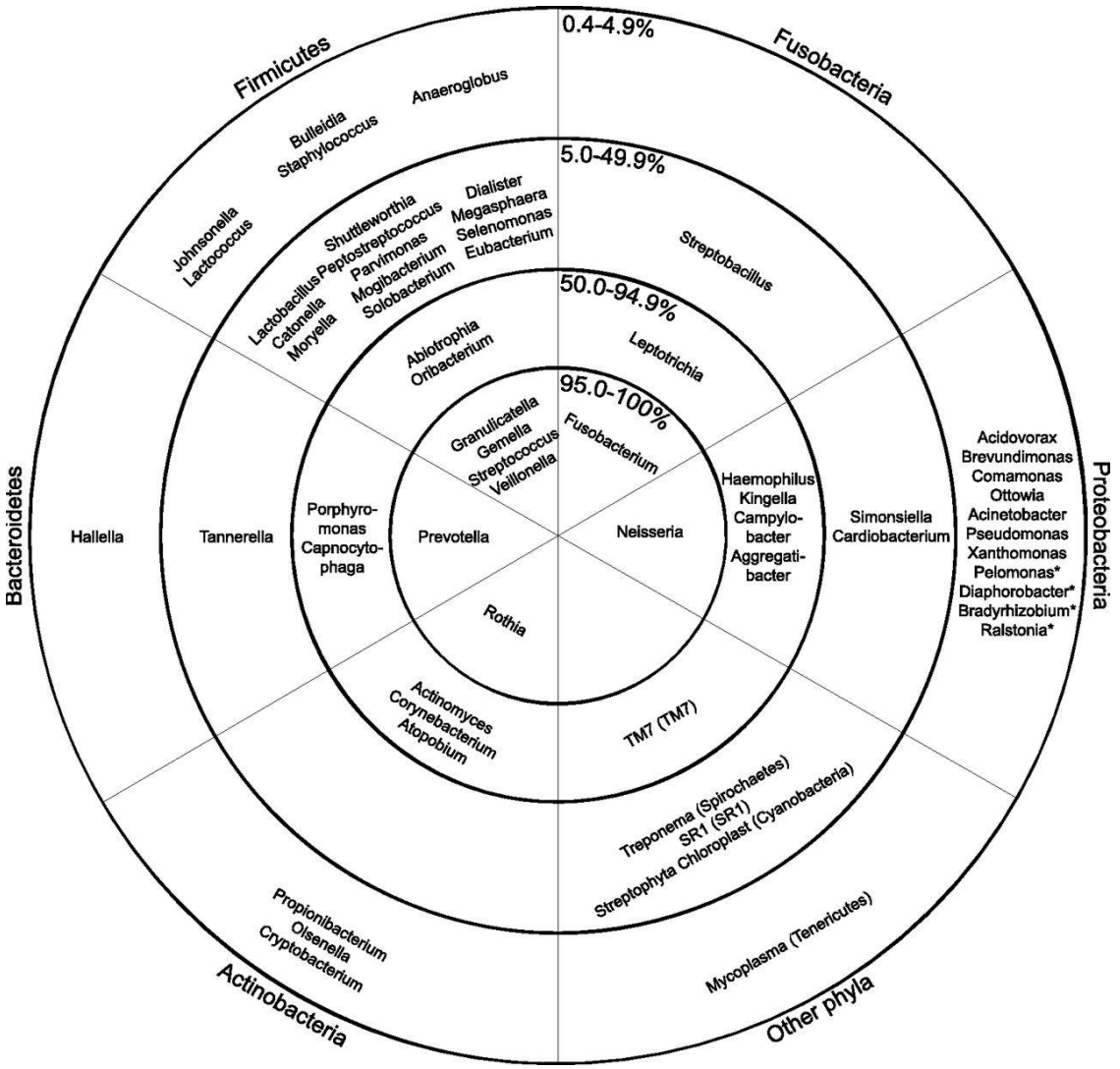


Figure 2.3: The "core" - Presence of bacterial genera based on occurrence in samples. Taxonomic identity is based on RDP classification (Cole et al. 2007). The rings represent the percentage of

samples where a given genus was observed (0.4%–4.9% means that genus was found in 1–12 samples, 5.0%–49.9% means that genus was found in 13–131 samples, 50.0%–94.9% means that genus was found in 132–250 samples, and 95.0%–100% means that genus was found in 251–264 samples). The pie slices subdivide the chart into the various bacterial phyla. Genera marked with an asterisk (*) have only been found in one sample.

While the number of members of the core is relatively small with 8 out of 61 total genera, they comprise on average 81.3% of sequence reads in each sample.

At a 97% identity level of OTUs, two OTUs (OTU 1 and OTU 578), both of which belong to the genus *Streptococcus*, were found in all but one sample each (99.6% of samples, 10.5% of all sequence reads). The species of OTU 578 could not be identified with certainty. The closest BLAST match to the Human Oral Microbiome Database (Chen et al. 2010) for OTU 1 was *Streptococcus mitis* (99.7% identity), the second most common bacterium isolated from the oral cavity by molecular cloning based on the HOMD database (accessed August 31, 2011). *Streptococcus mitis* was also the only oral bacterium found on all oral surfaces from at least four of five individuals examined in Aas et al.(2005). This result shows that the salivary microbiota is dominated by a small number of OTUs. The 42 OTUs shared across > 80% of samples comprise more than 50% of all sequence reads. A detailed list of all OTUs including DNA sequencing can be found in the supplemental information of my published study (Stahringer et al. 2012).

Many members of the core are well known oral bacteria important for biofilm succession as discussed in Chapter 1.3.2 (Biofilm). *Streptococcus* and *Actinomyces* (found in 94.7% of samples) are the two initial colonizers, which are able to bind to the salivary pellicle on teeth. Early colonizers, including four of the core genera (*Rothia*, *Prevotella*, *Neisseria* and *Veillonella*) are able to bind to the initial colonizers. Another member of the core, *Fusobacterium*, attaches to the fresh biofilm as the bridging organism to allow binding of later colonizers, including oral pathogens (Kolenbrander 2011b, pg. 5-8). The role of *Granulicatella* and *Gemella* in the oral microbial communities is still unclear, but they might have a similar keystone function in the colonization of novel biofilms.

2.3.3 Monozygotic twins are statistically not more similar to each other than dizygotic twins

Is the composition of the oral microbiome heritable? To answer this question, a twin design was utilized. Comparison of the sharing of microbiota of MZ and DZ twins permits a powerful assessment of heritability (i.e., the influence of the human genotype on phenotype). MZ twin pairs, who share 100% of their alleles, are expected to have oral microbiomes that are more similar to each other than do DZ twin pairs, who share on average 50% of their alleles. I compared 59 MZ and 39 DZ same-aged twin saliva sample pairs obtained between the ages of 12, 17, and 22. The metric used for comparison was the unweighted UniFrac distance (Lozupone and Knight 2005; Lozupone et al. 2006). A slight trend toward more similarity among MZ pairs than DZ pairs is observable (Figure 2.4), but as previously shown in the gut (Turnbaugh et al. 2009), this difference is not statistically significant. This observation could be due either to a small genetic influence relative to overall variation or to other cofounding effects. However, if I compare both MZ and DZ pairs to unrelated individuals who live in different homes at the same age, the difference becomes highly significant. Because the MZ–DZ comparison was nonsignificant, MZ and DZ twins were pooled together for all following analyses.

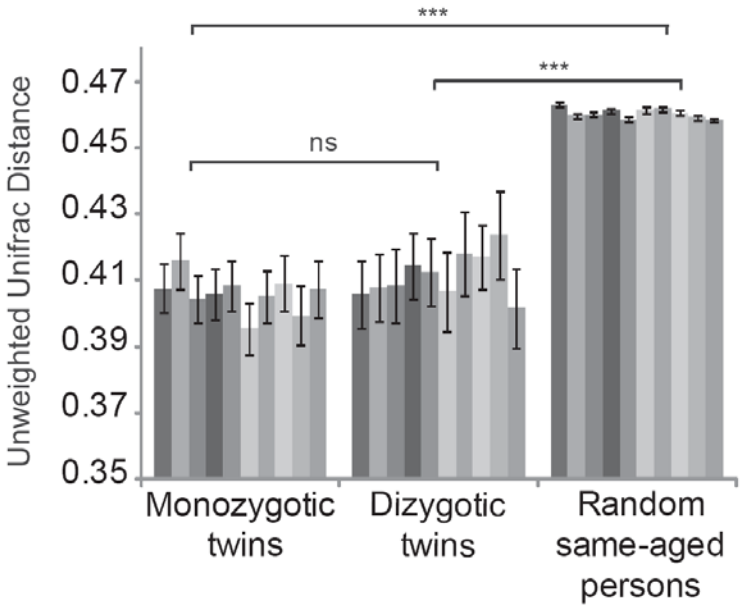


Figure 2.4: Genetic effect on the salivary microbiome at the ages of 12, 17, and 22 years. Averaged pairwise unweighted UniFrac distances of same aged MZ (n = 59) and DZ (n = 39)

twin pairs and same aged sample population (n = 7882 pairs) at ages 12–13, 17–18, and 22–24 (± standard error of the mean - SEM). The data set was randomly subsampled 10 times at a sequencing depth of 800 sequences/sample, and each subsampling is shown as a separate bar. The statistical analysis was a Mann-Whitney U-test. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10 rarefactions is presented.

2.3.4 Twins become more dissimilar when moving apart

The longitudinal design allows for the assessment of changes in development during adolescence and young adulthood. To detect patterns of dynamic, temporal changes in the microbiome during adolescence, I analyzed the salivary microbiome of 82 individuals over time (198 saliva samples). I assessed the unweighted UniFrac distance between twin pairs at the ages of 12, 17, and 22 and compared them to unrelated individuals at the same ages. At each time point, the salivary microbiota of twins is significantly more similar than the random population. At the earlier time points, age 12 and age 17, when it is assumed that the twins are still cohabitating, the similarities are approximately equal. However, at age 22, when most twin pairs move apart and do not share the same houshold, the differences between the pairs increases, even if not signficantly (Figure 2.5).

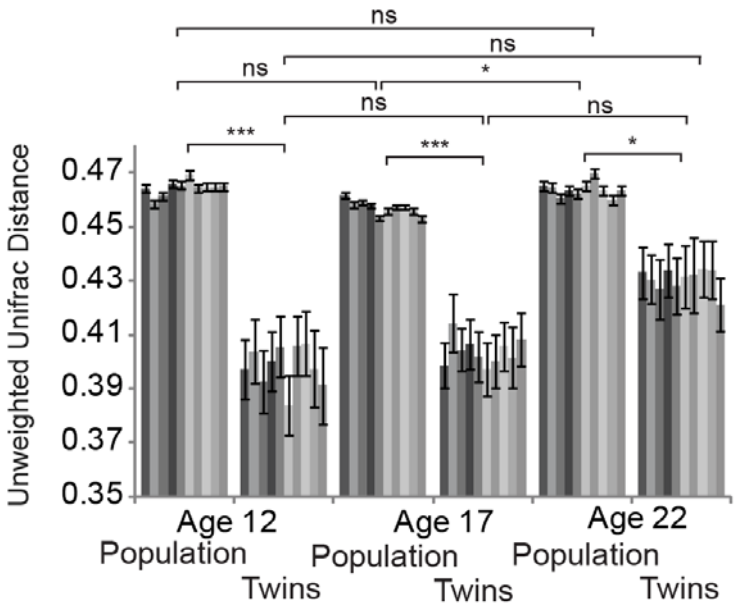


Figure 2.5: Similarities based on age of sampling (ages 12, 17, and 22 years) of twins and the sample population. Pairwise unweighted UniFrac distances twin pairs (ages 12: n=30, 17: n=40, and 22: n=26) and the sample population at different ages (ages 12: n=2211, 17: n=4186, and 22: n=1485) were calculated (±SEM). The data set was randomly subsampled 10 times at a sequencing depth of 800 sequences/sample, and each subsampling is shown as a separate bar. The statistical analysis was a Mann-Whitney U-test. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10 rarefactions is presented.

I assessed whether the twin pairs were living apart at age 22 – 24 and the information was not available for each pair. Twenty-one pairs were living apart, 1 pair (S14968 and S14969) was living with their parents and the living status for 4 pairs was unclear (S14895 and S14922, S14896 and S14897, S14973 and S14974, S14976 and S14977). The average unweighted UniFrac distance for the pairs living apart was 0.432, the average unweighted UniFrac distance for the pair which certainly lived together at that time was 0.374. As expected, their oral communities are more similar. One individual (S14896) of the pair (S14896 and S14897) reported to be in poor health and lived with her parents, whereas the other individual (S14897) reported to be in excellent health and information about her living arrangements were not available. Their oral microbiotas are distinct with an UniFrac distance value of 0.520. This high beta diversity value

indicates that they might not have been living together at the time of sampling or that the disease caused a change of microbial composition in the sick twin. If I exclude these two pairs, the average UniFrac values of the remaining 3 pairs with unclear living arrangement is 0.430, which is slightly lower than the average UniFrac distance for pairs living apart.

Another finding was a slightly signficant increase in average beta diversities from age 17 to age 22 in the random population. It is likely that an exposure to different microbial sources and other influencing factors cause an increase in beta diversity when reaching young adulthood and independence from parents' homes and supervision.

2.3.5 Temporal intra-individual similarity during two 5 year spans

This sample set allows for a comprehensive temporal study of the oral microbiome of the same individuals spanning a period of over 10 years. In the same dataset as above I compared each individual to itself and to its twin sibling at a later time-point. I also compared the salivary microbiome of the cohort population of the same age from age 12 to age 17, from age 17 to age 22, and from age 12 to 22, spanning a period of 5 and 10 years. After both 5-year spans the oral microbiome of an individual resembles itself more closely than that of the population based on unweighted UniFrac distances. Therefore, even in the human oral microbiome where one may anticipate frequent environmental perturbations there is remarkable stability over long time periods during development up to 5 years. After 10 years (from age 12 to 22) the oral microbiome still has a trend toward self-similarity, but this trend is not statistically significant (Figure 2.6). This finding is supported by a DGGE based study which has shown stability in one individual over a time span of 7 years (Rasiah et al. 2005).

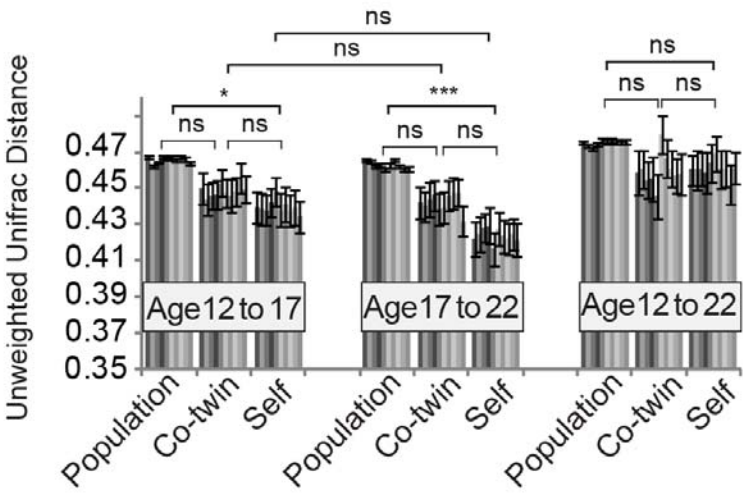


Figure 2.6: Time progression of individuals, their twins, and the sample population at ages 12, 17, and 22. Age 12–17: Population (n = 6165 pairs), Co-twin (n = 58 pairs), Self (n = 64 pairs); age 17 to 22: Population (n = 5060 pairs), Co-twin (n = 50 pairs), Self (n = 52 pairs); Age 12 to 22: Population (n = 3685 pairs), Co-twin (n = 28 pairs), Self (n = 34 pairs). Error bars: ± SEM. The data set was randomly subsampled 10 times at a sequencing depth of 800 sequences/sample, and each subsampling is shown as a separate bar. The statistical analysis was a Mann-Whitney U-test. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10 rarefactions is presented.

Within the twin sample, I compared the oral microbiome of one individual twin at a younger age with the oral microbiome of the co-twin at an older age (e.g., twin A at age 12 to his co-twin B at age 17). I found that comparing microbiomes at age 12 to age 17 between the twins is statistically no different than comparing the microbiome in the same individual going from age 12 to age 17. The similarity across the twin pairs is reduced between the ages of 17 and 22, when at least 21 out of 25 twin pairs (84%) no longer cohabitate. It should be noted that changes in individuals occurring between age 17 and 22 tend to be less extreme than changes seen between age 12 and 17. In this time interval, there are significant developmental changes that occur, e.g., puberty (Guncu et al. 2005) or behavioral changes, which could be the contributing factors.

2.3.6 Age effects on the oral microbiota

Bacterial taxa change with age

Besides exploring how whole bacterial communities change over time using the UniFrac distance metric, it is important to test for changes in taxa at different levels. To account for differences in sample numbers at each age and to aid in visualization, I grouped the individual samples into four age groups (ages 12–14, ages 15–17, ages 18–21, and ages 22–24) for an ANOVA. The Pearson product-moment correlation to the actual ages and ANOVA to the age groups were applied. All reported significant families and OTUs were recovered from both statistical tests with a $p < 0.05$ with Bonferroni correction. On a bacterial family level, a negative correlation with the Pearson product-moment correlation on the actual ages of Veillonellaceae (Firmicutes, Pearson product-moment correlation coefficient [Pearson's r] = -0.28700, Bonferroni corrected $p = 3.98 \times 10^{-4}$) and a positive correlation of Actinomycetaceae (Actinobacteria, Pearson's r = 0.32188, Bonferroni corrected $p = 3.02 \times 10^{-5}$, Figure 2.7 A) with age was observed. There is a significant and substantial increase in the proportion of Actinomycetaceae with age, especially in young adults at age 22–24. Actinomyces species have been found preferentially in early-stage caries in children and young adults (Aas et al. 2008), and this may reflect a general decline in dental health with age. Even though *Actinomyces* and *Veillonella* have been shown to coaggregate (Shen et al. 2005), their abundance changes do not follow the same pattern in my study. In addition, there are OTUs at a 97% identity level that are positively and negatively correlated with age in the genera *Actinomyces*, *Veillonella*, and *Streptococcus* (Figure 2.7 B, C, and D).

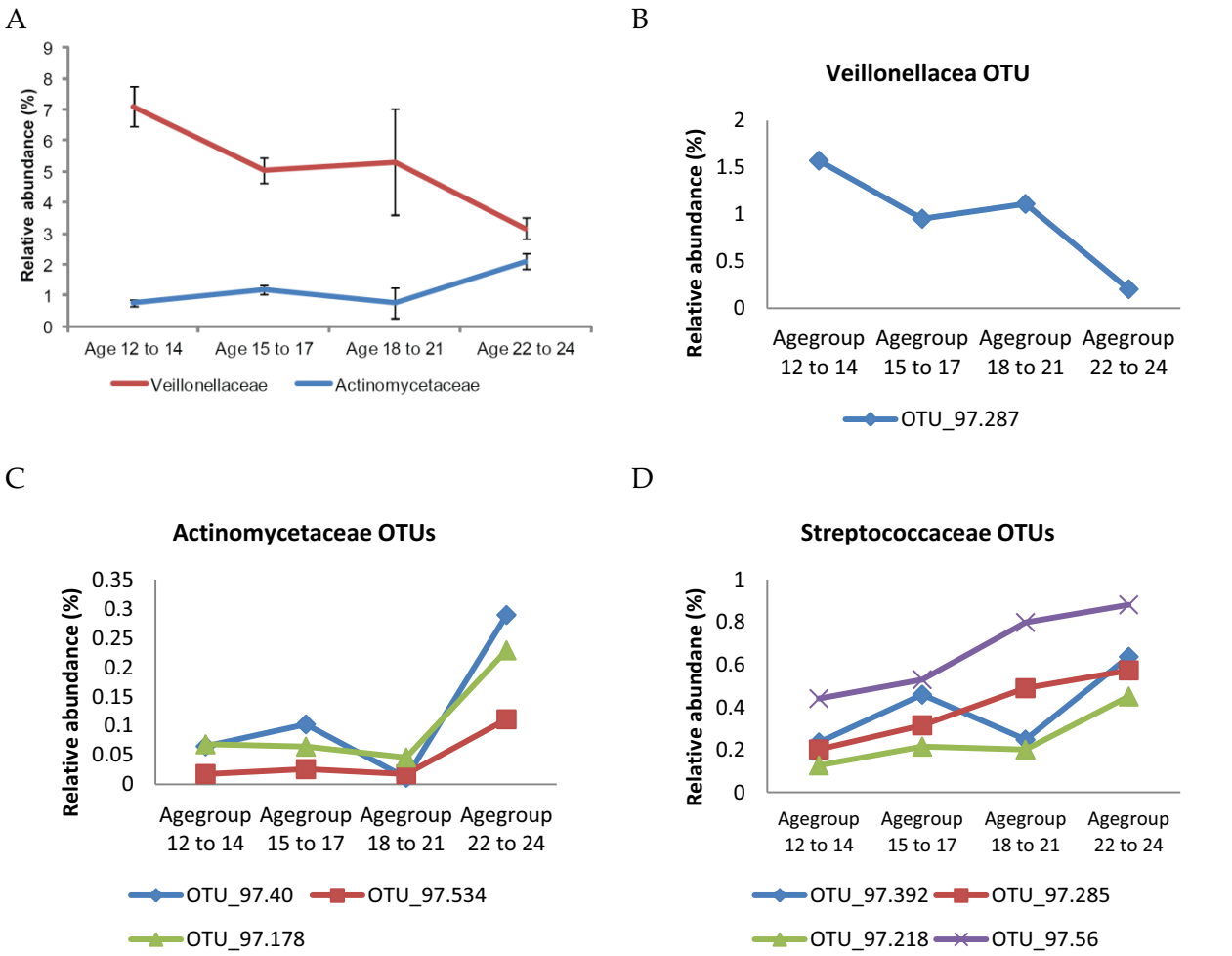


Figure 2.7: A. Relative abundance of bacterial families Veillonellaceae and Actinomycetaceae and their negative and positive correlation with age from adolescence to early adulthood. Error bars: ± SEM. B-D. Relative abundance or bacterial OTUs in the families A) Veillonellaceae, B) Actinomycetaceae, and C) Streptococcaceae. Sample sizes: Age 12 to 14: n = 72, age 15 to 17: n = 88, age 18 to 21: n= 19, age 22 to 24: n = 55. All of the reported significant Pearson product-moment correlation coefficients were at a p-value of <0.05 with Bonferroni correction for multiple testing.

Table 2.2 shows the exact p-values with Bonferroni correction from a Pearson correlation, the Pearson product moment correlation coefficient, and the closest BLAST match in the HOMD database (Chen et al. 2010). However, the species assignments are not reliable at the current length of sequence reads, especially streptococci which are known to have a very similar 16S rRNA gene sequence unable to discriminate different species (Kawamura et al. 1995).

Table 2.2: OTUs correlated with age during adolescence and young adulthood. This table lists OTUs that were found to be significantly correlated with age. It states the p-value with Bonferroni correction from a Pearson correlation, the Pearson product moment correlation coefficient, and the closest BLAST match in the HOMD database (Chen et al. 2010).

OTU Number	p-value (Bonferroni corrected)	Correlation coefficient	Closest BLAST match (HOMD, (Chen et al. 2010))
OTU_97.285	0.000582	0.32922	*Streptococcus infantis/ S. sp.*
OTU_97.287	0.001522	-0.30078	*Veillonella dispar/ Veillonella parvula*
OTU_97.218	0.00204	0.29855	*Streptococcus australis*
OTU_97.534	0.003468	0.306533	*Actinomyces odontolyticus*
OTU_97.392	0.00416	0.29686	*Streptococcus peroris/ S. mitis*
OTU_97.56	0.012416	0.285265	*Streptococcus sp.*
OTU_97.40	0.020507	0.27445	*Actinomyces odontolyticus*

A similar cross sectional study looked at the abundance of bacterial taxa in school children, ages 3-18 and found partly conflicting results (Crielaard et al. 2011). They observed an increase in Veilloneallaceae, whereas my study shows a decrease with age. However, their study focused on the progression from deciduous (milk) dentition via mixed dentition to permanent dentition, whereas my study cohort is assumed to have mostly a full permanent dentition by age 12.

Pre-adolescent exhibit decreased diversity. Age effect or sampling effect?

While the full study includes samples from individuals as early as 8 years of age (Figure 2.1), these samples were not used for any part of the presented analysis in this chapter except to show inter-individual variation and the core. Because this sample set was not collected with microbiome analysis in mind, discrepancies in sampling methods could not be excluded. Recall from the sampling staff at the Institute of Behavioral genetics suggests that individuals 11 years or younger (years 1997 and 1998) have not received Scope mouthwash; instead, they were likely sampled with a cotton swab across the mucosa. A preliminary study indicates that different

sampling methods do not have a directed bias (see Appendix B: The effect of sampling and extraction methods on the oral microbiota).

A principal coordinate analysis (PCoA) plot of the unweighted beta-diversities showed that samples obtained from these children cluster very closely together indicating a high similarity between children, but not adolescents (Figure 2.8).

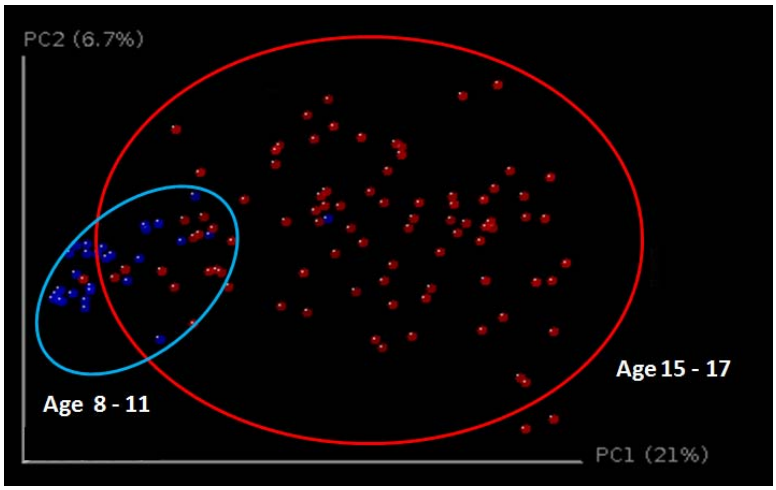


Figure 2.8: Close clustering of children age 8-11 (blue) compared to adolescents age 15-17 (red) using PCoA of the unweighted UniFrac distance matrix. Each point corresponds to a sample and the distances correspond to beta diversity values. The percentage of variation explained by the principal coordinate 1 (PC1) and 2 (PC2) is indicated on the axes.

This finding is intruiguing and requires further investigation. Large-scale projects such as the HMP will soon be expanded to include different states beyond the healthy adult cohort, and will be able to answer the question of whether the effect I saw in this dataset is real. Given the circumstances of uncertainty about the sampling method of these samples, I do not want to attempt to interpret this finding further.

2.3.7 The effects of obesity on the oral microbiota

The interplay between human weight (lean vs. obese) and the gut microbiome is under intense investigation (Ley et al. 2005; Ley et al. 2006b; Turnbaugh et al. 2006; Turnbaugh et al. 2009; Vijay-Kumar et al. 2010; Ridaura et al. 2013). However, very little is known about the oral microbiome in this context. There is evidence of association between obesity and periodontal disease (Linden et al. 2013) and an independent increase of particular oral bacteria, *Tannerella forsythia* (Haffajee and Socransky 2009) or *Selenomonas noxia* (Goodson et al. 2009), with BMI although there has been no broad 16S rDNA sequencing approach applied to compare phylum level correlations. To test for associations between saliva microbiome and BMI, I divided subjects into underweight, normal, overweight, and obese weight classes for boys and girls based on the age-appropriate target BMI (Kuczmarski et al. 2000). In contrast to the gut microbiome (Ley et al. 2006), I found no significant correlation between any OTUs at a 97% or higher identity taxonomic level and overall weight class in human saliva in this sample set of adolescent and young adult individuals.

CHAPTER III

SALIVARY MICROBIAL COMMUNITIES DIFFER BY AGE AND SMOKING BEHAVIOR

3.1 Introduction

Chapter 2 has shown that environmental influences such as cohabitation have a significant influence on the microbiota of the oral cavity. However, it is unclear what these shared environmental factors are. Chapter 3 aims to examine internal and external factors that have the potential to influence the salivary microbiota, namely age, smoking, alcohol consumption, and stimulant usage. In this chapter, I discuss the results from a bacterial 16S rRNA gene based study of 210 saliva samples obtained from individuals age 12 to 65, who exhibit varying degrees of smoking, alcohol and stimulant consumption. The first subchapters present possible effects of environmental variables with emphasis on tobacco smoking, alcohol consumption, and methamphetamine as an example for stimulant use. The result section provides evidence of an effect of tobacco smoking and age on the beta diversity of the salivary microbiome and the fitness of certain bacterial genera, including *Neisseria*, which decreases in smokers, as well as *Fillifactor*, which thrives in smokers. The effects of alcohol and stimulants are not discussed due to dominant effects of smoking and age.

3.1.1 The effect of environmental variables on human microbial communities

The human oral microbiota underlies daily fluctuations (Caporaso et al. 2011). This is likely due to a combination of factors that expose the individual to new subsets of microbes and exhibit selective forces on individual microbes. At this point, it is unknown what factors influence the microbiota, but oral hygiene, diet, and life style are good candidates.

Oral hygiene, which includes daily tooth brushing and flossing, has the aim to reduce the biofilm on teeth temporarily to prevent the accumulation of potential pathogenic species in the maturing biofilm as discussed in Chapter 1.3.2. However, the effect of undisturbed supragingival plaque which can be managed by non-professional dental care on the salivary microbiota might be limited. It has been shown for children (Papaioannou et al. 2009) and adults (Zaura et al. 2009; Segata et al. 2012) that plaque communities cluster separately from mucosal or salivary communities. Some transfer of bacteria from plaque bacteria to saliva is possible, but has not been quantified to my knowledge.

It has been shown in the gut microbiota, that diet has a significant influence on infants and adults. In the developing infant gut, the introduction of solid food causes a dramatic shift in the gut microbiota (Koenig et al. 2011). Research on the adult gut microbiota has demonstrated that short-term dietary changes alter the abundance of certain bacteria taxa, but long-term dietary changes are required for substantial alterations of the bacterial community (Wu et al. 2011). Long-term dietary habits can lead to horizontal gene transfer of specialized genes for digestion of complex dietary polysaccharides. Seaweed is a traditional food product in the Japanese population and is regularly consumed. Hehemann et al. have identified porphyranases and agarases that aid in the digestion of seaweed in a Japanese population, but these genes were absent in the gut microbiota of North Americans (Hehemann et al. 2010). Geographic gut microbiota differences have been shown in a study comparing individuals with very different diets and life styles from the US, Malawi, and Venezuela (Yatsunenko et al. 2012).

The oral microbiota is also likely influenced by dietary habits. It is well known that a diet high in fermentable carbohydrates promotes acidification of plaque due to the presence of acidogenic and aciduric organisms (e.g. *Streptococcus mutans*) which are involved in tooth decay (Wilson 2005, pg. 364). The infant oral microbiota is also directly influenced by diet, as differences between breastfed and formula-fed infants have demonstrated (Holgerson et al. 2013). Changes in the oral microbiota have even been shown for gross dietary changes during

human evolution. Ancient calcified dental plaque from skeletons between the Mesolithic, Neolithic, Bronze Age, and Medieval period could show a shift in diversity and composition from the hunter-gatherer to the farming life style, and a clear difference to the modern microbiota with an abundance of tooth decay causing *Streptococcus mutans* (Adler et al. 2013). Individual food products high in plant polyphenols such as tea, cocoa, and wine have been proposed to influence the oral microbiota (Ferrazzano et al. 2009; Signoretto et al. 2010; Daglia et al. 2011; Gazzani et al. 2012). In contrast, a study of 10 individuals, each from 12 different countries worldwide, failed to show geographical clustering in the salivary microbiota (Nasidze et al. 2009a). However, the study was based on 120 sequences per sample, which would have missed rare species. In addition, it was not discussed whether all individuals lived a similar Western lifestyle, which would explain the lack of geographic differences.

Furthermore, the composition of the oral microbiota might be influenced by lifestyle including non-food consumed substances. The next subchapters will discuss the potential effect of recreational drug use (tobacco, alcohol, stimulants) on the oral bacterial communities.

3.1.2 The effect of smoking

Significance – health and economical impact. Active smoking, alcohol consumption, and passive smoking are the three leading preventable causes of death in the US (Arbes et al. 2001). Smoking tobacco is detrimental to the health of the smoker and people exposed to the tobacco smoke (passive smoke). Smoking tobacco is the leading health risk factor in high-income North America and Western Europe, listed before high blood pressure, overweight, high fasting glucose levels, and physical inactivity (Lim et al. 2012). On a national level, it caused 443,000 premature deaths yearly with 49,000 of those deaths caused by second hand smoke (2000-2004) (Roger et al. 2012). That corresponds to about 1 in 5 deaths, which can be attributed to tobacco smoking in the US. Male smokers die on average 13.2 years earlier, whereas female smokers die on average 14.5 years earlier compared to their sex matched

nonsmokers (Roger et al. 2012). In 2010, 19.3% of the adult US population (≥ 18 years) were current cigarette smokers (female 17.5%, male 21.2%) (Roger et al. 2012). The estimated yearly cost of smoking totals $96 billion in direct medical costs and $97 billion in productivity loss to a total of $193 billion (2000-2004) (Roger et al. 2012).

Smoking is not only a developed world problem; in contrast, it is a global public health issue. Worldwide, smoking tobacco with the inclusion of second hand smoke is the second leading risk factor for global disease burden; 5.7 million deaths worldwide were attributed to smoking and second hand smoke and accounted for 6.3% of all death and disability-adjusted life years in 2010 (Lim et al. 2012). Smoking is prevalent in all ages. Smoking and second hand smoke are the second leading risk factors for both sexes in all age groups studied, 15-49 years of age, 50-69 years of age, as well as 70 years and older in 2010 (Lim et al. 2012). However, there are gender differences. On a global scale, tobacco smoking is a more serious problem in men than in women. While smoking, including second hand smoke, is the leading risk factor for disability, life adjusted for men (8.4% of total), it is the fourth highest risk factor for women (3.7%) (Lim et al. 2012).

Effect on the human body. Tobacco is the drug with the most frequent usage pattern. Many smokers expose themselves to tobacco smoke voluntarily several times a day, every day, for many years or even decades (Palmer et al. 2005). The exposure to nicotine, the psychoactive component in cigarette smoke, is the main compound for the addiction potential of tobacco leading to chronic, long-term use (Palmer et al. 2005; Swan and Lessov-Schlaggar 2007). Nicotine is readily absorbed from the lungs and then transported to the location of action, the brain (Palmer et al. 2005). Nicotine activates the nicotinic acetylcholine receptors leading to a short time enhancement of cognitive performance (Swan and Lessov-Schlaggar 2007).

Besides nicotine, there are about 4000 components in cigarette smoke, many of which are toxic or carcinogenic (Smith and Hansch 2000). While the smoke is inhaled through the oral cavity, trachea and ultimately the lungs, many negative health outcomes do not originate from a

topical exposure, but a systemic distribution of the toxic compounds in the human body (Palmer et al. 2005).

Today, it is well known that smoking is the leading cause of preventable disease and deaths in the US and a serious problem worldwide. However, this is not a new finding. The negative impact on human health has been observed as early as 1604. King James I wrote in the "Counterblaste to Tobacco" about smoking: "a custom Lothesome to the Eye, hatefull to the Nose, harmfulle to the Braine, dangerous to the Lungs" (Northrup 1957).

Consequences of cigarette smoking are an increase in chronic conditions and disease of multiple organs. This includes cardiovascular diseases and coronary heart disease (such as angina, myocardial infarction, sudden death, and congestive heart failure), cerebrovascular disease (such as transient ischemic attacks, stroke), vascular disease (such as claudication, aortic aneurysm, and atherosclerosis), impotence, and hypertension (Swan and Lessov-Schlaggar 2007; Bagaitkar et al. 2008). It is also the primary cause of chronic-obstructive pulmonary disease (COPD), such as mucous hypersecretion, interference with ciliary function, and alveolar destruction (Swan and Lessov-Schlaggar 2007). Premature death is mainly due to various cancers, especially lung cancer, stroke and ischemic heart disease, as well as COPD, see Figure 3.1 (Adhikari 2008; Rostron 2013). Other fatal outcomes include premature birth and low birth weight of neonates born to smoking mothers, sudden infant death syndrome (SIDS), pneumonia, pancreatic and esophageal cancer, cancer of the upper respiratory tract as well as residential fires (Adhikari 2008; Bagaitkar et al. 2008).

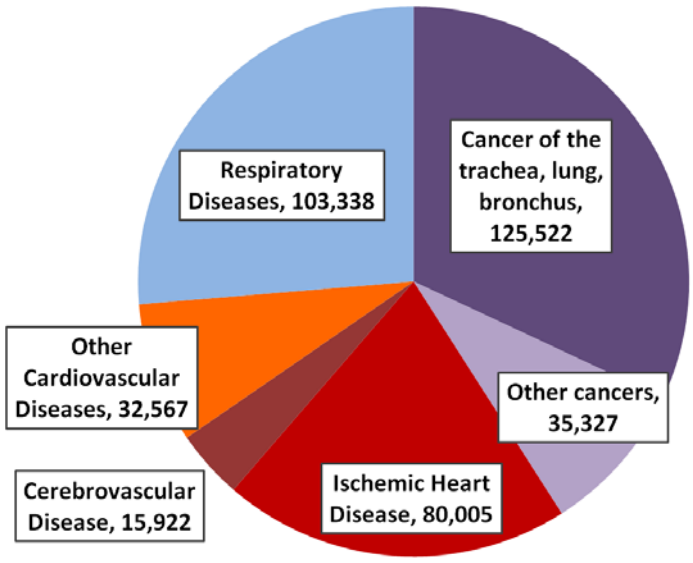


Figure 3.1: Average number of deaths attributable to cigarette smoking in the US (2000-2004) Data from Adult SAMMEC — Five-Year Report (Centers for Disease Control and Prevention, accessed Sept. 2013)

Cancer. Tobacco smoke contains at least 4000 different components including chemicals that were declared Group I carcinogens by the International Agency for Research on Cancer such as benzene, arsenic, chromium, 2-naphthylamine, vinyl chloride, 4-aminobiphenyl, and beryllium (Smith and Hansch 2000). Tobacco smoke also contains acetaldehyde, which is likely to contribute to carcinogenesis in the oral cavity (see Chapter 3.1.3 on alcohol). The average concentration of acetaldehyde in tobacco smoke is 709ug per cigarette (Smith and Hansch 2000).

Smoking and bacterial diseases. In addition to the large number of diseases caused or aggravated by smoking, active smokers and individuals exposed to second hand smoke have an elevated risk of bacterial infection (Bagaitkar et al. 2008). As expected, bacterial infections in the respiratory tract and nasopharynx are increased. Smokers harbor an increased number of pathogens in their nasopharyngeal tract, for example, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Streptococcus pyogenes* (Bagaitkar et al. 2008). Smokers are at an increased risk to develop bronchitis, bacterial pneumonia, Legionnaires disease, as well as tuberculosis and meningitis (Bagaitkar et al. 2008). An increase in post-surgical and nosocomial infections, *Helicobacter pylori* infections, meningitis, otitis media and also diseases of the

reproductive organs such as bacterial vaginosis and sexually transmitted diseases (chlamydia and gonorrhea) have been observed in smokers (Bagaitkar et al. 2008). In addition, periodontitis incidences are more frequent and more severe in smokers, as discussed in the next paragraphs (Bagaitkar et al. 2008).

Effect on oral health. Smoking is the most important environmental risk factor for periodontitis (Palmer et al. 2005). 41.9% of periodontitis can be attributed to smoking (Kumar 2012). The form of periodontitis observed in smokers is more severe, with deeper pockets, more serious attachment loss due to destroyed collagen to the root surface and bone destruction (Barbour et al. 1997; Kinane and Chestnutt 2000; Kumar 2012). Periodontal treatment often fails in smokers, and 80-90% of treatment failures occur in smokers (Kinane and Chestnutt 2000). The increase in periodontitis caused by smoking is usually estimated to be in the 2-6 fold range, some even say even up to 15 fold, in a dose and exposure length dependent matter (Barbour et al. 1997; Kinane and Chestnutt 2000; Kumar 2012). This increase in risk is independent of oral hygiene practices as a study in dental hygienists showed (Kinane and Chestnutt 2000). The amount of exposure to nicotine can be established by measuring cotinine, a nicotine metabolite with a longer half-life, in plasma or saliva. This is a better estimator of exposure than self reported cigarette exposure (Barbour et al. 1997; Kinane and Chestnutt 2000; Palmer et al. 2005). Typical cotinine concentrations of smokers are > 14 ng/ml, but could be as high as 1000 ng/ml (Palmer et al. 2005). Nicotine levels are typically in the 5-50 ng/ml range (Palmer et al. 2005).

Causes of oral health deterioration. On a whole body level, three potential mechanisms for an increased risk of infection in smokers have been proposed (Bagaitkar et al. 2008):

1. Physiological and structural changes in humans
2. Dysregulation of immune function
3. Increase in bacterial virulence

These mechanisms can lead to tissue damage and compromised tissue repair (Barbour et al. 1997).

1. Physiological and structural changes in humans

Effect of smoking on the human body. Physiological and structural changes in humans have been described primarily in the respiratory tract and vasculature. In the respiratory tract, ciliary function is impeded which results in an inadequate clearance of bacterial biofilm (Bagaitkar et al. 2008). Vasoconstriction is observed in peripheral blood vessels, whereas vasodilation is observed in the brain (Bagaitkar et al. 2008).

Effect of smoking on the periodontal tissue. As mentioned before, periodontal disease is more severe in smokers. However, smokers also exhibit reduced clinical inflammation, see Figure 3.2 A (Kinane and Chestnutt 2000). Diminished inflammatory response and impaired wound healing are two possible reasons for this contradictory observation on the host side.

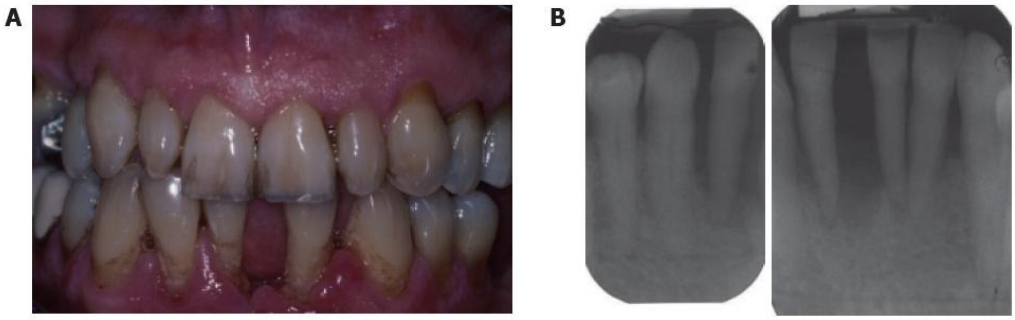


Figure 3.2: Dental status of a 55 year old male smoker. A. While the teeth are visibly neglected (plaque, staining, receding gingiva), the upper gingiva does not exhibit any clinical signs of inflammation. There are some inflammation symptoms (redness, swelling) visible on the lower gingival. B. The bone loss is clearly visible. The remaining lower incisors are inevitably lost. Photograph and x-rays from Scott and Martin (2006).

The effect of smoking on the gingival vasculature is not completely understood, however angiogenesis (formation of new blood vessels) seems to be suppressed (Bagaitkar et al. 2008). Contrary to popular belief, smoking likely does not induce gingival vasoconstriction. However, it has been repeatedly shown that gingival inflammation and bleeding on probing is decreased in smokers compared to nonsmokers with similar levels of periodontitis (Lie et al. 1998; Bagaitkar et al. 2008). Smokers who quit smoking experience an increase in bleeding even

with visibly reduced inflammation symptoms. It is suspected that smokers have a suppression of the normal inflammatory response to plaque, which is visible as a reduced number of gingival vessels (Palmer et al. 2005).

Smokers exhibit a reduction in gingival crevicular fluid (GCF) compared to nonsmokers (Kinane and Chestnutt 2000).The effect of tobacco smoke seems not to be the topical, direct effect of the smoke blowing over teeth. The lingual side of the teeth which are directly exposed to smoke do not show significant difference in the amount of GCF compared to the buccal side of the teeth (Kinane and Chestnutt 2000). The effect has to be either systemic or distributed via saliva (Kinane and Chestnutt 2000). One possible mechanism could be a reduction in blood flow (Kinane and Chestnutt 2000). The impaired angiogenesis in smokers may lead to a reduced inflammatory response against pathogenic bacteria (Bagaitkar et al. 2008).

Smoking impairs the healing abilities after surgical and non-surgical treatment in periodontology. Fibroblasts are important structural cells and are essential for wound healing. They produce type 1 collagen and fibronectin as well as collagenase. Nicotine and other tobacco smoke components have been shown to have a negative effect on gingival and periodontal ligament fibroblast recruitment and attachment (Palmer et al. 2005). These findings could explain the diminished healing response of the periodontal tissue of smokers (Kinane and Chestnutt 2000; Palmer et al. 2005).

2. Dysregulation of immune function

Smoking has a modulatory effect on neutrophils, monocytes, T-cells, and B-cells which can lead to an increase risk in infections (Palmer et al. 2005; Brook 2011). Changes in cytokine, chemokine, and growth factor concentrations in the gingival crevicular fluid have been observed (Kumar 2012). An increase in white blood cells, leukocytes in smokers (“Smokers' leukocytosis”) has been observed for more than 30 years (Barbour et al. 1997); however, the function of leukocytes is greatly reduced (Sopori 2002).

Smoking and professional phagocytes (neutrophils, monocytes, and macrophages). Both neutrophils (also called polymorphonuclear leukocytes) and macrophages are able to phagocytose and kill invading organisms. Monocytes are precursors of macrophages and are also able to phagocytose. These professional phagocytes kill their "prey" in the phagosome via fusion with the lysosome. The invader is killed via release of defensins and a highly reactive oxygen species such as superoxide (O_2–), hypochlorite (HOCl), hydrogen peroxide (H_2O_2), and hydroxyl radicals (-OH). The sudden increase in oxygen consumption necessary to produce these reactive oxygen species (ROS) is called respiratory burst. In addition, the rising pH activates potent proteases (Alberts et al. 2008, pg 1532). Furthermore, macrophages are also able to kill invaders extracellularly (Barbour et al. 1997).

In healthy tissue, neutrophils protect the gingiva against periodontal plaque bacteria (Palmer et al. 2005). While the number of neutrophils is increased systemically after exposure to tobacco smoke, it is not increased in the gingival sulcus and oral cavity. There is evidence that it might be even reduced in the gingival sulcus. This suggests that the migration across the periodontal microvasculature is impaired in smokers (Palmer et al. 2005). Tobacco smoke or nicotine has been shown to directly impair the function of neutrophils and monocytes, including phagocytosis of bacteria and intracellular killing via the generation of respiratory bursts (Sopori 2002; Bagaitkar et al. 2008). Neutrophils were found to have a reduced ability to adhere and phagocytose oral bacteria when exposed to tobacco smoke (Barbour et al. 1997). A potential mechanism is the perturbance of the cytoskeleton in oral neutrophils which might affect the motility and phagocytosis of oral neutrophils, impairing the immune response (Palmer et al. 2005).

In the lung, alveolar macrophages are responsible for exacerbating the harmful effects of smoking. The number of alveolar macrophages in the lung is increased in smokers and their elastase activity could partly be responsible for the increase in COPD in smokers (Sopori 2002).

In addition, the phagocytosis and bacterial killing ability are impaired along with lower levels of proinflammatory cytokines (Sopori 2002).

Tobacco smoke may be able to directly influence neutrophil and monocyte function through receptor binding (Bagaitkar et al. 2008). Substances found in tobacco smoke and their metabolites, such as nicotine, cotinine, and aryl hydrocarbons bind to receptors on neutrophils (Palmer et al. 2005). These receptors are upregulated in smokers and decrease after cessation (Palmer et al. 2005). Endogenous factors such as IL-8, ICAM-1 and TNF-alpha also bind to receptors on neutrophils and their natural agonists have been found to be misregulated in the oral cavity in smokers (Palmer et al. 2005). The changes in ICAM-1 may affect the recruitment of host defense cells to areas of inflammation and microbial colonization through an impairment of leukocyte binding to capillary endothelial cells (Kinane and Chestnutt 2000). Macrophages also express lower levels of class II MHC molecules in smokers, which may lead to a reduced humoral immune response (Kinane and Chestnutt 2000). In addition, monocytes respond to tobacco smoke with a down regulation of pathogen recognition receptors such as TLR-2 and MARCO (Bagaitkar et al. 2008).

Proteolytic enzymes produced by neutrophils such as matrix metalloproteinase and elastase are preferentially released from neutrophils after exposure to tobacco smoke and are suspected to degrade pulmonary vessels and tissues (Palmer et al. 2005).

This process might play a role in the destruction of periodontal tissue (Palmer et al. 2005). The release of reactive oxygen species might also play a role in the different immune responses in smokers compared to nonsmokers. Tobacco smoke is thought to increase fMLP receptor expression on neutrophils, which leads to a hyper-reactive response of release of reactive oxygen species and elastase (Palmer et al. 2005). This hyperinflammatory response leads to excessive tissue damage (Palmer et al. 2005). Others have shown that elastase levels in gingival crevicular fluid are lower in smokers (Kinane and Chestnutt 2000). The experimental

findings are not clear; no conclusion about the effect of reactive oxygen species can be drawn today (Palmer et al. 2005).

In conclusion, while neutrophils are an important part of immune response, hyperactivity can lead to destruction of gingival tissue (Palmer et al. 2005).

Smoking and other immune response. Dendritic cells process antigens and present them to adaptive immune cells. Tobacco smoke can hinder their maturation. This ultimately leads to a reduced antigen presentation through reduction in antigen uptake, antigen and costimulatory molecules (MHC) presentation as well as a decrease in T-cell stimulating cytokines in response to LPS (Bagaitkar et al. 2008).

Smokers have been found to have a lower number of natural killer cells with an impaired function. This finding could aid in the understanding of increased lung cancer incidence in smokers through reduced clearance of precancerous cells in addition to the immediate effect of carcinogens in tobacco smoke (Barbour et al. 1997; Sopori 2002).

As discussed in the Chapter 1.3.4, periodontitis is caused by an overreaction of the immune system to bacterial triggers (Kumar et al. 2011). The effect of smoking on the levels of pro- and anti-inflammatory cytokines is controversially discussed, where some findings demonstrate a decrease in proinflammatory cytokines including TNF-alpha and IL-6 and an increase in anti-inflammatory cytokines (IL-10) (Barbour et al. 1997; Bagaitkar et al. 2009), others find the opposite (Kinane and Chestnutt 2000; Kumar et al. 2011).

The effect of tobacco smoke on the expression of cytokines might not be solely dependent on the reaction of leukocytes, but also on the effects of smoking on the bacterial community. Evidence suggests that *Porphyromonas gingivalis* that have been exposed to cigarette smoke extract suppress proinflammatory cytokine response (Bagaitkar et al. 2009) and simultaneously promote the release of the anti-inflammatory cytokine IL-10 (Rehani et al. 2008). This induction was reversible by culture of *Porphyromonas gingivalis* in smoke free medium and

the level of cigarette smoke extract tested did not interfere with *Porphyromonas gingivalis* growth (Bagaitkar et al. 2009).

The research on lymphocyte number and function yields very controversial findings and it is not possible to draw any conclusions, presently (Palmer et al. 2005). Part of the reason might be a complex interplay between smoking, race, periodontal diagnosis, and age. In addition, tobacco smoke contains immunosuppressives, such as nicotine and benzo[a]pyrene, as well as immunostimulatory components such as metals (Palmer et al. 2005). Therefore the net effect on the immune system depends on the frequency, dosage, and the duration of exposure (Palmer et al. 2005).

Some evidence suggests that tobacco smoke may hinder T-cell proliferation after induction from antigens. It is also suspected that B cell proliferation and maturation is inhibited (Barbour et al. 1997; Bagaitkar et al. 2008). Studies on the effectiveness of the hepatitis B vaccine showed that smokers have a reduced immune response to the vaccination (Kinane and Chestnutt 2000). The production of different immunoglobulin by B-cells is influenced by tobacco smoke. Levels of IgE are increased (Bagaitkar et al. 2008), but IgG levels (anti-bacterial) are decreased (Barbour et al. 1997; Kinane and Chestnutt 2000; Bagaitkar et al. 2008). Specifically, IgG levels against oral pathogens such as *Prevotella intermedia*, *Fusobacterium nucleatum*, and *Aggregatibacter actinomycetemcomitans* are lower in smokers (Kinane and Chestnutt 2000). Taken together this could lead to a reduced ability to fight bacterial infections in smokers (Bagaitkar et al. 2008).

Summary. To summarize, there is evidence that smokers exhibit a reduction in inflammatory response. Further evidence comes from epidemiological studies, which found that smoking seems to have a protective effect on diseases many of which have an inflammatory component such as ulcerative colitis, sarcoidosis, pigeon breeders' disease/Bird fancier's lung, and Sjögren's syndrome, but not on others such as Crohn's disease (Sopori 2002). There is evidence that the immunosuppressive effect is caused by nicotine, at least partially (Sopori

2002). In addition to a direct effect on leukocytes, nicotine may modulate the immune system through the central nervous system via activation of the nicotinic anti-inflammatory pathway (Sopori 2002). This connects the nervous system directly with the inflammatory response (Scott and Martin 2006).

3. Increase in bacterial virulence

Besides changes in the host tissue, the oral bacteria are influenced by tobacco smoke. It has been shown that the expression profile of *Porphyromonas gingivalis* changes under culture conditions involving tobacco smoke (Bagaitkar et al. 2009). *Porphyromonas gingivalis* belongs to the "red complex" of organisms commonly associated with periodontitis (Chapter 1.3.4). A microarray experiment showed that 6.8% of *Porphyromonas gingivalis* genes were differentially expressed upon exposure to cigarette smoke extract. Cigarette smoke causes environmental stress to the organism and significantly induces operons of virulence genes, detoxification and oxidative stress-related genes as well as DNA repair genes. Enhanced virulence in pathogen-host interaction can be partly due to the significant upregulation of major fimbriae and putative lipoproteins required for fimbriae assembly, which are important for bacterial adhesion, invasion and upregulation of multiple proteases. Upregulated genes are found in the functional families of DNA replication and repair as well as insertion and transposition of genetic material. Tobacco smoke contains many known carcinogens and therefore an upregulation in DNA repair as observed is not surprising. The environmental stress of tobacco smoke might encourage genetic rearrangements. Lastly, resistance to xenochemicals in tobacco smoke with could be harmful to the cells could be obtained via the induction of efflux transporters (bacterial secretion system) and other transport proteins. Amongst the downregulated genes, the researcher found capsular genes as well as many unknown genes. Capsular genes can elicit pro-inflammatory response. This observation is in accordance with the clinical reduced signs of inflammation (Bagaitkar et al. 2009).

Adherence. Adherence is the first essential step in biofilm formation, whether the organism is beneficial, commensal, or a pathogen. Adherence works via specific receptor binding to the mucosa or other structures of the host (Brook 2011). Studies found a selective increase in adhesion properties of *Streptococcus pneumoniae* and other pathogenic species in smokers compared to nonsmokers (Brook 2011). Diluted concentrations of smoke extract were found to increase binding capacities, while the undiluted smoke extract reduced binding to epithelial cells (Brook 2011). This is in accordance to the upregulation of fimbriae and essential helper proteins in *Porphyromonas gingivalis* (Bagaitkar et al. 2009).

Bacterial toxins. There is experimental evidence that the combination of bacterial toxin and nicotine can have synergistic toxic effects on the host organism. The mixture of nicotine and bacterial toxins from *Staphylococcus aureus*, *Klebsiella pneumonia*, and *E. coli* strains which were obtained from sudden infant death syndrome cases showed a lethal synergistic effect on chick embryos at concentrations where no effect was observed when applied individually (Sayers et al. 1995). The same group showed a few years later that cell-free extracellular extracts of a number of periodontal pathogens, namely *Porphyromonas gingivalis*, *Porphyromonas asaccharolytica*, *Fusobacterium nucleatum*, and *Fusobacterium necrophorum* also resulted in synergistic toxicity with nicotine in the chick embryo model (Sayers et al. 1997). However, as others pointed out, mechanistic studies on the direct influence of tobacco smoke on the bacterial pathogenicity are rare in the literature (Bagaitkar et al. 2008; Brook 2011).

Bacterial growth. Bacterial growth can be stimulated or inhibited by exposure to tobacco smoke. The growth of *Streptococcus mutans* and *Streptococcus sanguis*, two common oral bacteria, was stimulated by cigarette smoke, however the study had experimental design flaws and did not address how much of the growth increase was due to a high CO2 content in the air or components in the tobacco smoke. However, the authors found substantial differences on bacterial proliferation between different brands of cigarettes (Zonuz et al. 2008). *Neisseria* growth is negatively influenced by tobacco smoke. The first evidence dates back to 1976, where

Neisseria were found to be statistically decreased on the tongue surface of smokers compared to nonsmokers (Colman et al. 1976). The finding of the sensitivity of *Neisseria* was confirmed by at least two other studies, but sensitivity of other organisms starts to emerge as well.

In vitro studies of the influence of tobacco smoke on the viability and growth rates of various oral organisms found that *Neisseria perflava*, *Neisseria sicca,* and *Moraxella catarrhalis* are very susceptible to cigarette smoke (Bardell 1981). The survival and growth of *Streptococcus* species including *Streptococcus mitis*, *Streptococcus salivarius*, and *Streptococcus sanguis* are only minimally affected by cigarette smoke (Bardell 1981). Ertel et al. showed one decade later again that the growth of *Neisseria*, *Moraxella catarrhalis*, and *Haemophilus influenza* were very susceptible to tobacco smoke, whereas *Pseudomonas aeruginosa*, *E. coli*, *Klebsiella*, *Enterobacter*, *Serratia*, *Providencia*, *Citrobacter*, *Enterococcus,* and *Acinetobacter* were not affected by tobacco smoke (Ertel et al. 1991). Among the susceptible organisms were several streptococci (*Streptococcus pyogenes* - Strep A, *Streptococcus agalactiae* - Strep B, *Streptococcus bovis*, *Streptococcus pneumonia*), and *Staphylococcus aureus* (Methicillin-sensitive - MSSA and Methicillin-resistant - MRSA) (Ertel et al. 1991). *Neisseria mucosa* levels were found to be significantly higher in periodontally healthy nonsmokers than in nonsmokers with periodontal disease or smokers in saliva, but not in other mucosal surfaces of the oral cavity (Mager et al. 2003a).

Because tobacco smoke changes bacterial gene expression and growth rates of specific bacterial entities, it is not surprising that whole shifts in the oral community can be observed. Besides direct host and bacterial changes, the gingival environment is altered. Tobacco smoke contains carbon monoxide (CO), which affects the oxygen saturation of haemoglobin. Research has shown that the oxygen tension in subgingival crevices is reduced, which could have a direct effect on the local bacterial community (Palmer et al. 2005; Kumar 2012). Furthermore, smoking increases the amount of free iron in the oral cavity and organisms that are commonly found in smokers thrive in iron-rich environments (Kumar 2012).

Biofilm formation. The subgingival plaque communities in smokers have been found to be distinct from nonsmokers (Kumar et al. 2011). As previously discussed (Chapter 1.3.2), an ordered biofilm formation begins within a few minutes after tooth eruption of the tooth and subsequent tooth cleanings. When controlling for other factors such as oral hygiene, smoking does not increase the amount of dental plaque (Lie et al. 1998; Palmer et al. 2005; Kumar et al. 2011). However, smoking does seem to alter the establishment of a healthy biofilm by enrichment of pathogens within the first 24 h (Kumar 2011), predisposing smokers to periodontitis. Niche saturation is disturbed, a phenomenon where a low number of species occupy an ecological niche and provide resistance to pathogen colonization (Kumar et al. 2011).

The initial plaque formation of sub- and supragingival plaque for the first 7 days of periodontally healthy smokers and nonsmokers have been studied (Kumar et al. 2011). Based on the alpha diversity index, the bacterial diversity was higher in smokers up to day 4, but approached nonsmoker levels by day 7 in the marginal plaque (Kumar et al. 2011). As the biofilm builds up over time, an overall trend of decreasing alpha diversity values was shown for smokers and nonsmokers, more so in smokers as they started out at a higher diversity level (Kumar et al. 2011). The microbial communities were less stable in smokers compared to nonsmokers. Bacterial communities exhibiting greater fluctuations as observed in smokers, could predispose the biofilm for easier colonization with pathogens (Kumar et al. 2011). Pathogenic genera were observed within 24 h of biofilm formation such as *Fusobacterium*, *Cardiobacterium*, *Synergistes*, and *Selenomonas*, but also genera which include systemic pathogens for example *Haemophilus* and *Pseudomonas* (Kumar et al. 2011).

Targeted microbiology studies. Many older studies on the oral microbiota of plaque on different oral sites of smokers and nonsmokers with and without periodontitis yielded conflicting results. Some studies found an increase in certain organisms (especially known periodontitis associated pathogens such as *Prevotella intermedia*, *Actinobacillus actinomycetemcomitans*, *Tannerella forsythia*, *Eubacteria nodatum*, *Micromonas micros*, *Prevotella*

nigrescens, *Porphyromonas gingivalis*, *Treponema denticola*, or *Campylobacter rectus*); while others did not find any significant differences (Palmer et al. 2005). Possible reasons for this controversy could include: 1. All studies relied on a culturable or known organism, 2. the sampling site and depths differed, and 3. different analysis methods differed (immunofluorescence, culture, DNA based methods) (Palmer et al. 2005). Prevalence of colonization is highly tooth location specific with a bilateral symmetry in smokers as well as nonsmoker as has been shown for *Tannerella forsythia* and *Prevotella nigrescens* (Haffajee and Socransky 2001). It is also important to distinguish between abundance and presence of certain organisms. While a study by Haffajee and Socransky did not yield differences in abundance based in DNA-DNA checkerboard assay, a significant increase in presence of members of the orange and red complex (*Eubacterium nodatum*, *Fusobacterium nucleatum ss vincentii*, *Prevotella intermedia*, *Peptostreptococcus micros*, *Prevotella nigrescens*, *Tannerella forsythia*, *Porphyromonas gingivalis*, and *Treponema denticola*) in smokers was observed (Haffajee and Socransky 2001). A selection of studies which found significant differences between smokers and nonsmokers is presented in Table 3.1. Another caveat in many of these studies might be statistical power. Palmer et al. stresses the importance of an adequately powered studies to overcome the intrinsic high variation, which is commonly observed and which may mask real effects (Palmer et al. 2005). This problem is inherent to all human microbiome studies, if effects are small.

Table 3.1: A selection of oral microbiology studies based on smoking status

Study	Study design	Location	Analysis	Increased in smokers	Decreased in smokers
Kamma and Nakou (1997)	Subjects: 25 smokers, 25 nonsmokers with early onset periodontitis	Subgingival plaque	Culture	*Staphylococcus aureus, Campylobacter concisus, Eikenella corrodens, Escherichia coli, Tannerella forsythia, Bacteroides gracilis, Campylobacter rectus, Porphyromonas gingivalis, Selenomonas sputigena* and *Candida albicans*	*Peptostreptococcus micros, Actinomyces naeslundii, Eubacterium lentum* and *Capnocytophaga gingivalis*
Haffajee and Socransky (2001)	Subjects: 124 never smoker, 98 past smoker, 50 current smokers	Subgingival plaque	29 bacterial species using checkerboard DNA-DNA hybridization	*Eubacterium nodatum*, *Fusobacterium nucleatum* ss *vincentii*, *Prevotella intermedia*, *Peptostreptococcus micros*, *Prevotella nigrescens*, *Tannerella forsythi* , *Porphyromonas gingivalis,Treponema denticola* and *Selenomonas noxia*	*Actinomyces naeslundii* genospecies 2 and *Eikenella corrodens*
Mager et al. (2003a)	Subjects: 72 active smokers, 72 involuntary smokers, 56 nonsmokers	Saliva and mucosal surfaces	40 bacterial species using checkerboard DNA–DNA hybridization	No significant differences found. Trend: *Prevotella nigrescens*, *Fusobacterium* and *Actinomyces s*pecies	*Neisseria mucosus* higher in periodontically healthy nonsmokers than smokers or nonsmokers with periodontitis in saliva. *Eikenella corrodens, Streptococcusconstellatus* higher on soft tissues of periodontically healthy nonsmokers.
Nishida et al. (2008)	Subjects: 72 active smokers, 72 involuntary smokers, 56 nonsmokers	Saliva	qPCR of 6 organisms		*Prevotella nigrescens*
Shchipkova et al. (2010)	Subjects: 15 smokers, 15 nonsmokers	Subgingival plaque	16S Sanger sequencing	*Parvimonas, Fusobacterium, Campylobacter, Bacteroides* and *Treponema* *Species: Uncultured Peptostreptococci, Parvimonas micra, Campylobacter gracilis, Treponema socranskii, Dialister pneumosintes* and *Tannerella forsythia*	*Veillonella, Neisseria,* and *Streptococcus* *Species: Veillonella* sp. oral clone B2*, Neisseria* sp. oral clone 2.24, *Streptococcus sanguinis,* and *Capnocytophaga* sp. clone AH015
Kumar et al. (2011)	Subjects: 15 smokers, 15 nonsmokers, all periodontically healthy.	Subgingival and supragingival plaque development during first 7 days	16S Sanger sequencing	Subgingival plaque: *Lactobacillus, Fusobacterium, Centipeda, Pseudomonas, Leptotrichia, Synergistes, Propionibacterium and Cardiobacterium.* Supragingival plaque: *Streptococcus, Haemophilus, Kingella, Selenomonas, Lachnospira, Pseudomonas, Lactobacillus* and *Treponema*	Subgingival plaque: *Neisseria, Actionomyces, Rothia, Lautropia*
Bizzarro et al. (2013)	Subjects: 15 smokers and 15 nonsmokers with periodontitis	Subgingival plaque	Culture, qPCR, 16S with 454	Abundance: Class: *Fusobacterium* OTUs belonging to *Fusobacterium, Prevotella* and *Selenomonas*	Abundance: *Peptococcus* and *Capnocytophaga*

16S RNA sequencing studies. Newer culture independent experimental approaches, such as sequencing of bacterial 16S rRNA genes allows for an open approach and generally yield more differences than experiments that rely on a known subset of organisms such as the ones mentioned above (Palmer et al. 2005). Bizzarro et al. compared the subgingival bacterial communities of smokers and nonsmokers with three different analysis techniques, namely culture, qPCR and 16S pyrosequencing (Bizzarro et al. 2013). While the more traditional techniques (culture and qPCR) did not show a difference based on smoking status, 16S rRNA gene sequencing revealed significant differences in abundance of various genera (Bizzarro et al. 2013, Table 3.1). This example stresses the importance of using an open-ended approach compared with a targeted analysis approach for exploratory studies. While culture based studies are essential for further investigations, open ended nucleic acid based techniques such as 16S rRNA gene sequencing are better suited for exploratory studies. Culture results, particularly, differed substantially from the DNA based techniques (Bizzarro et al. 2013).

As Table 3.1 shows, many studies show an increase in periodontal pathogens such as *Bacteroides*, *Treponema*, *Campylobacter*, and *Fusobacterium* (Socransky et al. 1998). While commensal such as *Veillonella neisseria* and *Streptococcus sanguinis*, which are often associated with periodontal health, are sometimes less abundant in smokers (Paster et al. 2001). Interestingly, *Neisseria*, which exhibited a strong sensitivity against tobacco smoke in vitro, is often reduced in vivo in smokers.

The periodontal health of the individual might be more important than smoking status. A study of subgingival bacterial communities of smokers and nonsmokers yielded two distinct clusters based on PCoA analysis. One exhibited reduced diversity (less taxa), more severe attachment loss, and an increase in *Porphyromonas*, *Selenomonas*, and *Capnocytophaga*, while the other exhibited an enrichment in *Fusobacterium*, *Paludibacter*, and *Desulfobulbus*. No significant difference was found between the average OTUs observed in smokers compared to nonsmokers (Bizzarro et al. 2013).

Very few studies have examined the saliva of smokers. Mager et al. studied saliva as well different tissue locations in the mouth (including buccal, tongue, palate, lip, gingival) of smokers and nonsmokers and found a reduction of *Neisseria* in the saliva of smokers (Mager et al. 2003a). Nishida et al. studied saliva and the effect of active, involuntary, and nonsmoking on the abundance and prevalence of six oral pathogens. They found a significant reduction of *Prevotella nigrescens* in active and involuntary smokers (Nishida et al. 2008).

Tooth decay. While there is clear evidence that smoking aggravates gum disease, the caries data are not clear. There is weak evidence that suggest an increased tooth decay risk with smoking (Benedetti et al. 2013).

Smoking cessation. Smokers exhibit enrichment in pathogens compared to nonsmokers (Shchipkova et al. 2010). Cessation of smoking alters the subgingival microbial communities (Fullmer et al. 2009) towards a level that is more similar to nonsmokers with an enrichment in health associated bacteria and a decrease in putative oral pathogens (Fullmer et al. 2009; Delima et al. 2010). The change observed is from a pathogen enriched environment with organisms such as *Parvimonas micra*, *Treponema denticola* and *Filifactor alocis* to a healthier community with higher levels of *Veillonella parvula* (Delima et al. 2010). After smoking cessation, periodontitis risk levels start to decrease, but do not approach periodontitis risk levels of never-smokers (Kinane and Chestnutt 2000).

Environmental tobacco smoke. Environmental tobacco smoke, better known as second hand smoke, is less studied, but also has been found to have detrimental effects. Cigarettes have two types of smoke, the mainstream smoke which is produced by puff-drawing the smoke into the mouth, and the sidestream smoke, which is directly emitted into the air (Sopori 2002). A cigarette burns at higher temperatures during inhalation and therefore the concentration of many toxic components is higher in sidestream smoke (Sopori 2002). Environmental tobacco smoke consist of sidestream smoke and exhaled mainstream smoke (Sopori 2002). An estimated ~50,000 death in the US are attributed to passive smoking (Arbes et al. 2001). Smoking parents

have been found to harbor pathogens in their oral cavity which can cause ear infections in their children at a higher frequency than nonsmoking parents (Brook 2011). Involuntary smoking also yields an increase in periodontitis risk compared to nonsmokers (Nishida et al. 2008). The increased risk was calculated to be 1.6 times higher for nonsmokers exposed to environmental tobacco smoke compared to nonsmokers who are not exposed to smoke (Arbes et al. 2001).

There are a number of potential mechanisms in which smoking alters the outcome of oral health, especially an increase in periodontitis severity. This chapter discussed main mechanisms for an increased risk of infection in smokers. The inflammatory cell response and wound healing are depressed through various mechanisms. Bacterial virulence is increased due to increased adherence, toxicity in combination with tobacco components, direct influence on bacterial growth rates, and ultimately community shifts which may predispose smokers to negative periodontitis outcomes. Figure 3.3 summarizes the negative network of potential influencing factors leading eventually to severe periodontal destruction.

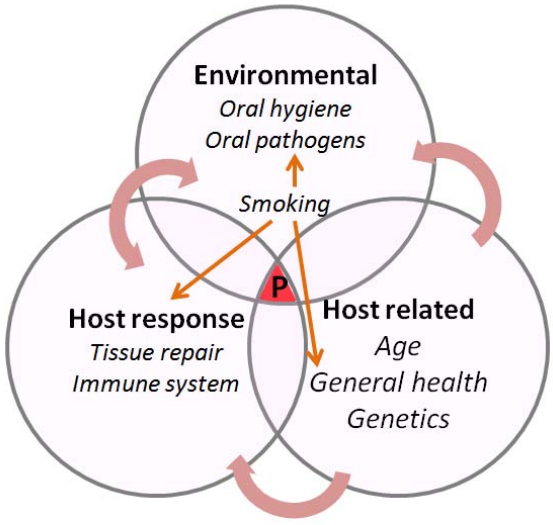


Figure 3.3: Factors that increase risk of periodontitis with emphasize on smoking. Modified from Barbour et al. and Kinane and Chestnutt (Barbour et al. 1997; Kinane and Chestnutt 2000).

3.1.3 The effect of alcohol consumption

Alcohol consumption is an acceptable social custom in today's society. More than 50% of persons 12 years and older in the US reported alcohol consumption during the past 30 days and 23.1 % were binge drinking (min. 5 drinks on one occasion), while 6.7% reported heavy drinking (binge drinking on at least 5 days within the last month) in 2010 (Substance Abuse and Mental Health Services Administration 2011). Alcohol use is not only common, but also detrimental to health. As the World Health Organization Global Burden of Disease Study 2010 points out, among the global leading risk factors of years lived with disability and years of life lost were tobacco smoking (No.2) and alcohol use (No.5) right after heart disease (No.1). Alcohol use accounted for 2.7 million deaths worldwide in 2010 and it was the leading risk factor for the age group 15-49 years (Lim et al. 2012). Furthermore, alcohol and tobacco smoking are also known risk factors for the development of head-and neck-cancer (Seitz and Stickel 2007). Drinking guidelines recommend not more than 28g of alcohol which equals 2 drinks with 14g alcohol each for men and a maximum of 14g of alcohol for women. The weekly dose should not exceed 14 units per week (196g alcohol) for men and 7 units per week (98g alcohol) for women (International Center for Alcohol Policies 2012, accessed July 19, 2013).

Alcohol, saliva and the salivary glands. While alcohol has various effects on the whole human body especially the liver and esophagus, alcohol also affects the oral cavity. Alcohol consumption has an effect on the salivary glands and saliva production. The largest salivary glands, the parotid glands located below and in front of both ears, are often enlarged in chronic alcoholics (see Figure 1.7 for anatomy). In the rat model, an accumulation of fat and a reduction in size and protein content of the parotid glands were found along with a reduction in salivary flow. It is likely that the reduced salivary flow is an effect of salivary gland atrophy. A reduced saliva flow allows for accumulation of carcinogens and bacteria in the human mouth, which could contribute to the increase in cancer. Saliva has buffering capacity, not only for pH, but potentially also for carcinogens (Riedel et al. 2003; Goodchild and Donaldson 2007).

Cancer. 3.5% of all cancer cases worldwide are attributed to chronic alcohol consumption, including cancers in the upper aerodigestive tract, the colorectum, liver, and breast (Seitz and Stickel 2007). The biggest risk factors for head and neck cancer are smoking and heavy alcohol consumption. While the involvement of smoking in carcinogenesis seems intuitively correct since tobacco smoke contains thousands of health damaging and carcinogenic components, ethanol itself is non-carcinogenic (Riedel et al. 2003). The combination of heavy alcohol and heavy smoking yields a synergistic, elevated risk (Salaspuro and Salaspuro 2004). Multiple studies suggest that alcohol alone increases the likelihood of carcinomas in the oral cavity and esophagus in a dose-response manner (Riedel et al. 2003). Alcohol abstinence as it is practiced by religious groups such as Seventh Day Adventists or Mormons have lower head and neck cancer rates than expected in these populations (Kato and Nomura 1994; Riedel et al. 2003).

Metabolism of ethanol. Three enzymes are the main metabolic components in ethanol elimination in the human body: Alcohol dehydrogenase (ADH), cytochrome P450 2E1 (CYP2E1), and catalase, with ADH being the major route of metabolizing (Riedel et al. 2003; Seitz and Stickel 2007), see Figure 3.4.

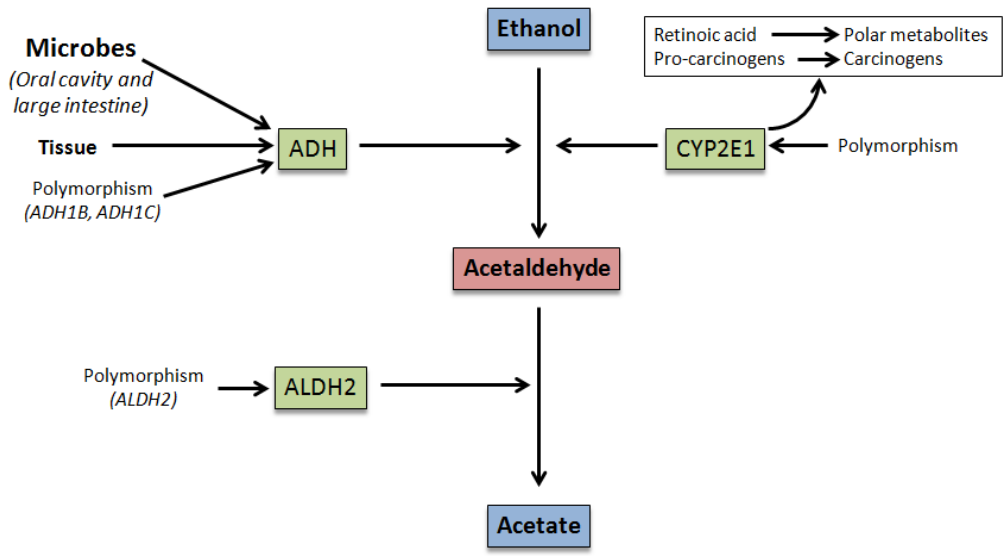


Figure 3.4: Metabolism of ethanol and its implication in cancer. For details, please see text. Abbreviations: ADH=Alcohol dehydrogenase, CYP2E1=cytochrome P450 2E1, ALDH=acetaldehyde dehydrogenase, modified from Seitz and Stickel (2007).

ADH, which is produced by the host tissue as well as oral microbes, oxidizes ethanol to acetaldehyde. There is genetic variation between the effectiveness of ADH alleles in humans (Seitz and Stickel 2007). Acetaldehyde is then oxidized to acetate via aldehyde dehydrogenase 2 (ALDH2). The metabolite enters the citric acid cycle as acetyl-CoA for energy production. CYP2E1 also converts ethanol to acetaldehyde, but it also produces reactive oxygen species. CYP2E1 is induced by chronic alcohol ingestion and can metabolize pro-carcinogens as well. For example, CYP2E1 can metabolize tobacco smoke to carcinogens and reduce retinoic acid (Vitamin A) levels in some tissues, which results in a lack of retinoic acid essential cell function. Acetaldehyde and reactive oxygen species bind to DNA forming stable DNA adducts that can be mutagenic in the context of stimulated DNA repair (Seitz and Stickel 2007).

As discussed above, ADH oxidizes ethanol to acetaldehyde, which is a known carcinogen in animal models and cell cultures and mimics the effects of alcohol in histological studies (Riedel et al. 2003). The levels of acetaldehyde in saliva have been found to be high enough to act as a carcinogen (50-100uM) after 0.5 g ethanol/kg body weight. Acetaldehyde concentrations in saliva are about 10-100 times higher than in blood (Seitz and Stickel 2007).

ADH is produced by various tissues of the human host, including the oral mucosa, but also by oral microbes (Homann et al. 2000). Yeast and streptococci have been shown to be able to convert ethanol into acetaldehyde (Kurkivuori et al. 2007). Another piece of epidemiological evidence supporting the role acetaldehyde in cancer risk, is that individuals with a highly active ADH variant and low ALDH2 activity, show elevated levels of acetaldehyde presumably due to hampered detoxification of acetaldehyde (Yokoyama et al. 2008). This results in an elevated risk of cancer in the upper gastrointestinal tract (Riedel et al. 2003; Seitz and Stickel 2007). Acetaldehyde is also a component in tobacco smoke (Salaspuro and Salaspuro 2004). ADH and ALDH2 are expressed at different levels in various tissues including the oral cavity (Riedel et al. 2003).

In the oral cavity the activity of ADH was found to be much higher than ALDH2, which allows for a prolonged exposure to acetaldehyde due to insufficient oxidation of acetaldehyde by ALDH2 (Dong et al. 1996). However, in this particular study by Dong et al., no precautions were taken to remove the bacterial biofilm on the specimen and therefore the activity of bacterial and human ADH was measured. Combined smoking and alcohol use potentiates the production of acetaldehyde compared to alcohol alone. In a study by Salaspuro and Salaspuro (2004), smokers and nonsmokers were asked to ingest 0.8g Ethanol/kg body weight, which corresponds to approximately 4-5 drinks with 14g pure alcohol for a 80kg person with and without consuming a cigarette at the same time. Saliva samples were taken every 20 minutes for 160 minutes thereafter. Smokers had two times higher acetaldehyde concentration in saliva than nonsmokers after alcohol alone. In combination with active smoking, smokers had a seven times higher acetaldehyde concentration compared to nonsmokers. The increase of acetaldehyde due to smoking is very short lived and drops rapidly after finishing a cigarette to almost, but not quite nonsmoking levels (Salaspuro and Salaspuro 2004). A regression analysis showed that heavy alcohol consumption as well as smoking are independent risk factors for higher acetaldehyde production (Homann et al. 2000).

The mechanism for the increase in ADH could be either through induction of the host metabolism through the presence of a larger bacterial load or specific organisms or through direct ADH production by the oral organisms, which could be inducible. An involvement of bacteria is likely as a study showed that individuals with a mutant ALDH2 allele showed higher cancer rates in tissues with a bacterial biofilm (oropharynx, stomach, colon, lung) but not in sterile tissues (liver and others) (Riedel et al. 2003). Therefore, one can speculate that the additional production of acetaldehyde by microbes increases its levels to a point where a reduced level of activity of the ALDH2 cannot metabolize it in a timely manner. This, then, can lead to an accumulation of acetaldehyde and a prolonged exposure to this carcinogen.

Oral microbes. There is a substantial variation of acetaldehyde production after ethanol consumption between individuals (Homann et al. 1997). This could be partly the result of variation in microbial oral biofilms among people. Treatment with antiseptic mouthwash with chlorhexidine led to a reduction of acetaldehyde production (Homann et al. 1997), which suggests that the oral microbiota has a substantial influence on the localized acetaldehyde production. Another study by the same group found smoking and heavy drinking as the strongest predictors for increasing microbial acetaldehyde production in saliva. A culture based microbial analysis of saliva, split into "low" and "high" acetaldehyde producing individuals, showed that high producers had an increase in total counts and aerobic Gram-positive bacteria and yeasts; specifically *Streptococcus salivarius*, hemolytic *Streptococcus viridians* var., *Corynebacterium* sp., and *Rothia* were increased (Homann et al. 2000). The abundance as well as the presence of yeast was increased in "high" producers. There was no bacterial species more frequent in the "low" acetaldehyde group. The differences were not based on smoking or alcohol consumption behavior.

Neisseria, a bacterial genus, that is affected by cigarette smoke (see Chapter 3.1.2 on the effects of smoking), was found to be a high producer of ADH activity compared to some other culturable members of the oral bacterial community (*Streptococcus*, *Moraxella*, *Rothia*) or *E.coli*. *Neisseria*'s growth rate was also found to be largely unaffected by increasing ethanol concentrations, in contrast to *Streptococcus*. Rinsing the mouth with a 10% ethanol solution led to a rapid increase in *Neisseria* numbers after 6 hours in selected individuals (Muto et al. 2000). However, no study has shown specific bacterial species associated with high alcohol exposure (Riedel et al. 2003).

Two possible causes of an increase of the total oral bacterial biomass are a reduction of saliva flow and poor oral hygiene. As described above, excessive alcohol intake as well as smoking reduces saliva flow (Riedel et al. 2003). Saliva is one vehicle to reduce the number of bacteria on the oral surfaces by constant flushing. If the saliva flow is reduced, the number of

bacteria can increase, which has been observed in smokers and drinkers (Homann et al. 2000). In addition, self reported dry mouth was a risk factor for an increase in acetaldehyde production (Homann et al. 2000). Dry mouth is also commonly observed in methamphetamine users (see Chapter 3.1.4 on the effects of methamphetamines).

Furthermore, heavy alcohol consumption is correlated with poor oral hygiene, which might lead to bacterial overgrowth and an elevation in acetaldehyde levels in saliva due to an increase in bacteria and yeasts (Riedel et al. 2003; Seitz and Stickel 2007). Alcohol consumption of more than 75g per day (the equivalent of more than 5 drinks) and smoking more than 20 cigarettes a day was negatively correlated with frequency of tooth brushing (Maier et al. 1993). Lack of or neglect of oral hygiene, that is suggested to be an expression of self-neglect, is more common amongst individuals with head and neck cancer. In addition, a larger plaque buildup, more decayed teeth, less frequent dental check-ups, and a higher incidence of chronic inflammation were observed in these individuals (Maier et al. 1993).

Plant based beverages. It is important to note that the type of alcoholic beverage is likely to modulate the oral microbial communities to different extents; some even have anti-caries or anti-gingivitis properties. The content of secondary plant metabolites such as polyphenols and other antioxidants, fibers, trace elements, and vitamins varies between beverages and these might override the detrimental effects during ethanol metabolism, leaving a net positive impact on overall health (Romeo et al. 2007). Due to relatively incomplete data in these studies, I present below, some of the tantalizing effects suggested by the studies for completeness but acknowledge that some of these conclusions need additional corroborating data in the future.

Evidence by DGGE suggests that red wine and coffee consumption can change the composition of microbes in the oral cavity (Signoretto et al. 2010). Other plant based beverages such as coffee, cocoa and tea, as well as various other plants or plant extracts have shown an effect on oral bacteria. This has been reviewed recently (Ferrazzano et al. 2009; Gazzani et al.

2012). Briefly, this modulating effect is likely due to the polyphenols and small molecules naturally occurring in plants. Polyphenols are present in a variety of drinks and foods (coffee, cocoa, apple juice, cranberry juice), including alcoholic beverages such as beer and wine. Several studies have found an antigingivitis or anticaries effect of plant or fungal based products. Some studies suggest that there is a difference in the risk level of head and neck cancer based on the type of alcohol consumed (beer vs. wine vs. whisky), but the results are controversial (Riedel et al. 2003).

Anticaries effects of plant and fungal extracts could be due to a variety of mechanisms including strengthening the tooth enamel, interfering with bacterial metabolism or adhesion, or elimination of certain pathogenic organisms. A study by Spratt et al. (2012) evaluated the effect on biofilm formation initiation and persistence, coaggregation, and antibacterial activities of a number of food products on a selected number of oral organisms. The foods tested included green and black tea, cranberry juice, raspberries, shiitake mushrooms, red chicory and, relevant for this dissertation, beer. Beer was found to inhibit all tested organisms including *Streptococcus mutans* and to a lesser extent, *Lactobacillus casei*. There was a dramatic inhibitory effect on the biofilm formation of *Veillonella dispar*, *Fusobacterium nucleatum*, *Actinomyces naeslundii*, *Prevotella intermedia*, *Neisseria subflava*, and *Streptococcus sanguinis*. The antibacterial activity and prevention of coaggregation was not pronounced. It was effective in disrupting existing biofilms of *Actinomyces naeslundii*, *Streptococcus mutans*, *Lactobacillus casei*, and *Prevotella intermedia* (sorted by increasing effectiveness). Adhesion of *Actinomyces naeslundii* to gingival epithelial cells was partially prevented and invasion was effectively prevented. There was a small effect of the prevention of adhesion to hydroxyapatite as well as a disruption of signal transduction of *Streptococcus mutans* (Spratt et al. 2012).

Dealcoholized beer and wine were studied in regard to oral disease. Beer has compounds that are known to aid oral health. Beer had the highest content in selenium of a number of food products tested and very high polyphenol content. In addition, the tested beer

was found to contain fluoride, zinc, strontium, and boron (Daglia et al. 2011). Fluoride is known to strengthen teeth. Zinc ions have antimicrobial properties. Strontium is chemically similar to calcium and can replace the calcium ions in hydroxyapatite in teeth, which can prevent caries. Boron is also thought to reduce the incidence of decayed teeth. In another study by Daglia et al., red and white wines were found to have antibacterial and bactericidal properties against oral streptococci (including *Streptococcus anginosus, S. constellatus, S. intermedius, S. mutans, S. oralis, S. salivarius, S. sanguinis,* and *S. vestibularis*). The study by Daglia suggests that the antibacterial properties are due to low-molecular organic acids, succinic, malic, lactic, tartaric, citric, and acetic acid which occur naturally in grapes or are synthesized during malolactic fermentation, not polyphenols. Red wine was more effective than white wine (Daglia et al. 2007).

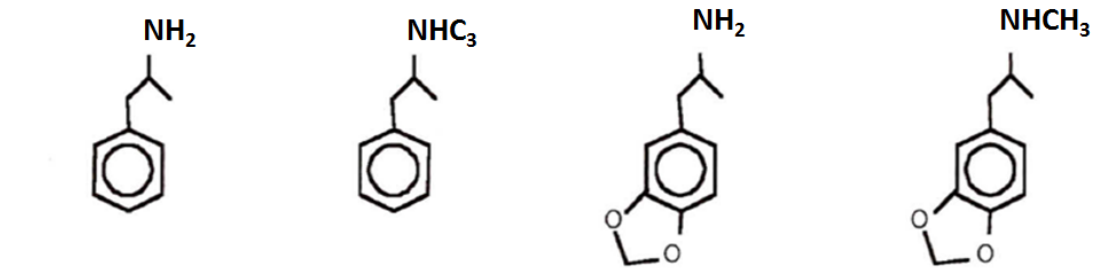
The effect of alcohol on the immune system. The effect of alcohol on the immune system is not well understood and it seems to be influenced largely by the amount, frequency and type of alcoholic beverage consumed as well as gender, genotype and nutrition status (Romeo et al. 2007). Moderate alcohol consumption (up to 3-4 drinks a day) have been linked to no increase or even a decreased risk of upper respiratory infections and up to 2 drinks per day have been proposed to have a beneficial impact on the immune system compared to total abstinence and alcohol abuse via an anti-inflammatory pathway, including an increase in production of cytokines (IL-2, IL-4, IL-10, interferon gamma) (Romeo et al. 2007).

In contrast to the consumption of moderate amounts of alcohol, alcohol abuse can lead to immunodeficiency and autoimmunity (Cook 1998). Alcoholics are more likely to acquire infectious diseases and have a suboptimal immune response; however, it is unclear if these observations are due to the direct effect of alcohol or comorbidities such as malnutrition or liver disease (Romeo et al. 2007). High doses of alcohol are known to lead to an increase in immunoglobulin A (IgA), IgG, and IgM levels, which likely result from a misregulation of antibody production as alcoholics are immunodeficient despite high antibody counts. Bacterial

LPS in combination with alcohol might increase the response of monocytes and macrophages and lead to an increase in damage to liver and other cells (Cook 1998).

3.1.4 The effect of stimulants using the example of methamphetamine

Discovery. Alphamethyl-phenylethylamin, later abbreviated as "amphetamine" was first isolated from the plant *Ephedra vulgaris* in 1887 and was chemically synthesized shortly thereafter in 1893 by the Japanese chemist Nagayoshi Nagai (Freye and Levy 2009, pg. 109). Methamphetamine, where an H molecule is replaced by a methyl group (Figure 3.5), was first synthesized in 1912 by Akira Ogata from ephedrine (Freye and Levy 2009, pg. 110). Its appearance are white to light brown crystals or powder (US Department of Transportation).



Amphetamine (Benzedrine, speed, Bennies, uppers)	**Methamphetamine** (speed, crank, meth, crystal meth, ice, uppers)	**3,4-Methylenedioxy-amphetamine** (MDA, love drug, love pill, ice)	**3,4-Methylenedioxy-methamphetamine** (MDMA, ecstasy, E, XTC, Adam, clarity, lover's speed, essence)
Medical uses: Narcolepsy, hyper-activity in children	**Medical uses:** Weight loss	**Medical uses:** None	**Medical uses:** None

Figure 3.5: Structure, street names, and medical uses of amphetamine and its analogs, adapted from Goodchild and Donaldson (2007) and Liska (2004, pg. 185ff.).

Chemical structures and pharmaceutical usage. Amphetamine and methamphetamine belong to the class of stimulant drugs. They are both in the Schedule II class, which means that they have a high potential for abuse, but there is also a currently accepted medical use and they are prescribed as treatment for attention deficit hyperactivity disorder (ADHD), narcolepsy, and

obesity (Goodchild and Donaldson 2007; Hamamoto and Rhodus 2009). They are sold under the brand name "Dexedrine" and "Adderall" for ADHD and narcolepsy (amphetamine) and "Desoxyn" for weight reduction (methamphetamine) (US Department of Transportation).

Amphetamine has structurally and functionally close relatives including methamphetamine, methylenedioxyamphetamine (MDA), methylenedioxymethamphetamine (ecstasy), and Ritalin (Hamamoto and Rhodus 2009). All of these compounds have a stimulatory effect on the central nervous system, but the potencies differ (Cho and Melega 2002).). Legally and illegally produced amphetamines are often abused for recreational use and abuse. Among all these different types of amphetamines, methamphetamine is the most often abused type (Hamamoto and Rhodus 2009).

Illegal drug usage. Amphetamine was originally used as a decongestant in a nasal inhaler, but the performance enhancing properties were quickly discovered (Freye and Levy 2009, pg. 109). The first non-medical, widespread use of methamphetamine was during World War II, when German military used it amongst all ranks and divisions to increase alertness and decrease fatigue. Methamphetamine was also used by US, British, and Japanese forces. It was still used by United States Air Force in the early years in 2000 for missions in Afghanistan and possibly still today (Freye and Levy 2009, pg. 112).

After a period of popularity in the 1960s as an illegal drug, it was not until the last decade when it was rediscovered by adolescents and young adults. "Home" laboratories started to emerge throughout the Pacific coast, Southwest, and West Central regions of the US (Goodchild and Donaldson 2007). The manufacturing is relatively easy and cheap, which helped spread its popularity (Hamamoto and Rhodus 2009). Even though current use of methamphetamines is on a decline, 105.000 first time users were reported in 2010 (Substance Abuse and Mental Health Services Administration 2011). Approximately 35 million people worldwide were using methamphetamine in 2000 (Hamamoto and Rhodus 2009) and 353,000 people 12 years and older in the US reported to have consumed methamphetamine within the

last month in 2010 (Substance Abuse and Mental Health Services Administration 2011). Methamphetamines are administered mainly by smoking or orally, but can also be snorted, or injected (Goodchild and Donaldson 2007). Population studies on either methamphetamine or amphetamines alone are difficult as street names are partly overlapping: for example methamphetamine as well as amphetamine are known as "speed" (Goodchild and Donaldson 2007), which makes the identification of the exact compound taken difficult, without laboratory tests on the original substance. However, the primary metabolite of methamphetamine is amphetamine in vivo (Cho and Melega 2002) and therefore, a strict distinction might not be necessary.

Effect on the human body. The primary and sought after effect of methamphetamine consumption is a release of dopamine and norepinephrine at the synapse and an inhibition of reuptake. This causes a feeling of euphoria, hyperactivity, hyperalertness, suppressed appetite, and increased physical and sexual endurance. Side effects include aggression, dizziness, hypertension, tachycardia, and tremor; furthermore toxic doses can lead to severe hallucinations, palpitations, convulsions, and coma (Goodchild and Donaldson 2007; Hamamoto and Rhodus 2009). Chronic methamphetamine use can lead to toxic effects by excessive dopamine release and result in long-term behavioral changes due to the suppression of dopamine transporter activity as well as other dopaminergic changes, such as an increase in extracellular Ca^{2+} concentration, glutamate, and endogenous opioids, as well as glia cell activation (Cho and Melega 2002). A study by Wang, Volkow et al. found that there was a long lasting decrease in striatal metabolism and nucleus accumbens which could be the underlying cause for a persistent amotivation and anhedonia (lack of pleasure) (Wang et al. 2004). PET scans of recent chronic methamphetamine users show a reduced activity within the dopaminergic basal ganglia (Figure 3.6) and even though there is a substantial improvement in brain activity, the brain does not fully recover even after one year of abstinence (Wang et al. 2004; Freye and Levy 2009).

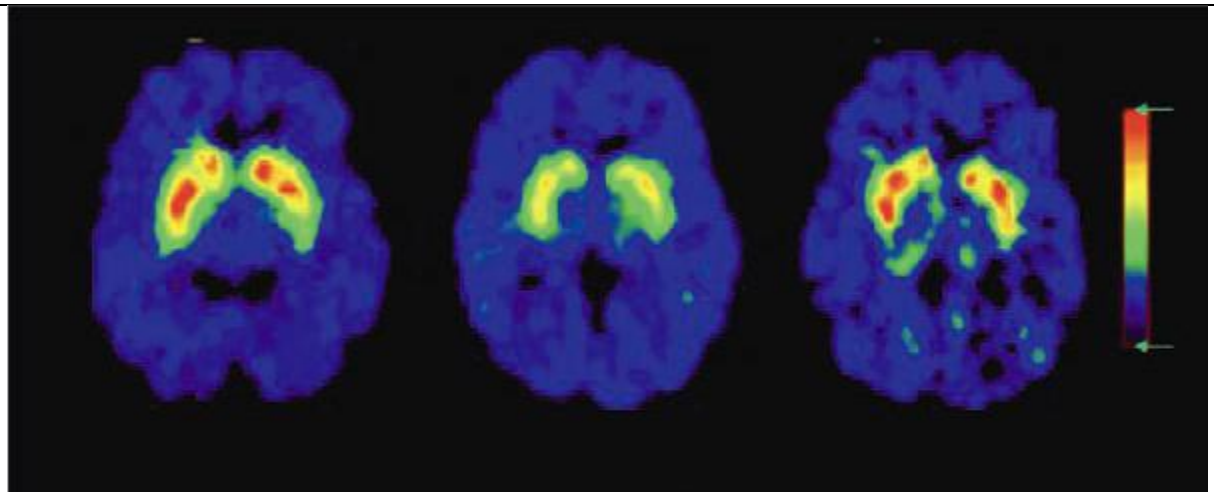


Figure 3.6: Reduced brain activity after short-term and long-term methamphetamine abstinence. PET scan from 1) Control brain, 2) methamphetamine user after 1 month of abstinence and 3) 24 months of abstinence from Freye and Levy (2009, pg. 116).

Effect on oral health. Methamphetamine consumption has a profound effect on oral health. The term "meth mouth" was invented to describe the rampant tooth decay progression, specifically enamel erosion, and excessive tooth wear, which is often seen in methamphetamine consumers. The teeth become "blackened, stained, rotting, crumbling, or falling apart", according to actual methamphetamine consumers (Hamamoto and Rhodus 2009), Figure 3.7.

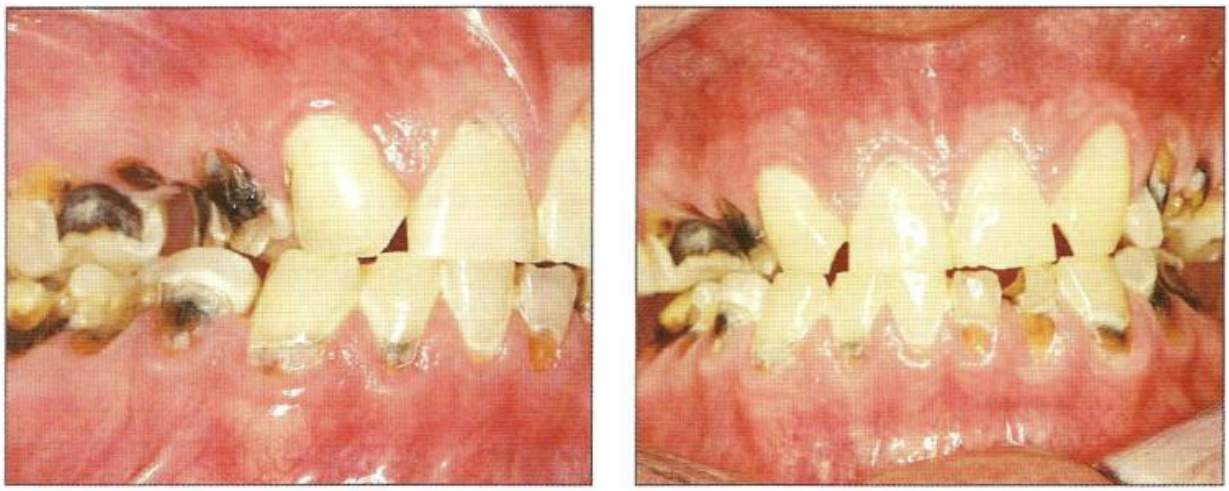


Figure 3.7: Photographs of the dental status of a 32-year old methamphetamine user. The severe caries on the canine and premolars as well as the excessive wear on the incisors are clearly visible. Photographs from Goodchild and Donaldson (2007).

The tooth decay may resemble baby-bottle caries (Goodchild and Donaldson 2007). Methamphetamine users also exhibit more missing teeth than non-users (Shetty et al. 2010).

Causes of oral health deterioration. Even though the effects of methamphetamine usage on oral health are dramatic, there is surprisingly little literature available besides anecdotal evidence. The exact cause for the deterioration is unknown, but it is suspected that it is a combination of the following behaviors (Hamamoto and Rhodus 2009):

- **Xerostomia**: Methamphetamine users suffer from dry mouth due to a lack of saliva production or chronic dehydration. Without the buffering capacity of saliva, the lactic acid produced by bacteria such as *Streptococcus mutans* can demineralize the tooth enamel at an even more rapid rate. Saliva is also an important vehicle to keep the biofilm accumulation at bay. Without sufficient saliva, increasing amounts of bacterial biofilm form.
- **Frequent consumption of sugar containing soft-drinks**: Due to the xerostomia and a craving for sugar, many methamphetamine users consume large amounts of non-diet sodas during an episode of drug consumption, which supply sugar and provide a substrate for lactic acid production by caries causing bacteria such as *Streptococcus mutans*.
- **Reduced/lacking oral hygiene**: Many methamphetamine users neglect oral and other hygiene during drug episodes. However, users are concerned about their appearance and brush in between episodes, which may lead to a slower decay than otherwise expected.
- **Bruxism**: Hyperactivity and neuromuscular activity in the jaw is common for methamphetamine episodes, which can lead to excessive wear on tooth surfaces.
- **Regurgitation**: Vomiting can worsen the tooth decay even further through the acidic pH of stomach acid.

Xerostomia may result from a sympathomimetic effect by stimulating the sympathic nervous system, which will cause reduced saliva secretion by stimulating inhibitory alpha-2 receptors. The plasma half life is approximately 12 hours which is much longer than for other drugs, for example cocaine (Cho and Melega 2002; Goodchild and Donaldson 2007). There are a number of other medical drugs which causes xerostomia, but the pattern of tooth decay differs to the one observed in methamphetamine consumers, which suggests that there are further factors influencing the tooth status than xerostomia (Hamamoto and Rhodus 2009).

The acidic pH of methamphetamine has been suspected to enhance the tooth decay. While some have argued that the pH drop of saliva in methamphetamine users is only from 7.4 to 6.9, which is above the clinical critical pH of 5.5 at which enamel demineralizes (Goodchild and Donaldson 2007), others have found that the pH of methamphetamine sold on the street can vary widely from 3.0 to 7.0 in a sample of size 29 in South Africa (Grobler et al. 2011). This suggests that the individual pH of a methamphetamine sample can have different effects on oral health. In the absence of an adequate amount of saliva through xerostomia, this drop in pH can further accelerate demineralization. In addition, it has been shown that intravenous injection of methamphetamine, which usually causes higher plasma levels, is associated with a significant increase in number of missing teeth compared to methamphetamine users who consume it by smoking only (Shetty et al. 2010). These observations suggest that the effect of methamphetamine is partly indirect, not exclusively through the oral exposure.

Two distinct patterns of methamphetamine consumption are reported in the literature which may have different effects on oral health. One distinguishes a) the chronic repeated self-administration during the day (small doses up to every 30 minutes, then rest during the night), and b) multiple doses during several days, a "binge" (Cho and Melega 2002). Especially during episodes of methamphetamine consumption that last for days, one would expect a negative outcome on oral health.

Effect on the immune system. Methamphetamine usage has been shown to be a risk factor for MRSA skin and soft tissue infections after controlling for sex, age, and race (Cohen et al. 2007). There is recent evidence that methamphetamine usage has a modulatory effect on the immune system. It causes direct immunosuppression on dendritic cells and macrophages (Talloczy et al. 2008). The molecular basis for this observation is thought to be an inhibition of endosomal acidification and therefore elimination of the pH gradient between organelle membranes due to the slightly alkaline nature of methamphetamine. Therefore, the antigen presentation and phagocytosis are impaired (Talloczy et al. 2008). It has an effect on viral as well as bacterial infections. In the mouse model it leads to a more rapid genital herpes simplex disease progression and leads to a disregulation of normal cytokine response (Valencia et al. 2012). Methamphetamine has been shown to reduce TNF-alpha levels in hepatocytes during hepatitis C infection (Ye et al. 2008). It also potentiates the effect of bacterial LPS, which causes increased IL-1beta production in human monocytes leading to an exacerbated form of periodontitis (Tipton et al. 2010).

Effect on oral bacteria. It is well known that oral bacteria are causing tooth decay, however to my knowledge there are no studies available that examine the composition of oral bacteria in methamphetamine users. There are very few studies which compared the supra- and subgingival plaque distribution of a selected number of organisms in khat users (Al-Hebshi et al. 2010). Khat is a cultivated plant in Yemen and Ethiopia that contains cathinone, an amphetamine-like substance (Al-Hebshi and Skaug 2005). In contrast to methamphetamine which is administered relatively pure, khat contains many other secondary plant products such as complex alkaloids and tannins and is usually chewed for hours a day (Al-Hebshi and Skaug 2005), which modulate the effect of cathinone on the oral bacterial composition through interactions with secondary metabolites as well as the constant friction on teeth and gums. This dissertation attempted to analyze the effect of methamphetamine usage on the oral microbiota.

3.2 Sample description

3.2.1 Overview

This chapter analyzes the bacterial composition of human saliva and the influence of age and regular tobacco smoking. Furthermore, the effect of alcohol consumption, and stimulant use was analyzed. Samples were identified from the salivary sample inventory of the Institute of Behavioral Genetics based on at least one of the following self-reported criteria:

- *Cigarette smoking*: daily cigarette smoking of a minimum of 20 cigarettes per day, smoking the day of the interview
- *Alcohol consumption*: reported alcohol consumption on 6 of 7 days
- *Stimulant use*: reported stimulant use within the last year (life time minimum usage is 5 times)
- *Controls*: no illicit drugs, no tobacco criterion, no recent alcohol, and no regular drinking

Overlap between the first three groups has been observed and therefore, I constructed new ternary (3) variables for each phenotype to account for this insufficient separation. Briefly: **Cigarette smoking** was coded as nonsmoker (0), irregular or former smoker (1), or regular smoker (2). Alcohol usage was divided into alcohol quantity and alcohol frequency. **Alcohol quantity** was coded as no alcohol (0), 1 to 4 days consuming alcohol the 7 days prior to the interview (1), or 5 to 7 days consuming alcohol the 7 days prior to the interview (2). Alcohol frequency was coded as no alcohol (0), 1 to 14 drinks in the 7 days prior to the interview (1), or 15 or more drinks in the 7 days prior to the interview (2). **Stimulant use** was grouped into no stimulant use (0), former stimulant use (1), and recent stimulant use (2).

3.2.2 Phenotype definitions

In order to assess the substance use habits of individuals in the various cohorts, different standardized questionnaires were administered. All individuals in this cohort received

the Composite International Diagnostic Interview Substance Abuse Module (CIDI-SAM) (Cottler et al. 1989; Cottler 2000), albeit different versions. Most individuals received the computerized version; however, some samples, especially samples collected before 2002, received one of two paper versions of the test. The CIDI-SAM supplement was administered to the majority of the cohort to assess substance experimentation. A list of the exact wording of the questions can be found in Appendix C.3.

Cigarette Smoking."Smoking" in the CIDI-SAM refers to tobacco cigarette smoking. Therefore, smoking, tobacco smoking and cigarette smoking are used interchangeably in this section. Other forms of smoking tobacco including cigars, pipes, or smoking other substances such as marijuana or illicit drugs are not examined. Tobacco use in the CIDI-SAM supplement includes other forms of tobacco than tobacco cigarettes, namely pipes, cigars, or chewing tobacco.

Nonsmokers (0): Individuals were coded as nonsmokers (0) if they fulfilled one of the following criteria:

- received the computerized CIDI-SAM and denied to have smoked more than 20 cigarettes in their lifetime (CIDI-SAM questions b1b, b1c, and b1d were not answered) and did not report any tobacco use within the last 180 days (CIDI-SAM supplement: samx1f or CIDI-SAM paper version: dsq1e)
- did not receive the computerized CIDI-SAM, but reported zero lifetime tobacco use (CIDI-SAM supplement: sam1a or CIDI-SAM paper version: dsq1a)

Irregular/Former Smokers (1): Individuals were coded as irregular or former smokers (1) if they fulfilled one of the following criteria:

- reported smoking in the computerized CIDI-SAM, but the last cigarette was more than one day ago (b1d = "2 to 6 days ago", "7 to 13 days ago", 14 to 20 days ago", "21 to 30 days ago", or "more than a month ago")
- reported smoking in the computerized CIDI-SAM, but the regular smoking pattern in the past 12 months was less than one day per week (b1b = "1 to 3 days a month" or "less than once a month")
- reported tobacco use at least once during the past 180 days (CIDI-SAM supplement: samx1f or CIDI-SAM paper version: dsq1e).

Regular Smokers (2): Individuals were coded as regular smokers (2) if they fulfilled all of the following criteria:

- received the computerized CIDI-SAM and reported smoking at least once per week (b1b = "every day", "5 or 6 days a week", "3 or 4 days a week", or "1 or 2 days a week")
- received the computerized CIDI-SAM and reported smoking the day prior or the day of the interview (b1d = "today" or "yesterday")

Alcohol consumption. The alcohol consumption measure is based on the last 7 days. The subjects were asked which and how many alcoholic beverages they consumed for each day of the last 7 days individually. The sum of drinking days and number of drinks was calculated. One drink equals one can or bottle of beer, one glass of wine or wine cooler, or one shot of hard liquor (individually or in mixed drinks). The alcohol variable was split into two measurements because alcohol consumption frequency does not necessarily correlate with alcohol quantity. Binge drinking, the consumption of large amounts of alcoholic drinks on single occasions, is common in young adults (Substance Abuse and Mental Health Services Administration 2011, pg. 28), while regular alcohol consumption of one to two drinks daily is recommended by the

American Heart Association due to its cardiovascular protective effects for adults (Pearson 1996).

No Alcohol (0): Individuals were coded as no alcohol (0) if they did not consume alcohol the week prior to the interview.

1-4 days (1): Individuals who consumed alcohol on 1, 2, 3, or 4 days the week prior were coded 1-4 days (1).

5-7 days (2): Individuals who consumed alcohol on 5, 6, or 7 days the week prior were coded 5-7 days (2).

Stimulant use. Ideally, our first intention was to restrict this study to the use of methamphetamines. However, street names overlap and it is therefore not possible to restrict my analysis to methamphetamine without testing the substance in a chemical laboratory. For example "speed" refers to amphetamine or methamphetamine and "uppers" refer to a number of stimulating drugs including methamphetamine and amphetamine (Goodchild and Donaldson 2007, http://www.urbandictionary.com/, accessed October 7, 2013). The definition of stimulants in the CIDI-SAM includes amphetamines, diet pills, ice, khat, methamphetamine, Ritalin, speed, and uppers. The questionnaire of the CIDI-SAM paper version and the supplement refer to the usage pattern of amphetamines, but none of the other substances. To unify the questionnaires, both are referred to as stimulants. The majority of the stimulants used in this cohort is likely methamphetamine (Professor Christian Hopfer, CU Denver, personal communication).

No stimulants (0): Individuals were coded as non-stimulant users (0) if they fulfilled all of the following criteria:

- did not report any lifetime stimulant use (CIDI-SAM supplement: samx5a or CIDI-SAM paper version: dsq5a)

- did report zero stimulant use during the past 180 days (CIDI-SAM supplement: samx5f or CIDI-SAM paper version: dsq5e)

Former stimulant use (1): Individuals were coded as former stimulant users (1) if they fulfilled any of the following criteria:

- received the computerized CIDI-SAM and reported using stimulants within the past 12 months (d2a2r = "past 30 days" or "not past 30 days, but in past 12 months"), but not within the last 180 days (CIDI-SAM supplement: samx5f or CIDI-SAM paper version: dsq5e)
- received the computerized CIDI-SAM and reported using stimulants prior to the preceding year (d2a2r = "more than 12 months ago")
- reported stimulant use at least once during the past 180 days (CIDI-SAM supplement: samx5f or CIDI-SAM paper version: dsq5e), but not within the past year in the computerized CIDI-SAM (d2a2r)

Recent stimulant use (2): Individuals were coded as recent stimulant users (2) if they fulfilled all of the following criteria:

- received the computerized CIDI-SAM and reported using stimulants within the past 12 months (d2a2r = "past 30 days" or "not past 30 days, but in past 12 months")
- reported stimulant use at least once during the past 180 days (CIDI-SAM supplement: samx5f or CIDI-SAM paper version: dsq5e)

The exact wordings used in the computerized CIDI-SAM, CIDI-SAM paper version and CIDI-SAM supplement can be found in Appendix C.

3.2.3 Characterization of the dataset

The saliva samples were collected between the years 1997 and 2008. They were derived from a number of different cohorts at CU-Boulder's Institute for Behavioral Genetics and CU-Denver's Division of Substance Dependence, including the Colorado community twin study (CTS), the Colorado longitudinal twin study (LTS)(Rhea et al. 2006; Rhea et al. 2013c), the Colorado Adoption Project (CAP) (Rhea et al. 2013a; Rhea et al. 2013b) and other studies focused on adolescent antisocial behavior and drug abuse (Stallings et al. 2003; Stallings et al. 2005; Kamens et al. 2013).

Informed consent was obtained from all individuals, and the anonymity of all subjects is ensured by separation of all personal information from the sample and replacing it with a numerical identifier. Written informed consent was obtained and the study was approved by the CU-Boulder Institutional Review Board and the Colorado Multiple Institution Review Board at CU-Denver. The use of both sets of anonymized data was approved by the CU-Boulder Institutional Review Board (protocol 0399.11). Sample collection and DNA extraction has been performed as described previously in the twin study (Chapter 2.2.1) and Appendix A.

Initially, this dataset was comprised of 247 individuals. Thirty seven samples were excluded for a number of reasons. First, 20 saliva samples were excluded because they did not yield the required minimum sample number of 498 reads. In addition, alpha diversity plots showed a reduction in diversities of samples obtained in the years 1997 and 2000 (Figure 3.8). As differences in sampling techniques could not be excluded, all samples prior to 2001 (n=17) were removed from the analysis, even though a preliminary experiment showed no reduction of alpha diversity of any method below the levels with Scope mouthwash (Appendix B, Figure B1). Furthermore, there was no drastic difference in mean age of the samples between the years 1997 (n=10, mean age: 16.3 years), 1998 (n=3, mean age: 18.2 years), 2000 (n=4, mean age: 17.8 years), 2001 (n=4, mean age: 17.3 years), and 2002 (n=10, mean age: 16.7 years).

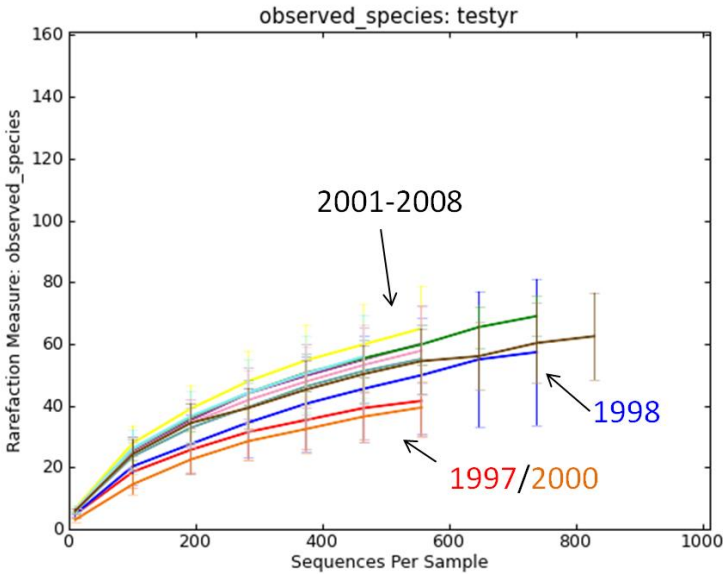


Figure 3.8: Reduced alpha diversity in samples prior to 2001. Alpha diversity rarefaction curves based on OTUs observed are plotted. The colors correspond to groups based on sample year. Samples from the years 1997-2000 (n=17) were excluded due to an unexplained reduction in alpha diversity of samples obtained in the years 1997 (n=10, mean age) and 2000 (n=4).

In addition, this dataset has an unintentional internal family structure. Of the 211 individuals, 63 have family members derived from 30 independent families. There are 28 family pairs, one family trio, and one family quartet represented. They are organized as following:

- 10 mother and father pairs
- 7 siblings pairs, both children are over 18 years old (including one DZ twin pair)
- 4 mother and child pairs (> 18 years)
- 2 father and child pairs (> 18 years)
- 2 mother and 2 children trios (> 18 years)
- 2 sibling pairs, where one is over 18 and the other under 18
- 1 father and child pair (< 18 years)
- 1 quartet of four siblings (> 18 years)
- father with the maternal grandmother (his "mother-in-law")

To assess the relationships, I ran Unweighted Pair Group Method with Arithmetic mean (UPGMA) based on the unweighted UniFrac values, which is a type of hierarchical clustering method using average linkage. Jackknifing (10 times) was performed to test robustness of the nodes. Most nodes were weak, which was supported by repeated analysis. No pair was repeatedly next to each other with a support > 25%. The only pair which was recovered next to each other in all repetitions (n=3), even with weak support, was the DZ twin pair. One member of this twin pair (S10490) was excluded from all analyses.

Table 3.2 shows the distribution of gender, smoking, alcohol phenotypes, and stimulant phenotypes of the final dataset with 210 individuals. The distribution of self-reported health status and race is also listed.

Table 3.2: Demographic characterization of the drug dataset. The total number of samples is 210.

Variable	Coding	n	%
Gender	Male	121	57.6
	Female	89	42.4
Health status	1 (excellent health)	44	21.0
	2 (very good health)	95	45.2
	3 (good health)	50	23.8
	4 (fair health)	11	5.2
	5 (poor health)	0	0.0
	NA	10	4.8
Race	Non-Hispanic White	162	77.1
	Hispanic	22	10.5
	American Indian	5	2.4
	African American	5	2.4
	Pacific Islander	2	1.0
	Multi Ethnic	14	6.7
Smoking	Nonsmoker	52	24.8
	Occasional smoker/Light smoker	20	9.5
	Regular Smoker	136	64.8
	NA	2	1.0
Alcohol Quantity	Zero drinks week prior	109	51.9
	1-14 drinks week prior	42	20.0
	15 or more drinks week prior	57	27.1
	NA	2	1.0
Alcohol Frequency	Zero drinking days week prior	109	51.9
	1-4 drinking days week prior	36	17.1
	5-7 drinking days week prior	63	30.0
	NA	2	1.0
Stimulant use	Nonuser	138	65.7
	Former user	20	9.5
	Recent user (last 6 months)	52	24.8

As seen in Table 3.2, the majority of the sample is non-Hispanic white with good to excellent health. About 25% are nonsmokers, while 65% are regular smokers. Compared to the general population (aged 12 or older) with ~25% smokers (Substance Abuse and Mental Health Services Administration 2011, pg. 39), this dataset has an excess of smokers, due to sample selection and comorbidities with other substances. The majority, 87%, of the regular smokers smoked at least 160 out of the past 180 days (Figure 3.9 A), and the majority smoked one pack with 20 cigarettes per day with a range from 2 to 60 cigarettes per day (Figure 3.9 B).

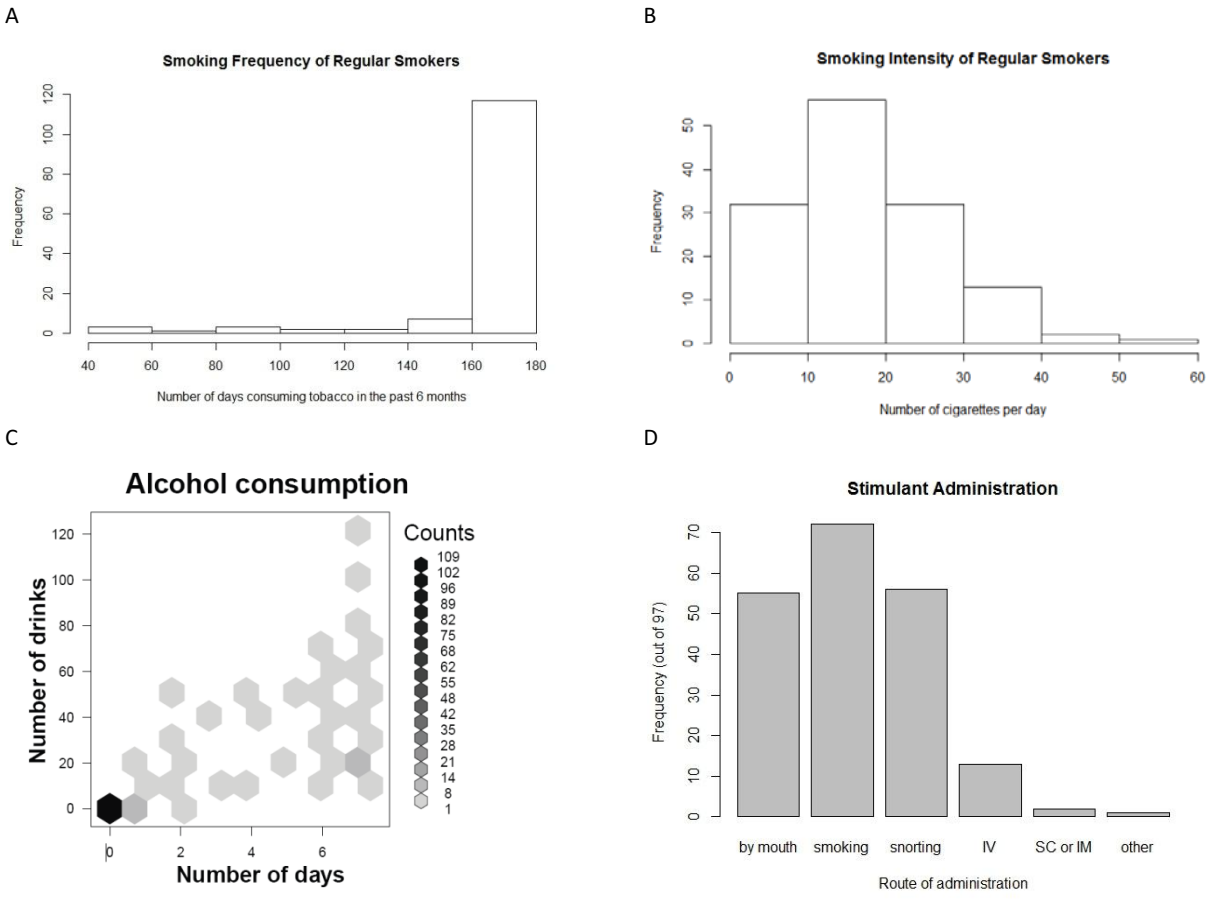


Figure 3.9: Smoking habits, alcohol consumption, and stimulant administration distribution. A and B depict the smoking habits of the regular smoker group (n=136). A) Frequency histogram of the number of days consuming tobacco in the past 6 months (range 45 to 180), B) Frequency histogram of the number of cigarettes per day (range: 2 to 60). C) Hexbin plot of number of drinking days and number of drinks of the full dataset (n=210). Number of days consuming alcohol in the past 7 days (range 0 to 7) and number of drinks consumed in the past 7 days (range 0 to 116). The darker the color, the more individuals are present in each group. D)

Distribution of administration routes of stimulants of current and former users (n=97). Multiple answers were possible. IV = intravenously, SC = subcutaneous, IM = intramuscular.

Approximately 50% of individuals have not consumed alcohol the week prior to testing. About 30% of the samples fell into the highest alcohol quantity or alcohol frequency group; however, these are not necessarily the same individuals, as Figure 3.9 C shows. A number of individuals drink regularly, but moderately, while some individuals exceed 20 drinks per day within a few days (binge drinking).

The majority of individuals in the US and in this dataset have never consumed stimulants in their life. Only 0.1% (353,000) of the US population were current methamphetamine users as defined by last month's usage in 2010 (Substance Abuse and Mental Health Services Administration 2011, pg. 13). In my selected dataset about 25% are recent stimulant users. I define recent as having consumed within the last 6 months. Only 23 individuals (11% of dataset) have consumed stimulants within the last month and 7 individuals have used stimulants the day prior to or the day of the assessment. Different routes of administration for stimulants exist, which include oral, smoking, snorting, injection into vein, skin or muscle, and others such as insertion into rectum or vagina (Figure 3.9 D). The main route of stimulant access in this dataset is via mouth, smoking, or snorting. As expected, a significant overlap and diversity in drug consumption habits is demonstrated in the dataset (Figure 3.10).

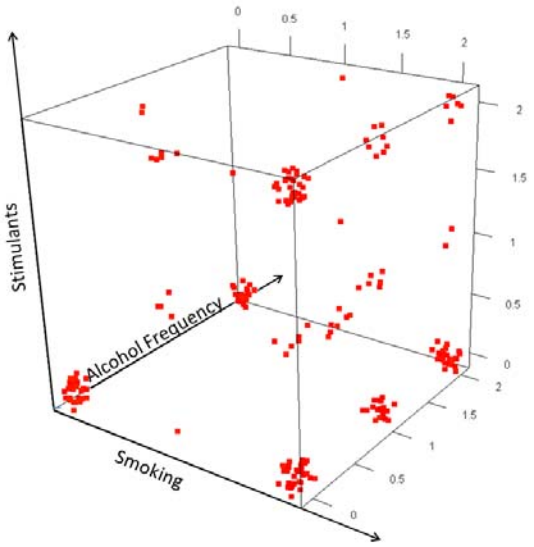


Figure 3.10: Overlapping drug consumption. This 3D scatterplot show the distribution of the three smoking, stimulant and alcohol consumption categories and their overlap. The categories are the following – Smoking: 0 = nonsmoker, 1 = occasional smoker, 2 = regular smoker; Stimulants: 0 = nonuser, 1 = former user, 2 = recent user; Alcohol Frequency: 0 = zero drinking days the week prior, 1 = 1-4 drinking days the week prior, 2 = 5-7 drinking days the week prior.

While there is an immense diversity of drug consumption combinations as Figure 3.10 shows, a few clusters can be identified. Besides the control cluster of nonsmokers without stimulant use or alcohol consumption, there are a few other common combinations. One is nonsmokers, who drink 5-7 days per week, no stimulants. Three clusters of smokers, who do not use stimulants, with all three alcohol drinking frequency are identified. Lastly, there is an accumulation of recent stimulant users, who smoked regularly but did not drink.

3.3 Results

3.3.1 Sequencing results and taxonomic overview

The final dataset analyzed was derived from 210 independent saliva samples, which yielded 201,165 sequence reads after quality control. The average read depth was 958 reads per sample with a range from 498 to 1980 reads per sample. The reads were binned based on a

minimum of 97% sequence identity and assigned to a total of 824 OTUs; taxonomy was assigned with Greengenes, version May 2013. Each sample contained, on average, 80 OTUs at the sequencing depth presented with a range from 33 to 138 OTUs. Two OTUs, OTU 239 and OTU 741, both assigned to the genus *Streptococcus*, were present in every sample. A plot of the main phyla and genera of the full dataset with the dataset split by age group and smoking status visualizes subtle differences (Figure 3.11 A and B).

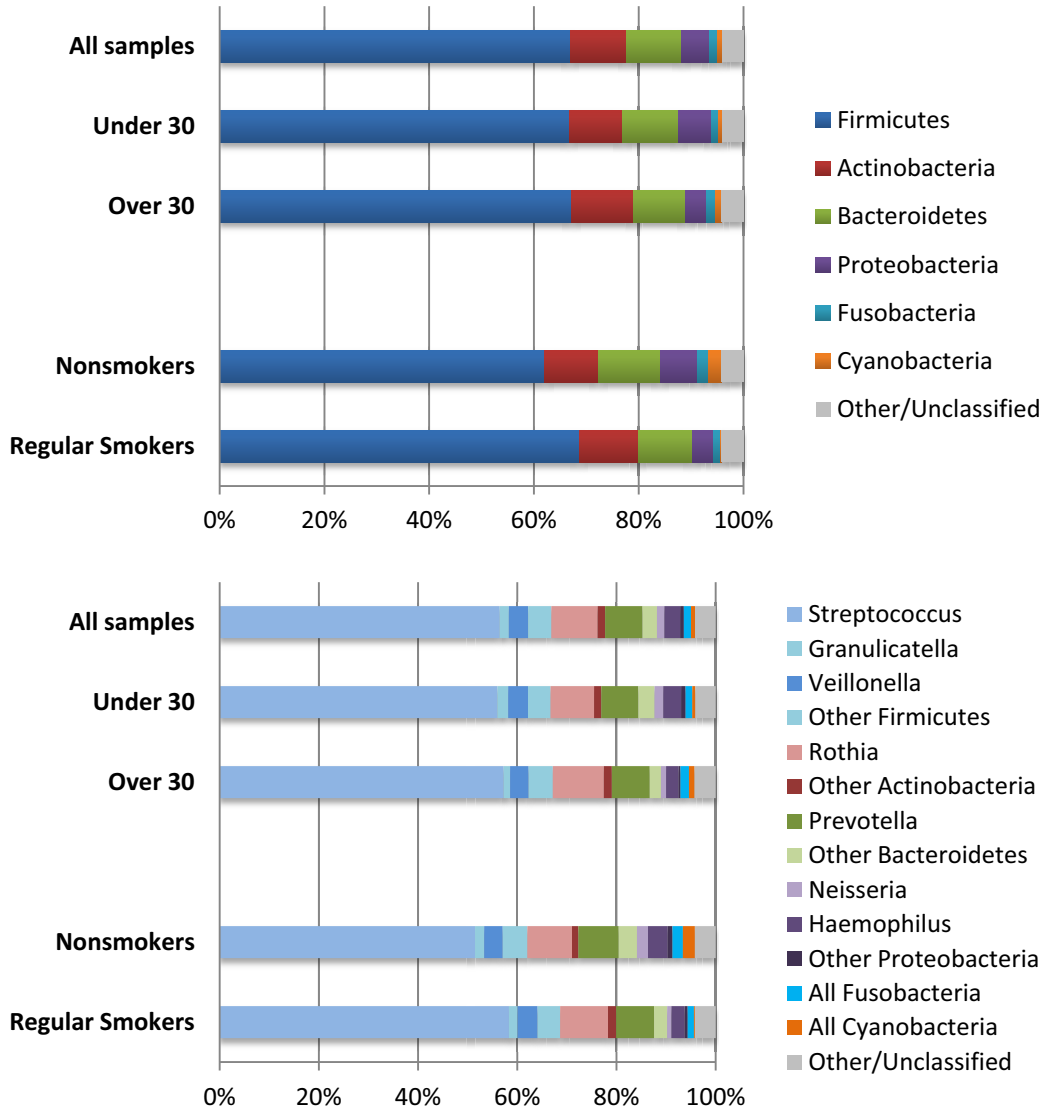


Figure 3.11: A) Average phyla distribution in full dataset (n=210), grouped by age (< 30 years: n=133; > 30 years: n=77), and smoking status (nonsmokers: n=52, regular smokers: n=136). B)

Average genera distribution in full dataset, grouped by age, and smoking status. Only the most common genera are shown (min 2% of sequences in one dataset).

The dataset is comprised of four main phyla (Firmicutes, Actinobacteria, Bacteroidetes, Proteobacteria) and two minor phyla (Fusobacteria and Cyanobacteria), but also contains sequences from members of Spirochaetes, Tenericutes, Synergistetes, Thermotogae, Chlorobi, Nitrospirae, Elusimicrobia, and Gemmatimonadetes and the candidate divisions TM7, SR1, OD1, GN02, WS5, WS3, and MVP-21. Visually, there are only minor changes in the main phyla and genera abundances between the two age groups. The smoking groups differ even on a gross level where regular smokers have an increase in the genus *Streptococci* and a reduction of *Haemophilus* and Cyanobacteria compared to nonsmokers.

3.3.2 Bacterial organisms of interest

Potential food contamination. The Cyanobacteria sequences in this dataset are mainly derived from the order of Streptophyta and are therefore, likely plant chloroplasts. The sequence blasts 99% identical to chloroplasts of common food grasses such as *Hordeum vulgare* (barely), *Zea mays* (maize or corn) or *Oryza sativa* (rice). Interestingly, the majority of reads (1438 out of 1564) come from only 7 saliva samples. One individual alone contributed 610 reads. This finding emphasizes the hypothesis of residual food debris, as they are seen sporadically and with high sequence counts in a few samples. Another potential food contamination could be the order Rhizobiales (Alpha-Proteobacteria) and members of Burkholderiales (Beta-Proteobacteria), which includes nitrogenfixing *Rhizobia* of legumes. However, only 7 reads were derived from Rhizobiales and the influence legume root nodule contamination is insignificant. Commonly, the root nodules are not consumed. Fifty-six total reads came from the order Burkholderiales, 50 of which were assigned to the genus *Lautropia* with only one known species *Lautropia mirabilis*, which is not associated with root nodules. This relatively unknown organism was significantly associated with oral health in a study of aggressive periodontitis in children (Shaddox et al. 2012) and adults (Colombo et al. 2009), and an indicator of therapeutic success in

periodontal treatment (Colombo et al. 2012). In my dataset, 13.8% of individuals harbored this organism enriched with young age (under 30) and nonsmokers (Table 3.4).

Potential human oral pathogens. Potential human oral and system wide pathogens were identified in the dataset by a BLAST search against the HOMD database (Chen et al. 2010).

The three members of the red complex, *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*, which are highly associated with severe periodontitis (Socransky et al. 1998), were identified as the first BLAST hit with a sequence identity of 100% (*Porphyromonas gingivalis*, OTU 611), 99.7% (*Treponema denticola*, OTU 567), and 98.4% (*Tannerella forsythia*, OTU 171), respectively. A heatmap of the relative abundances shows that especially OTU 611, assigned to *Porphyromonas gingivalis*, is common in a selected number of saliva samples (Figure 3.12).

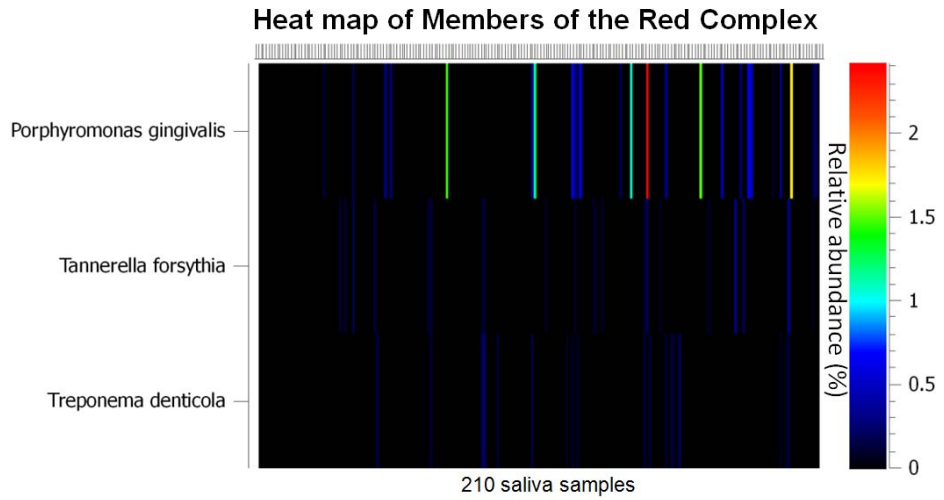


Figure 3.12: Heatmap of sequences which may correspond to the three members of the red complex associated with severe periodontitis (*Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*). The colors correspond to the relative abundance in percent.

Possibly, the individuals from whom the samples were obtained suffered from severe periodontitis, but information on oral health status of these individuals is not available. However, the sequence length and taxonomic resolution does not allow for a definite taxonomic

assignment and further studies with sequence specific primers or longer sequence reads are needed for further confirmation.

The BLAST search to the HOMD database (Chen et al. 2010) revealed further possible human oral pathogens, for example *Streptococcus mutans* (OTU 240) and *Scardovia wiggsiae* (OTU 920), which are associated with severe childhood caries (Tanner et al. 2011). A large number of organisms were found which are also potentially associated with periodontitis, including members of the orange complex (*Prevotella nigrescens* – OTU 259 and *Prevotella intermedia* - OTU 466, and *Fusobacterium nucleatum/periodonticum* –OTUs 712, 720, 729, 773) (Socransky et al. 1998). All reported members were the top BLAST hit and had a sequence identity of 98% or more. It is important to stress again, that the current level of taxonomic resolution does not allow for assignment to species and therefore, these results need further validation.

3.3.3 Alpha diversity

Alpha diversity was compared between different groups determined by smoking status, gender, age group, alcohol and stimulant consumption, but I observed no significant differences. I cannot conclude that no differences exist, but they were not detectable at my current level of sequencing depth and taxonomic resolution. Kumar (2011) found a decrease in alpha diversity in the first few days of plaque development (marginal as well as subgingival) after professional tooth cleaning, however the effect decreases during their 7 day study and is non- detectable at the end of the week, which suggests that the effect is transient. This cohort was not controlled for dental health.

3.3.4 Age increases the beta diversity of the oral microbiota

A development of the oral microbiota with age has been reported in children (Crielaard et al. 2011), but not adolescents and adults. In order to assess the distribution of age, the PCoA of the unweighted UniFrac distance matrix was visualized. Unexpectedly, this dataset exhibits a

clear age gradient along PC1 and PC2, where adolescents pictured in red and orange are spatially closer and therefore, more similar than individuals aged 40-65 depicted in blue colors (Figure 3.13).

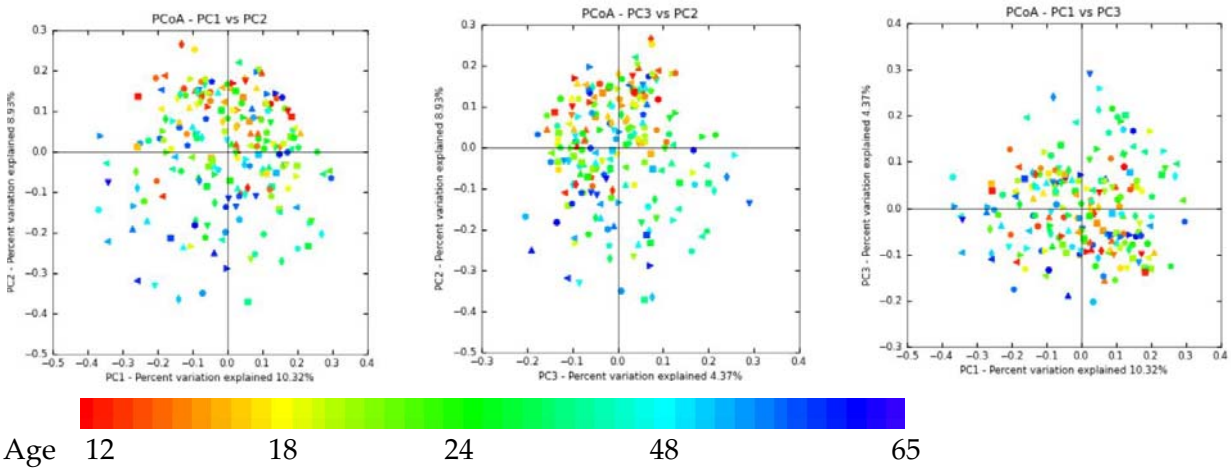


Figure 3.13: Uneven distribution of ages using PCoA of unweighted UniFrac distance matrix. Adolescents (age 13-19) in red and yellow colors are more similar than their parents' generation (40-60 years in age in blue colors).Each point corresponds to a sample and the distances correspond to beta diversity values. The percentage of variation explained by the PCs is indicated on the axes.

The age histogram reveals a bimodal age distribution, which corresponds to the collection of the proband generation (adolescents and young adults) with a peak age around 20 and their parents' generation with a peak age around 50 (Figure 3.14 A). A split of the dataset at age 30, to separate adolescence and young adults from adults demonstrates a clear difference in averaged, unweighted UniFrac values where individuals younger than 30 (n=133) are more similar to each other than individuals over 30 (n=77), see Figure 3.14 B. To my knowledge, this effect has never been shown before.

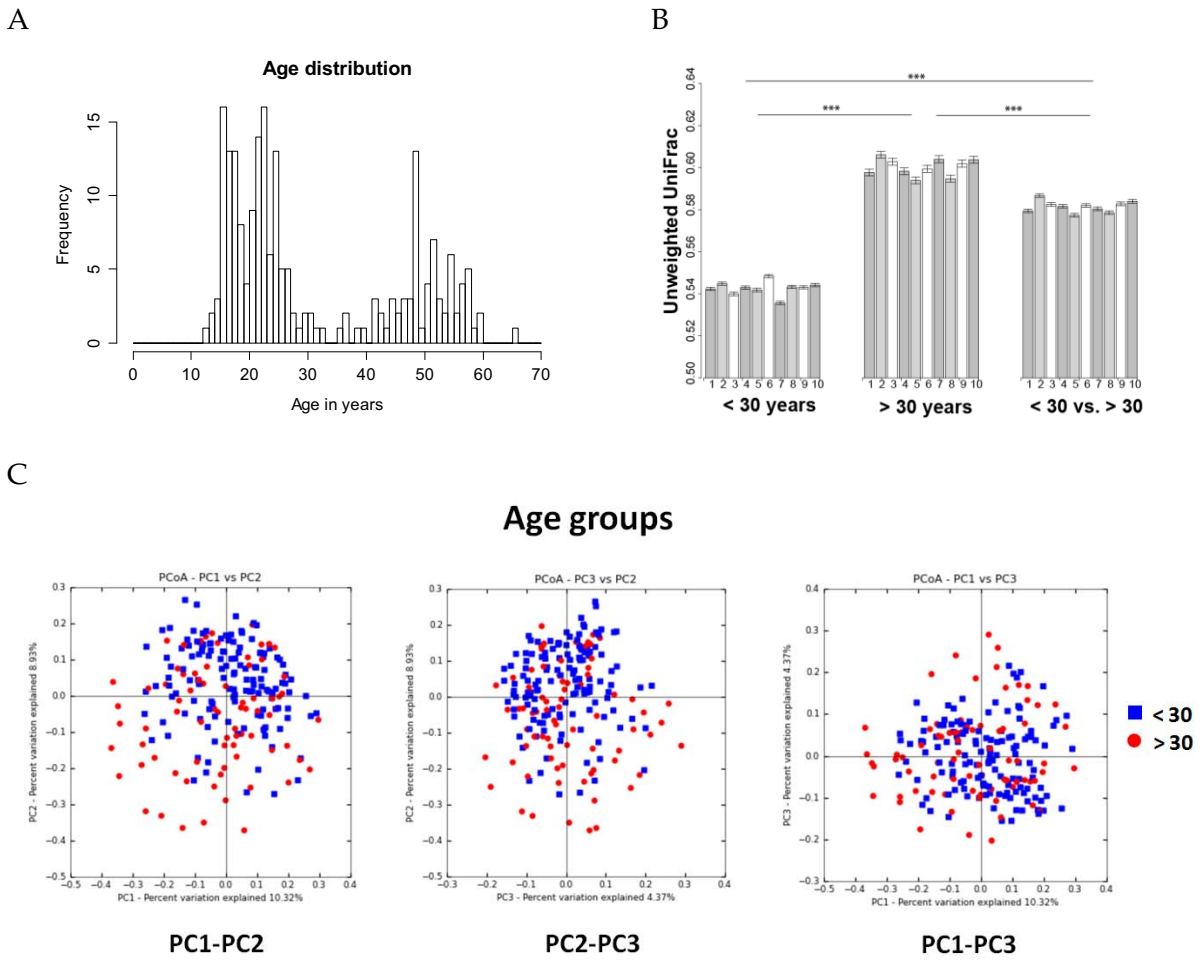


Figure 3.14: Differences in beta diversity based on bimodal age distribution A) Frequency histogram of ages in the dataset (n=210). The bimodal distribution shows peak ages around 20 and 50 years of age. B) Age effect on the salivary microbiome below age 30 (n=133), above age 30 (n=77) and between. Averaged pairwise unweighted UniFrac distances of individuals below age 30 (n=8778 pairs), above age 30 (n=2926 pairs) and between individuals below and above age 30 (n=10241 pairs) (± SEM). The data set was randomly subsampled 10 times at a sequencing depth of 498 sequences/sample, and each subsampling is shown as a separate bar. The statistical analysis was a Mann-Whitney U-test. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10 rarefactions is presented. C) Distribution of age groups in PCoA plot of unweighted UniFrac distance matrix. Individuals under 30 are depicted in blue (n=133), over 30 in red (n=77). Each point corresponds to a sample and the distances correspond to beta diversity values. The percentage of variation explained by the PCs is indicated on the axes.

Separation of age groups

To clarify the effect of these "age groups", the dataset was split and each part was analyzed separately. Figure 3.14 C shows the PCoA plots of the full dataset (A), colored by age group, which reveals an increased density of younger individuals especially along PC2, whereas older individuals are more dispersed and exhibit reduced similarity. However, there is no clear separation but an overlap between these age groups. Next, individual calculations of beta diversity for each of the two age groups were performed. Figure 3.15 shows that the age gradient is reduced in the age group specific plots, but it is not eliminated.

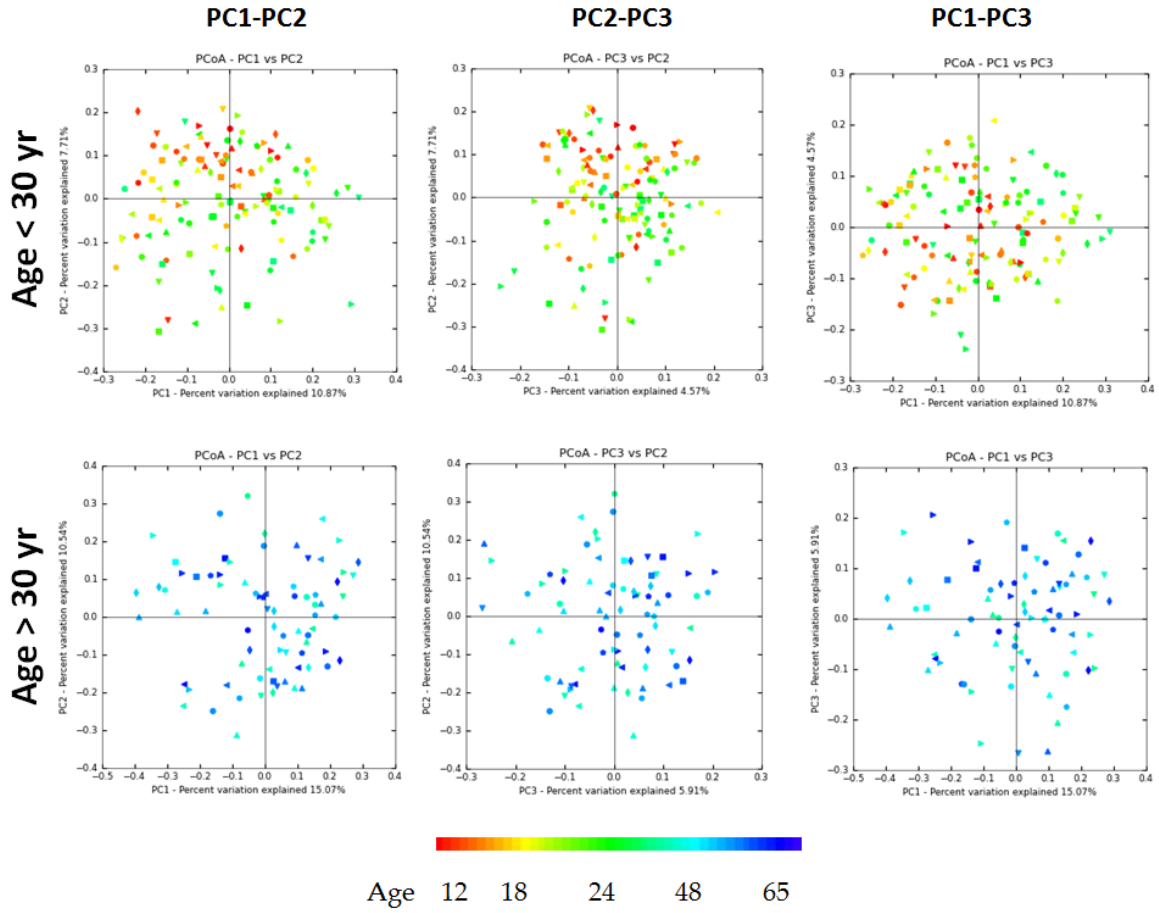


Figure 3.15: Distribution of ages in subgroups under (n=133, top panels) and over 30 years (n=77, bottom panels) based on PCoA of unweighted UniFrac distance matrices. A slight age gradient is still visible. Each point corresponds to a sample and the distances correspond to beta diversity values. The percentage of variation explained by the PCs is indicated on the axes.

3.3.5 Smoking increases the beta diversity of the oral microbiota

Tobacco consumption is known to have an effect on the oral microbiota (reviewed in Chapter 3.1.2). I therefore analyzed the effect of tobacco smoking on the full dataset, as well as on the individual age groups. Nonsmokers as well as irregular smokers clearly cluster closer in my dataset than regular smokers. This increase in similarity is observable in the PCoA plots along PC1 and PC2, but also in the results of the averaged beta diversities (Figure 3.16 A and B).

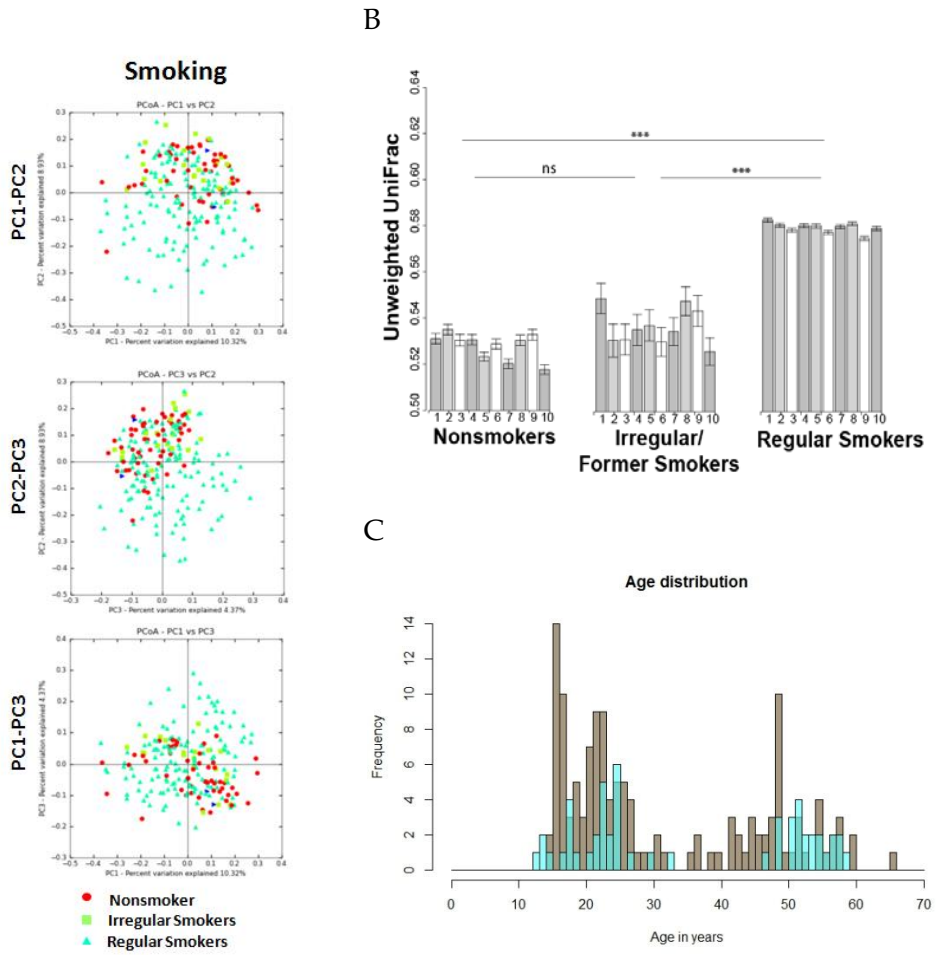


Figure 3.16: Nonsmokers (n=52) and irregular smokers (n=20) are more similar to themselves than regular smokers (n=136). A) Distribution of nonsmokers (red), irregular smokers (green) and regular smokers (blue) based on PCoA of unweighted UniFrac distances. Nonsmokers and irregular smokers are group closer than regular smokers. Each point corresponds to a sample and the distances correspond to beta diversity values. The percentage of variation explained by the PCs is indicated on the axes. B) Similarities based on smoking status (nonsmokers, irregular

smokers, regular smokers). Pairwise unweighted UniFrac distances between nonsmokers (n=1326 pairs), irregular smokers (n=190 pairs), and regular smokers (n=9180 pairs) were calculated (±SEM). The data set was randomly subsampled 10 times at a sequencing depth of 498 sequences/sample. Each subsampling is shown as a separate bar. The statistical analysis was a Mann-Whitney U-test. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10 rarefactions is presented. C) Frequency histogram, color coded by nonsmokers (n=52, blue) and regular smokers (n=136, brown), shows a similar age distribution for nonsmokers and regular smokers.

Smoking yields an increase in dissimilarity in the full dataset based on PCoA plots and averaged beta diversities similar to age. While the difference between nonsmokers and irregular smokers is not significant, both groups are significantly less diverse than regular smokers (Figure 3.16 B). As Figure 3.16 C shows, the age distribution between regular smokers and nonsmokers is similar. There are 21.1% (n= 28) nonsmokers, 15.0% (n=20) irregular smokers, and 62.4% (n=83) regular smokers in the young dataset and 31.2% (n=24) nonsmokers, no irregular smokers, and 68.8% regular smokers (n=53) in the old dataset. However, the patterns of nonsmokers overlap with the pattern of younger individuals in Figure 3.13, which could cause this correlated effect. I therefore utilized the split age group dataset to examine the smoking effect further.

Based on unweighted UniFrac distances and assessed by visual PCoA plot inspection, as well as averaged beta diversities (Figure 3.17 A, B, C, D), nonsmokers group significantly closer than smokers in the stratified age groups independently. Interestingly, the irregular/former smokers are not significantly different from nonsmokers in the full dataset and the younger dataset. This suggests that the recentness of tobacco consumption is crucial and the long-term effects are relatively low. There are no irregular/former smokers in the older dataset based on my definition.

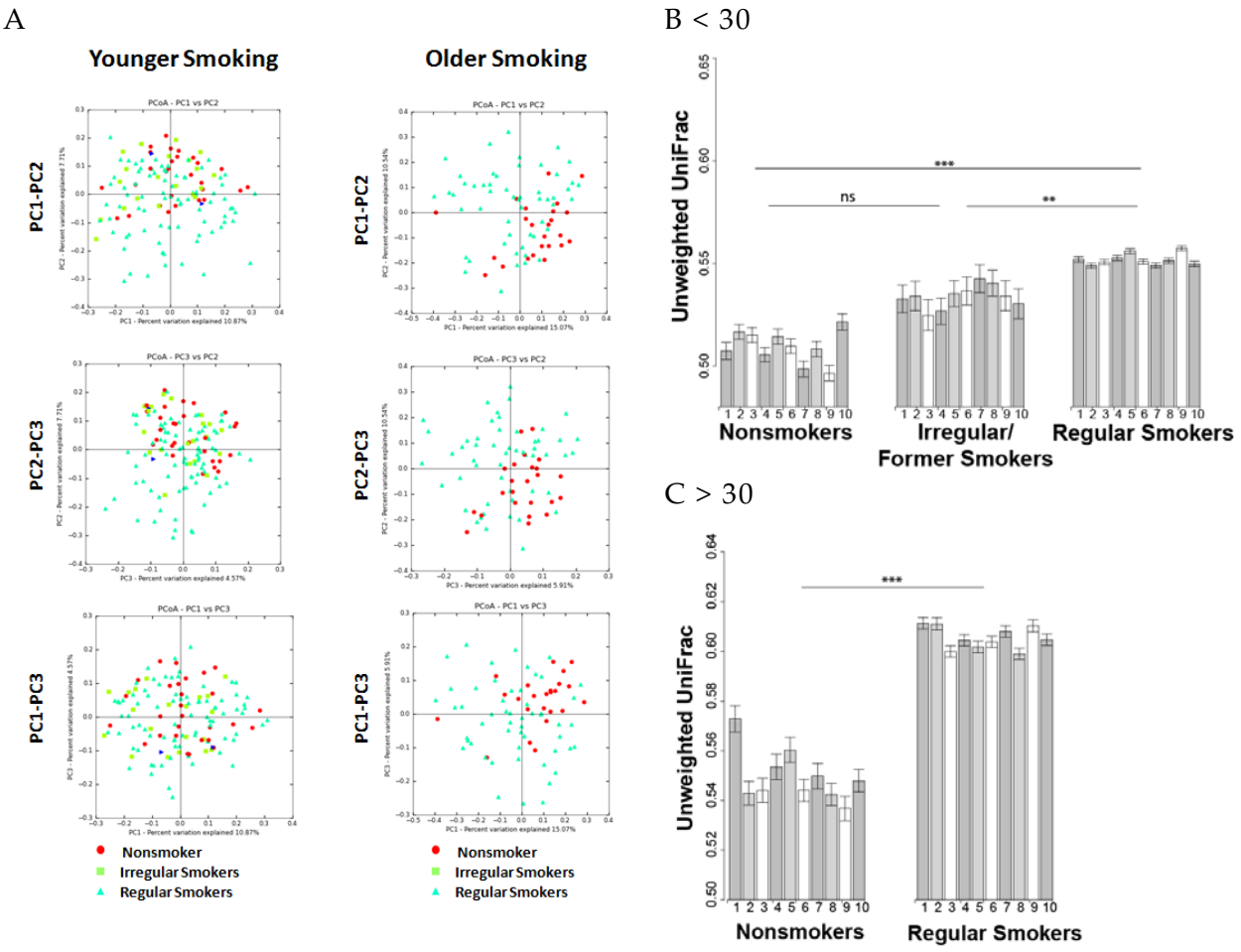


Figure 3.17: Smoking effect is demonstrated in both age groups. A, B) PCoA plots of the unweighted UniFrac reveals uneven distribution for nonsmokers in the subset of individuals A) < 30 years (nonsmokers (n=28), irregular smokers (n=20), and regular smokers (n=83)) and B) > 30 years (nonsmokers (n=24) and regular smokers (n=53)). There are no irregular smokers in the age group over 30. Each point corresponds to a sample and the distances correspond to beta diversity values. The percentage of variation explained by the PCs is indicated on the axes. C) Averaged pairwise unweighted UniFrac distances of nonsmokers (n=378 pairs), irregular smokers (n=190 pairs), and regular smokers (n=3403 pairs)) in the younger cohort and D) averaged pairwise unweighted UniFrac distances of nonsmokers (n=276 pairs) and regular smokers (n=1378 pairs) in the adult dataset (± SEM). The data set was randomly subsampled 10 times at a sequencing depth of 498 sequences/sample, and each subsampling is shown as a separate bar. The statistical analysis was conducted using a Mann-Whitney U-test. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10 rarefactions is presented. Please note the change in scale between C and D.

3.3.6 Multiple linear regression reveals independent age and smoking effects

In order to assess the effects of different phenotypes independently without the need for further stratification, I applied multiple linear regressions on the first three principal coordinates PC1, PC2, and PC3 individually. The first three PCs explain by definition the highest proportion of variance. For the full dataset, the first three PCs explain approximately 10%, 9%, and 4% of variance, respectively. In the dataset consisting of individuals younger than 30 years, the first three PCs explain approximately 11%, 8% and 4.5% of variance. Finally, in the dataset consisting of individuals aged 30 and up, the first three PCs explain approximately 15%, 11% and 6% of variance, respectively. In the discovery model, the phenotypes age, gender, race, smoking, alcohol and stimulant use as explanatory variables and each PC as dependent variable were used.

The initial model took the form:

$$PC_i = \beta_1 * age + \beta_2 * gender + \beta_3 * race + \beta_4 * smoking + \beta_5 * alcohol + \beta_6 * stimulant + \varepsilon_i$$

with i = 1, 2, 3. Gender, race, smoking, alcohol, and stimulants were modeled as unordered factors and age as continuous numeric variable.

Race was split into Non-Hispanic White and Others ("non white"). The three smoking (nonsmoker, irregular, regular) and stimulant groups were used, as defined earlier. I modeled alcohol usage based on alcohol frequency, alcohol quantity and as a binary alcohol variable (have you used alcohol the week prior? yes/no), but alcohol use did not explain significant parts of the model. The same was observed for stimulant use and therefore, these two variables were dropped from the model.

The final model consisted of the following covariates: age, gender, smoking (nonsmoker, former/irregular, current), and race (Non-Hispanic White, Other), see Table 3.3.

Table 3.3: Results of multiple linear regression for PC1, PC2 and PC3 according to the model PC = age + gender + smoking + race + ε for the full dataset (n=206) and the two stratified age groups individually (< 30 years: n=129, >30 years: n=77). Due to missing values, four data points were excluded from the multiple linear regression. The reference group for gender was female; the reference group for race was white. The reference group for smoking was nonsmoker (0) with factors irregular smoker (1) and regular smoker (2). The individual p-value outcomes for the covariates are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$. Multiple R^2 for the Full dataset PC1: 0.03781, PC2: 0.3009, PC3: 0.07453 with model p-values: 0.1696, 3.713×10^{-14}, 0.008063, respectively. Multiple R^2 for the younger dataset PC1: 0.03682, PC2: 0.3091, PC3: 0.101 with model p-values: 0.4575, 9.025×10^{-09}, 0.02107. Multiple R^2 for the older dataset PC1: 0.04374, 0.2317, 0.1677 with model p-values: 0.5145, 0.0007084, 0.009482.

Full dataset

PC1	Coefficient	P Value	Significance
Intercept	0.0843263	0.0156	*
age	-0.0015585	0.0265	*
gender	-0.0062325	0.7588	ns
smoking (1)	-0.0435123	0.2756	ns
smoking (2)	-0.0450979	0.0572	ns
race (non white)	0.0010109	0.966	ns

PC2	Coefficient	P Value	Significance
Intercept	0.1683307	3.54E-09	***
age	-0.0032953	9.32E-09	***
gender	0.0060484	0.705	ns
smoking (1)	-0.0164214	0.601	ns
smoking (2)	-0.1149186	3.47E-09	***
race (non white)	0.0280808	0.134	ns

PC3	Coefficient	P Value	Significance
Intercept	-0.0485535	0.02824	*
age	0.0002067	0.64158	ns
gender	-0.0007171	0.95567	ns
smoking (1)	0.0438663	0.08446	ns
smoking (2)	0.048236	0.00151	**
race (non white)	0.0340535	0.0249	*

< 30 years

PC1	Coefficient	P Value	Significance
Intercept	-0.039776	0.559	ns
age	0.004442	0.134	ns
gender	-0.019227	0.426	ns
smoking (1)	-0.030483	0.438	ns
smoking (2)	-0.030466	0.296	ns
race (non white)	0.001501	0.955	ns

PC2	Coefficient	P Value	Significance
Intercept	0.308078	7.31E-09	***
age	-0.011366	5.52E-07	***
gender	0.002649	0.880256	ns
smoking (1)	0.005918	0.836246	ns
smoking (2)	-0.08278	0.000154	***
race (non white)	0.03242	0.09984	ns

PC3	Coefficient	P Value	Significance
Intercept	-0.123378	0.00462	**
age	0.004448	0.01799	*
gender	-0.009642	0.52538	ns
smoking (1)	0.03974	0.1094	ns
smoking (2)	0.036981	0.04529	*
race (non white)	0.03525	0.03867	*

> 30 years

PC1	Coefficient	P Value	Significance
Intercept	0.174547	0.238	ns
age	-0.003433	0.214	ns
gender	0.015935	0.683	ns
smoking (2)	-0.063878	0.125	ns
race (non white)	0.006454	0.894	ns

PC2	Coefficient	P Value	Significance
Intercept	0.1377079	0.247	ns
age	-0.0019737	0.373	ns
gender	-0.0005636	0.986	ns
smoking (2)	-0.1530981	1.66E-05	***
race (non white)	0.023693	0.543	ns

PC3	Coefficient	P Value	Significance
Intercept	0.119312	0.1902	ns
age	-0.003494	0.0416	*
gender	0.025712	0.2849	ns
smoking (2)	0.053795	0.0372	*
race (non white)	0.030752	0.3032	ns

It is clearly visible that PC2 has the greatest explanatory potential for all three datasets. Overall, the results of the multiple linear regression shows that only age, regular smoking (2) and race have significant effects in the model, while gender and irregular smoking (1) never reach significance. PC1 does not have significant explanatory power and age barely reaches significance in the full dataset. The overall p-values are not significant for PC1. In contrast, PC2 is highly significant for all three datasets. Significant variables in the model are age and regular smoking, except for the older dataset where smoking, but not age is significant. As expected from the visual inspection of the PCoA plots, increasing age and smoking drive the PC2 in the same direction. PC3 reaches significance and the influencing factors are age for the individual age groups (<30 and >30), but not the full dataset, regular smoking for all three and race for the

full dataset and the younger, but not the older. Based on the R^2 values, very little variance is explained, except for PC2. However, due to the many factors that can possibly influence the oral microbiota, it is surprising that the model can even explain the percentages shown. This result is in accordance to the results obtained with the averaged beta diversities and PCoA. Based on these results, studies of the oral microbiota need control for age, race, and regular smoking, but not irregular smoking.

3.3.7 Changes in the core microbiota based on age and smoking status

In order to assess a shared core microbiota in this selected dataset, genera were visualized based on their sorted presence in the full dataset and further stratified based on age group and smoking status. As Table 3.4 demonstrates, younger individuals below the age of 30 as well as nonsmokers share a higher proportion of genera, especially more common genera than older individuals or regular smokers do.

Table 3.4: Presence of bacterial genera based on occurrence in the full dataset (n=210) and stratified into subgroups based on age (< 30 years: n=133; > 30 years: n=77), and smoking status (nonsmokers: n=52, regular smokers: n=136). The color codes represent the rings in Figure 2.3 where blue refers to genera which appear in ≥ 95% of samples, genera in yellow appear in 50% to 94.9% of samples, and genera in green appear in 5% to 49.9% of samples. Genera that appear in less than 5% are not highlighted or not shown. Genera, which appear in 100% of samples are highlighted in bold. Sequences that could not be assigned to genus level with Greengenes (version May 2013) are not shown.

Taxon	*All samples*	*Under 30*	*Over 30*	*Nonsmokers*	*Regular smokers*
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Rothia	99.5	**100.0**	98.7	**100.0**	99.3
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella	99.0	99.2	98.7	**100.0**	99.3
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella	98.6	98.5	98.7	**100.0**	97.8
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella	98.1	100.0	94.8	**100.0**	97.1
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces	91.4	93.2	88.3	86.0	94.1
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus	89.5	97.7	75.3	96.5	85.3
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium	88.1	89.5	85.7	94.7	86.0
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas	80.0	87.2	67.5	82.5	78.7
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia	76.2	84.2	62.3	87.7	69.9
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria	65.7	75.2	49.4	87.7	55.1
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium	58.1	54.9	63.6	56.1	62.5
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Aggregatibacter	50.0	57.9	36.4	61.4	47.8
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Oribacterium	46.2	48.9	41.6	54.4	41.2
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga	44.8	48.1	39.0	64.9	36.8
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter	44.8	48.9	37.7	54.4	40.4
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	40.0	45.1	31.2	38.6	37.5
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megasphaera	38.6	35.3	44.2	42.1	39.0
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Actinobacillus	37.6	45.9	23.4	54.4	27.9
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Abiotrophia	36.7	47.4	18.2	40.4	32.4
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Parvimonas	34.8	32.3	39.0	31.6	37.5
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia	34.8	33.1	37.7	36.8	36.0
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Moryella	30.0	24.8	39.0	35.1	27.9
k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__Gemella	27.6	33.1	18.2	33.3	25.0
k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema	27.6	26.3	29.9	15.8	33.8
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas	27.1	30.8	20.8	31.6	24.3
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus	24.3	16.5	37.7	10.5	33.1
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Filifactor	23.8	19.5	31.2	5.3	33.1
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Lactococcus	18.6	17.3	20.8	12.3	22.1
k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta (food contamination)	17.1	12.8	24.7	17.5	16.9
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus	17.1	16.5	18.2	22.8	16.9
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella	17.1	20.3	11.7	21.1	16.2
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister	16.7	17.3	15.6	12.3	19.1
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia	13.8	17.3	7.8	22.8	9.6
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Catonella	13.3	12.8	14.3	12.3	13.2
k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Dethiosulfovibrionaceae;g__TG5	12.4	12.0	13.0	0.0	17.6
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Mycoplasma	12.4	8.3	19.5	5.3	16.2
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Tannerella	11.0	13.5	6.5	12.3	10.3
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia	10.0	11.3	7.8	3.5	12.5
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];g__Mogibacterium	8.1	7.5	9.1	10.5	8.1
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella	8.1	9.8	5.2	14.0	4.4
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Paludibacter	7.1	2.3	15.6	7.0	7.4
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Peptococcus	4.8	4.5	5.2	5.3	5.1
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Cardiobacteriales;f__Cardiobacteriaceae;g__Cardiobacterium	4.8	5.3	3.9	7.0	3.7
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Acholeplasmatales;f__Acholeplasmataceae;g__Acholeplasma	3.3	3.8	2.6	0.0	5.1

This finding supports the earlier results of an increased beta diversity of older individuals and regular smokers. The lower diversity in younger individuals and nonsmokers could be an indication of health. The shared genera of this dataset were compared to the twin dataset presented in Chapter 2 and there is significant overlap (see also Figure 2.3). The Firmicutes *Streptococcus*, *Veillonella*, and *Granulicatella* were found in ≥ 95% of samples in both the twin dataset and the full drug dataset. The reduced presence of *Gemella* can be explained by

differences in taxonomy assignment, as there is a high abundance of the family Gemellaceae (94.8%) in the full drug dataset where a genus could not be assigned. The presence of *Rothia* and *Prevotella* is comparable between the datasets. The presence of *Fusobacterium* and *Neisseria* is reduced in the overall dataset, especially in the smokers compared to the twin dataset. The majority of the twins in Chapter 2 were nonsmokers, so the reduction of these organisms could be a consequence of exposure to tobacco smoke.

3.3.8 Specific bacterial organisms respond to age and smoking status

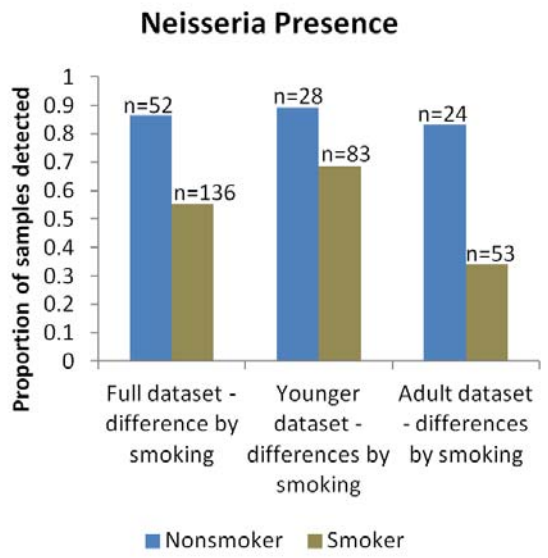
The two part statistic (Wagner et al. 2011) implemented in Explicet (Robertson et al. 2013b) was utilized to address the question whether certain genera experience change in abundance or presence or both based on the smoking behavior or age. Only taxonomic groups were included which could be assigned to genus level with the Greengenes database (Version May 2013). A total of 22 genera were influenced by smoking or age group (Table 3.5).

Table 3.5: Genera influenced by smoking or age group. The two part test (Wagner et al. 2011) was utilized to identify genera which changed by smoking status in the full dataset, as well as in the individual age groups. "Increase" refers to an increase from nonsmokers to smokers, and "decrease" refers to a decrease from nonsmokers to smokers, or younger to older, respectively. Genera had to occur in >5% of samples (n=11) to be included in the test. In addition, genera which changed significantly between the two age groups split at age 30 are shown. All reported genera have a p-value < 0.05 and changes with a p-value < 0.001 are bold. The results are not corrected for multiple testing.

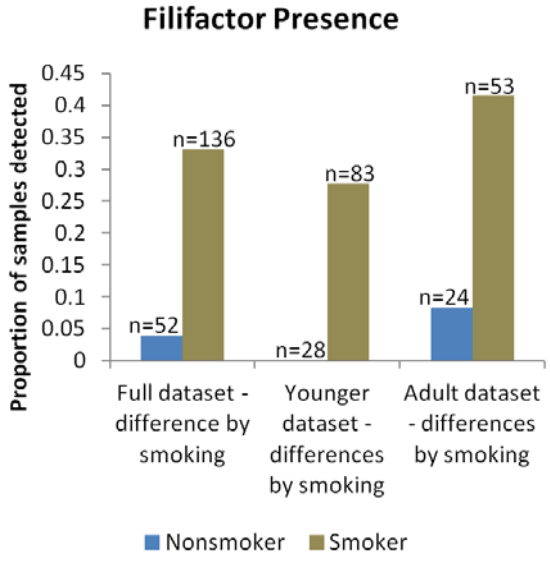
		Full dataset - difference by smoking		Younger dataset - differences by smoking		Adult dataset - differences by smoking		Differences in Age groups	
Phylum	**Genus**	**Abundance**	**Presence**	**Abundance**	**Presence**	**Abundance**	**Presence**	**Abundance**	**Presence**
Actinobacteria	*Rothia*			increase	no change				
Bacteroidetes	*Paludibacter*							increase	increase
Bacteroidetes	*Porphyromonas*							decrease	decrease
Bacteroidetes	*Capnocytophaga*	decrease	decrease			decrease	decrease		
Firmicutes	*Abiotrophia*							**decrease**	**decrease**
Firmicutes	*Granulicatella*					decrease	decrease	**decrease**	**decrease**
Firmicutes	*Lactobacillus*	increase	increase			decrease	increase	decrease	increase
Firmicutes	*Lactococcus*			increase	increase				
Firmicutes	*Streptococcus*	increase	no change						
Firmicutes	*Moryella*							increase	increase
Firmicutes	*Oribacterium*	decrease	decrease						
Firmicutes	*Filifactor*	**decrease**	**increase**	increase	increase	decrease	increase		
Fusobacteria	*Fusobacterium*	**decrease**	**decrease**			decrease	decrease		
Fusobacteria	*Leptotrichia*							increase	decrease
Proteobacteria	*Lautropia*	increase	decrease						
Proteobacteria	*Neisseria*	**decrease**	**decrease**	decrease	decrease	**decrease**	**decrease**	**decrease**	**decrease**
Proteobacteria	*Actinobacillus*	increase	decrease	increase	decrease			**decrease**	**decrease**
Proteobacteria	*Aggregatibacter*	decrease	decrease					**decrease**	**decrease**
Proteobacteria	*Haemophilus*	decrease	decrease			decrease	decrease	increase	decrease
Spirochaetes	*Treponema*	increase	increase						
Synergistes	*TG5 (Dethiosulfovibrionaceae)*	decrease	increase						
Tenericutes	*Mycoplasma*					increase	increase		

Most effects observed are decreases in abundance and/or presence with age and smoking, which suggests that common genera including the core genera are susceptible to smoking and aging. Only very few genera increase with smoking, which suggests that presumably healthy nonsmokers are more similar to other nonsmokers than smokers are to other smokers. Similar to the beta diversity results, there are some overlaps between smoking and age but overall lower than indicated in the beta diversity results. A number of genera change significantly in the full dataset, the younger dataset, or the adult dataset with smoking, but not consistently. This could be due to lack of statistical power or due to an age effect. The only two organisms that change consistently and significantly in all smoking groups are *Neisseria* and *Filifactor* (Figure 3.18).

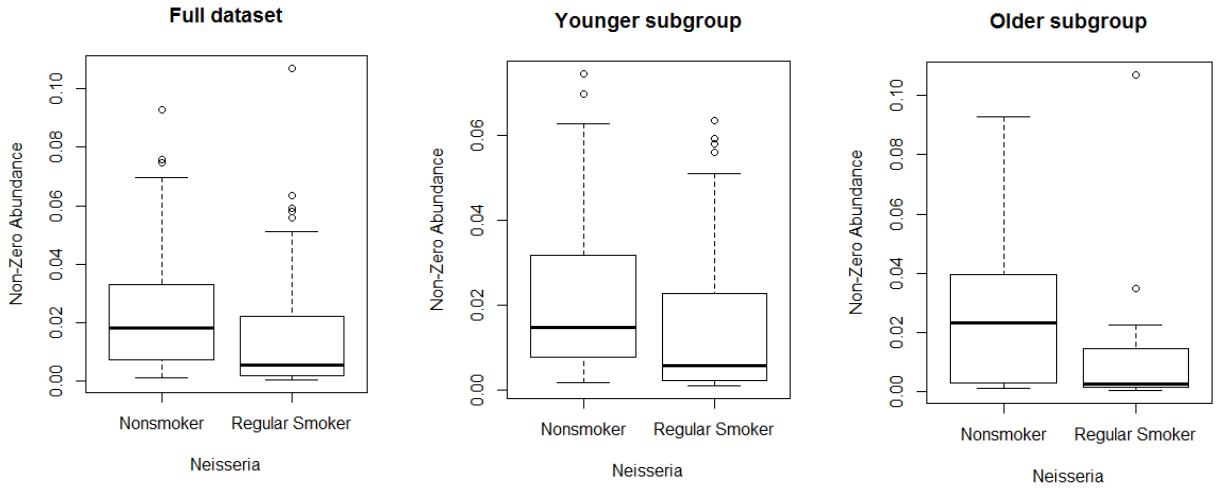
A


Neisseria Presence
Proportion of samples detected
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
n=52
n=136
n=28
n=83
n=24
n=53
Full dataset - difference by smoking
Younger dataset - differences by smoking
Adult dataset - differences by smoking
Nonsmoker
Smoker

B


Filifactor Presence
Proportion of samples detected
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
n=52
n=136
n=28
n=83
n=24
n=53
Full dataset - difference by smoking
Younger dataset - differences by smoking
Adult dataset - differences by smoking
Nonsmoker
Smoker

C


Full dataset
Younger subgroup
Older subgroup
Non-Zero Abundance
Nonsmoker
Regular Smoker
Neisseria

D

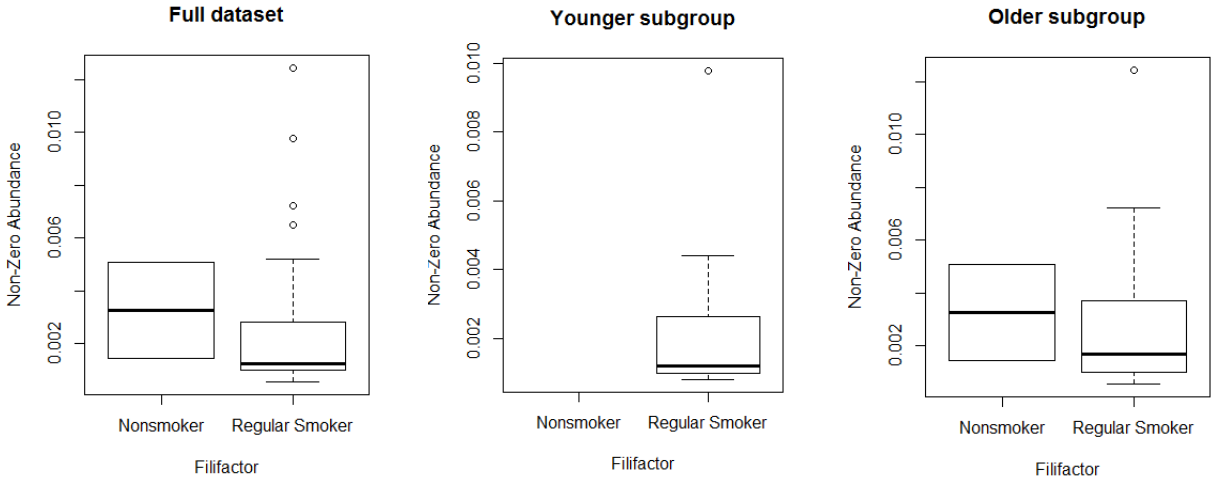

Full dataset
Younger subgroup
Older subgroup
Non-Zero Abundance
Nonsmoker
Regular Smoker
Filifactor

Figure 3.18: Presence and non-zero abundances of the genera *Filifactor* and *Neisseria* change significantly with smoking in all age groups. A, B)Proportion of samples with non-zero counts split in nonsmokers and regular smokers for the full dataset as well as stratifications into age groups for genus A) *Neisseria* and B) *Filifactor*. Sample sizes are indicated above the bars. C, D) Non-zero abundances split in nonsmokers and regular smokers for the full dataset as well as stratifications into age groups for genus C) *Neisseria* and D) *Filifactor*. Sample sizes are identical to A and B.

Figure 3.18 A and C demonstrate a consistent decrease in presence and abundance of *Neisseria* in regular smokers compared to nonsmokers. A reduction of *Neisseria* has been observed in vitro (Bardell 1981; Ertel et al. 1991) and in vivo before (Colman et al. 1976; Mager et al. 2003a; Shchipkova et al. 2010; Kumar et al. 2011). The tobacco susceptibility of *Neisseria* was confirmed by my study.

A dramatic increase in presence with smoking has been observed in the genus *Filifactor* across all age groups as Figure 3.18 B and D shows. A study of smoking cessation found a significant decrease of *Filifactor alocis* levels after quitting (Delima et al. 2010). Furthermore, the organism has been detected predominantly in subgingival plaque of periodontitis patients; however, the proportion of individuals who were smokers was not reported and could be a significant factor (Schlafer et al. 2010).

It is likely that smoking and age affect bacterial strains or taxa with certain chemical sensitivities or biochemical properties. The current level of taxonomic resolution is not sufficient to resolve such differences. Furthermore, even on a species level, members of the same genus behave differently in regard to smoking status. While *Porphyromonas endodontalis* was significant associated with smoking in a study by Delima et al., its close relative *Porphyromonas gingivalis*, did not change in abundance or presence (Delima et al. 2010).

However, it is interesting to think about signature genera, which suggest the involvement of known oral colonizers. For example, the genus *Aggregatibacter* includes the species *Aggregatibacter actinomycetemcomitans*, famously known for its involvement in juvenile chronic periodontitis, decreases in abundance and presence with age as well as smoking.

Paludibacter which increases with age in this study has been found to be associated with periodontitis in subgingival plaque (Bizzarro et al. 2013). The 16S rRNA gene sequence blasted to the HOMD is most similar (99%) to *Bacteroidales[G-2] sp. oral taxon 274* also known as *Oral Clone AU126 (*Human Oral Taxon ID 274). This organism was found to be increased in subgingival plaque in periodontitis and gingivitis and had a higher disease association ratio than the common bacterial species commonly associated with periodontitis including the red complex (Kumar et al. 2003; Li et al. 2006). Periodontitis incidence is positively correlated with age, which may explain this association (U.S. Department of Health and Human Services 2000). The increase of *Treponema* in the full dataset based on smoking behavior is in accordance with other studies, which observed an increase in the periodontal pathogen *Treponema denticola* in the sub- and supragingival biofilm of smokers (Haffajee and Socransky 2001; Kumar et al. 2011). Furthermore, a decrease in *Capnocytophaga gingivalis* has been reported between nonsmokers and smokers with early onset periodontitis, aged 25–38 years (32.3 ± 3.8) previously (Kamma and Nakou 1997). A significant reduction of the genus *Capnocytophaga* has been observed in this full dataset and adult dataset, but not in the younger dataset. The age range in Kamma and Nakou overlaps significantly with this adult subgroup.

More speculative, the abundance of the genus *Haemophilus* decreases with smoking in the full and adult dataset as well as with aging. Sequences in this genus are most similar to *Haemophilus parainfluenzae* based on the comparison with HOMD (Chen et al. 2010). *Haemophilus* is not known as an indicator for health and includes the known pathogen *Haemophilus influenzae (Musher 1996)*. However, it is also present in 100% of the oral cavity of HMP subjects, a presumably healthy cohort (Segata et al. 2012). Others have shown an association of *Haemophilus parainfluenzae* with oral health in children (Shaddox et al. 2012) as well as with periodontitis treatment success (Colombo et al. 2012). *Haemophilus parainfluenzae* can occasionally be found in systemic diseases such as pneumonia or endocarditis (Musher 1996). Smokers are known to be more susceptible to periodontitis and have a poor treatment response

(Palmer et al. 2005), which could explain these correlations. It is therefore possible that *Haemophilus,* specifically *Haemophilus parainfluenzae* is an indicator of oral health, however there are conflicting results in the literature. An in vitro study showed that *Haemophilus influenza* was highly susceptible to tobacco smoke (Ertel et al. 1991), while other studies on contrast have observed an increase in *Haemophilus* colonization in developing plaque in smokers (Kumar et al. 2011).

3.3.9 Effect of alcohol and stimulants

My analyses indicates that the effect of alcohol and stimulant consumption is minor compared to the dominant effects of smoking and age. Exploratory analysis of alcohol and stimulant effects in small subgroups showed inconsistent results. Due to prohibitive low samples numbers and multiple testing, subgroups of other phenotypes, for example young smoking stimulant users, have not been reported. A larger study controlled for age and smoking is necessary to assess these effects.

CHAPTER IV
DISCUSSION

This dissertation analyzed influence of host genetics, age, and environmental variables such as cohabitation and tobacco use, on the oral microbiota.

4.1 General remarks about the studies

Sample description. I successfully analyzed the bacterial communities of a total of 474 salivary samples obtained via 16S rRNA gene sequencing of variable regions 1 and 2 with 454 pyrosequencing in two studies. The samples were derived from the salivary DNA collection of the Institute of Behavioral Genetics at the University of Colorado. I showed that banked salivary samples can be used successfully to analyze the human salivary microbiota, which allows large-scale microbiota studies without the costs associated with sampling and DNA extraction. Furthermore, a vast array of supplemental information about these individuals is available, specifically on stimulant use and experimentation, as well as key psychological measurements. In addition, most subjects were analyzed in their family context and a number of subjects are being followed longitudinally.

Limitations. Because these samples were collected with the goal of human genetic analysis, not microbial analysis, a number of limitations exist. Samples were collected with a variety of methods. To avoid the effect of different collection and extraction methods in the sample, I restricted my study to samples collected with Scope mouthwash. However, a prelimary study did not find identifiable biases between the methods (Appendix B). The DNA of dissolved bacterial and human cells was extracted with commercial reagents (Gentra Puregene, Qiagen). Although no mechanical disruption method was used in the extraction of the DNA from saliva, I readily detected known "hard to lyse" species including *Actinomyces*

and *Streptococcus* in comparable abundances to previously published studies that used zirconia beads to disrupt bacterial cell walls (Keijser et al. 2008).

Because of the study design, information on important modifiers of microbial composition such as oral health and antibiotic use are not available. While oral health potentially has an effect on the salivary communities, no consistent significant difference between saliva from periodontal healthy and diseased individuals have been detected (Mager et al. 2003a) and despite changes in subgingival bacterial composition after periodontal treatment, the composition of saliva remained relatively unchanged within the two year span before and after treatment (Yamanaka et al. 2012). Antibiotic use is estimated with 1-3% of individuals taking antibiotics at any given time, which corresponds to two to eight individuals in each of the two datasets (Goossens et al. 2005).

Choice of beta diversity metric. The unweighted UniFrac distance was selected as the beta diversity metric of choice (Lozupone and Knight 2005; Lozupone et al. 2006). The unweighted UniFrac distances (presence based) was preferred to the weighted UniFrac (presence and abundance based) because it generally performs better in discriminating human body sites based on my observation and others' (Costello et al. 2009). Preliminary analysis of my dataset revealed that due to the dominance of streptococci in most samples, the distance matrix is skewed and does not give good resolution.

4.2 The effect of host genetics and cohabitation on the oral microbiota

Host genetics. Chapter 2 analyzed microbial communities of saliva derived from MZ and DZ twins in a longitudinal design. Twin studies allow one to dissect the influence of environmental and host genetic effects based on MZ-DZ genomic differences. Heritable factors could include innate immune factors, physiological or morphological factors. In addition to direct modulation of the oral environment, it has been shown that factors that are commonly described as environmental are also heritable. This includes dietary preferences (Teucher et al.

2007) and external factors such as childhood environment, social environment and behavior, leisure time activities and life events (Vinkhuyzen et al. 2010). While the identity of specific genetic effects is likely to be too complex to be revealed by genome wide association studies, if these heritable behaviors influence the composition of the microbiota, twin studies could expose the heritability and it is expected that MZ are more similar than DZ twin pairs. Although there was a trend toward a higher similarity among MZ twin pairs in my study, this similarity was not statistically significant. Results obtained by these findings suggests that the overall genetic make-up of the host has little or no apparent role at ages 12–24 in explaining salivary phylogenetic composition measured by unweighted UniFrac distance.

This finding is supported by another twin study of the oral microbiota of twins, which compared bacterial communities in supra- and subgingival plaque via T-RFLP (Papapostolou et al. 2011). The same has been shown for the gut microbiota in various independent twin samples from the US and Malawi (Turnbaugh et al. 2009; Yatsunenko et al. 2012). The failure to show a significant difference between MZ and DZ twins suggests that the overall heritability based on beta diversity measures (unweighted UniFrac) is low in a complex human environment.

The environment has a higher impact on the overall composition of the oral microbiome than host genetics. This finding does not rule out the possibility that individual gene variations may have an effect on the overall composition or specific strains, species, or higher taxonomic groups. For example the presence of a methanogenic archaea, *Methanobrevibacter smithii*, had a higher concordance rate in MZ than in DZ twins (Hansen et al. 2011), which does suggest that the presence of individual organisms have a genetic predisposition. Mouse studies have shown that besides cohort and litter influence, QTLs exist which are significantly associated with the abundance of certain bacterial taxa (Benson et al. 2010). Human environmental influences are more complex than environmental influences in laboratory mice, which could override the heritable effects.

Intra-individual differences and similarities. In addition, even though twins have a more similar microbiota compared to unrelated individuals, there are still differences between them, which suggest postnatal influences on composition. For example, the gut viromes were found to be individual specific, regardless of their genetic relatedness (Reyes et al. 2010). Despite having a shared environment, short-term variation exists. While I and others have shown that self-similarity exists up to 5-year time spans during adolescence and young adulthood (from age 12 to 17 and age 17 to 22), relatively large intra-individual variation has been observed even short term (Turnbaugh et al. 2009; Caporaso et al. 2011; Human Microbiome Project Consortium 2012).

Shared environment. As this study and other studies showed, shared environment, measured by cohabitation, is an important influence on the microbiota of various human habitats (Stahringer et al. 2012; Yatsunenko et al. 2012; Song et al. 2013). I showed that twins share more of the phylogentic diversity measured by unweighted UniFrac of their salivary microbiota than unrelated individuals do. In addition, the similarity decreases once they cease cohabitating at age 22. This suggests a dominant effect of environment. However, the similarity between twins is still significant at this point, which suggests that early life influences or learned habits for example dietary patterns or oral hygiene influence the composition. It is also not known how much of the oral microbiota is shared long-term or on sporadic direct contact such as family gatherings. Turnbaugh et al. (2009) also showed that adult twins as well as the (adult) children and their mothers, many of which did not share the same household, still shared significant proportions of their gut microbiota. On a temporal scale, individuals and their co-twins are not significantly distinguishable during their progression from age 12 to age 17 and progression from age 17 to age 22. However, the same individuals differ significantly from unrelated individuals at ages 12–17, and even more at ages 17–22. As the majority of twins changed their environment (i.e., move to a new home) between the ages of 17 and 22, which may increase differences. These results support the shared environment hypothesis.

Other cohabitation examples. Gut microbiome research reveals more evidence of cohabitation effects. It has been shown that related individuals (parent and child, siblings) as well as unrelated individuals (partners) share a greater proportion of their gut microbiota than unrelated individuals when living in the same household (Yatsunenko et al. 2012). However, cultural differences exist. While children and their mothers in the US and Venezuela (Amerindians) had lower beta diversity values than unrelated children to adults and are therefore more similar to each other, this difference did not exist in Malawi. Group rearing practices might be more common in Malawi than the US or Venezuela. Interestingly, Yatsunenko showed that same gender DZ twins were significantly more similar than mixed gender DZ twins in US teenagers (Yatsunenko et al. 2012). The same trend has been observed in US infant twins, but the difference failed to reach significance. Same gender DZ twins were even more similar than MZ twins. This finding was not discussed in this study. One could speculate that same gender DZ twins are more prone to engage in the same activities together (self-chosen or by parents), but MZ twins chose to be more individualistic because of their identical visual appearance to be perceived as separate beings to their peers. This finding stresses the importance of environmental influence. The finding that cohabitation leads to shared microbial composition extrapolates to other body habitats. A recent study by Song et al. (2013) demonstrated that cohabitating individuals (spouses, parent and child) have a significantly lower beta diversity based on unweighted UniFrac and are therefore more similar than non-cohabitating individuals on various body sites, including stool, two skin locations (forehead, palm), and tongue. It is currently unclear what factors of shared environment or cohousing are important. Examples are exposure to the same microbial sources, diet, or hygiene practices. Evidence of direct salivary contact comes from the study from Song et al. (2013). While spouses and mother and child pairs shared significantly more of their tongue microbiota than random individuals, father and child did not. A speculation is that spouses exchange saliva on a regular bases (kissing), and children might share utensils with their mother more

likely than with their father. Another scenario is cross-contamination during meal preparation, which is traditionally done primarily by females.

4.3 The effect of age on the oral microbiota

This dissertation analyzed the effect of age on the oral microbiota in various ways. The longitudinal design of the twin study allows us to follow the intra-individual changes over two five year spans, from age 12 to 17 and 17 to 22. The drug study with a large range of ages (12 to 65) allows a cross-sectional approach in assessing the influence of age on the oral microbiota.

Self-similarity for up to 5 years. In Chapter 2, I demonstrated that individuals exhibit a self-similarity over two five year spans in the twin dataset based on unweighted UniFrac distances. The long-term stability of the oral microbiome over many years is remarkable. One expects changes in diet, oral hygiene, or romantic partners to occur in these years, and yet stability remains high. These findings are similar to those for the gut, where the microbial community is in flux only in the first few months of life, mainly determined by environmental events (Koenig et al. 2011). Later on, the microbiome stabilizes and becomes less susceptible to disruptions (Spor et al. 2011).

More changes are observed between the ages of 12 to 17 than ages 17 to 22. Changes within an individual during adolescence (age 12 to 17) have been found to be more drastic than changes from adolescent to young adult (age 17 to 22). Hormonal (puberty) and behavioral changes are likely the important factors. Sex hormones have shown to influence the gingiva in male and female (Guncu et al. 2005). It is currently unknown whether the oral microbiome becomes increasingly stable after the teenage and preadolescent years, when hygiene and eating routines develop.

Cross sectional age differences. Very little is known of the microbial dynamics in adults (Cho and Blaser 2012). Differences in children (Kononen 2000; Papaioannou et al. 2009; Crielaard et al. 2011; Ling et al. 2012) and the elderly (Percival et al. 1991; Marsh and Percival

2006; Kraneveld et al. 2012) have been studied, but it is generally assumed that the oral microbiota does not change during adulthood. Data from Chapter 3 suggests that an age gradient exists in cross sectional studies from adolescence to mature adulthood. The salivary microbial communities of this selected dataset with enrichment in antisocial behavior and drug consumption exhibits a strong age gradient across PCs based on a PCoA of unweighted UniFrac distances. The effect is strongest in the subpopulation of under 30 year olds, but exists in individuals over 30 years as well. Individuals up to 30 years are significantly more similar to each other than individuals over 30 years. This result was further supported by the multiple linear regressions on the full dataset as well as the stratified datasets at age 30, which identified age as one of the strongest predictors of several PCs.

Limitations and further evidence. Several cofounding factors could have an influence in this cross sectional study of a selected population including drug experimentation or neglect of oral health. Nevertheless, these factors are unlikely to explain the whole age gradient and further research needs to be conducted to examine this effect. A small but significant increase in beta diversity was observed in the randomized population in the twin study between age 17 and 22 and indicates an increase in diversity with age in unrelated individuals (Figure 2.5). A similar effect has been shown by Papapostolou et al. (2011). In the supragingival plaque bacterial communities a significant trend of decrease in similarity with increasing age between 20 and 40 years was observed (Papapostolou et al. 2011). This and my findings suggest that the shared oral microbiota "core" decreases with age and each individual acquires an increasing number of unique organisms. This hypothesis is supported by a reduction in the shared core with aging. Furthermore, more genera were found to decrease significantly in abundance and presence than increase from the subpopulation under 30 years to the subpopulation over 30 years, which also suggests that each individual loses members of the common core and acquires his own collection of microorganisms. An important cofounding factor could be periodontitis, which is positively correlated with age (U.S. Department of Health and Human Services 2000).

However, due to lack of information on oral health in this dataset I could not address the association with periodontitis.

Further evidence of an age effect was obtained in the twin study. During the adolescent and young adulthood years, a systematic pattern of change in the genera *Veillonella* (decrease with age), *Actinomyces* (increase with age), and *Streptococcus* (increase with age) was observed. Furthermore, unpublished results of the cohort of 8-11 year olds in the twin study were significantly more similar within themselves than later samplings. However, due to questions regarding differences in the sample collection methods, these findings should be regarded with caution.

My results show that it is crucial to age match cases and controls for correct conclusions to be drawn, given changes in the oral microbiome over adolescence and adulthood. This study thus highlights the requirement for a broad sampling of humans of different ages and lifestyles for microbiome studies, especially longitudinal designs to assess the stability of oral microbial communities during adulthood in individuals. The twin cohort with its ongoing follow-up and extensive longitudinal information would be ideally suited for such an experiment.

4.4 The effect of environmental variables on the oral microbiota on the example of drug consumption

Environmental factors play an important role in determining the oral microbiota. I attempted to elucidate the effect of life style choices, namely consuming various drugs. I successfully analyzed the salivary bacterial communities of 210 individuals with different substance use habits with respect to tobacco, but not alcohol and stimulants.

Study limitations. Due to the polyfactorial nature of the study and the lack of control for oral status, this study is not suited to address all proposed effects of various substances. Unexpected strong age effects challenged the assessment of subtle effects such as alcohol consumption and stimulant use. It is currently unknown what drives the age effect.

The cutoffs for different drug usages were highly debated. The definitions for smoking, alcohol and stimulants usage utilized in this dissertation were basic and refer to the most recent usage behavior. Accounting for other similar substances, such as other tobacco products or marijuana smoking and prescription stimulant use, as well as long-term alcohol consumption behavior would be ideal. For example, a large proportion of subjects from this dataset (approximately 50%) used marijuana within the past 180 days at least once. This includes two nonsmokers, who consumed marijuana daily. I decided to use the core definitions, mainly based on recentness, for this exploratory analysis due to restrictions in sample size. Further studies with bigger sample sizes will allow one to control for such cofounding factors.

Tobacco effect. Tobacco smoking has a strong effect on oral health which results in a more severe form of periodontitis (Brook 2011; Kumar 2012), an altered immune state of the host (Brook 2011), and changes in the gene expression of oral bacteria (Bagaitkar et al. 2009; Brook 2011). It has been shown that smoking results in enrichment in periodontal and systemic pathogens (Kamma and Nakou 1997; Shchipkova et al. 2010; Brook 2011; Kumar et al. 2011; Kumar 2012).

Chapter 3 demonstrated that tobacco smoking increases the beta diversity in the salivary bacterial communities independent of age. The PCoA plot based on unweighted UniFrac beta diversity metric shows an increase in beta diversity in smokers compared to nonsmokers and irregular smokers. The two groups do not form two separate clusters, instead they overlap and the nonsmokers and irregular smokers form a subgroup in the observed diversity of smokers. This suggests that healthy oral microbiota are more similar to each other and there is a healthy state represented by the nonsmokers. In contrast, perturbations, for example, through smoking can cause the salivary microbiota to become more diverse. This result was supported by findings of the averaged beta diversity which is increased in smokers compared to nonsmokers, as well as a reduction of the most common genera found in smokers compared to nonsmokers. While my current level of taxonomic resolution is not sufficient to identify pathogenic species,

evidence on genus level suggests that an increase in periodontal pathogens is observed. Furthermore, the known tobacco smoke susceptible genus *Neisseria* was significantly reduced in smokers compared to nonsmokers. *Neisseria* is thought to be an indicator genus of oral health (absence of periodontitis) (Kumar et al. 2005). Nonsmokers as well as the twin population, which are predominantly nonsmokers share a greater proportion of "core" organisms than smokers do. In addition, a larger number of organisms are depleted in smokers, but only a small number of organisms consistently appear.

The multiple linear regression model based on the first three PCs of the PCoA plot showed that smoking was significantly associated with multiple PCs, independently of age. My result that nonsmokers comprise a subgroup of diversity within smokers and exhibit a gradient along one PC has been found previously in bacterial communities in subgingival plaque of smokers and nonsmokers (Bizzarro et al. 2013), but was not interpreted this way in the study by the authors.

Biological model. These findings suggest that each smoker acquires his or her own diverse set of microbes, which differs from nonsmokers and other smokers, while nonsmokers are more similar to each other. It has been suggested before that the microbiota that associate with disease are more complex than healthy microbiota (Jenkinson 2011). This has been found to be true in intestinal diseases, such as inflammatory bowel disease (Morgan et al. 2012), but also in the initial plaque formation of smokers (Kumar et al. 2011). A model has been proposed, where smoking disturbs the formation of a healthy bacterial colonization and potentially opens up the biofilm for pathogenic organisms, which is supported by my and other findings of a reduced core in smokers and a potential increase in oral pathogens (Kumar et al. 2011).

Overlapping effects of age and tobacco. Age as well as smoking causes an increase in beta diversity and a reduction in the shared core. In the presented dataset, it is difficult to separate the effects. However, results from multiple linear regressions as well as stratification into age groups suggest that the effects are independent. It is likely that both age and tobacco

smoking independently cause a loss of equilibrium, which could be associated with periodontal disease or other factors. These results together with the results from the young and predominantly nonsmoking twin dataset indicate that a young and nonsmoking core microbiota with little beta diversity exists and aberrations can be caused by smoking as well as age. This parallels the Anna Karenina principle (Leo Tolstoy), which states that several key aspects or conditions must be fulfilled in order to flourish. For the oral microbiota it can be adapted from "Happy families are all alike; every unhappy family is unhappy in its own way" to "Happy salivary communities are all alike; every unhappy salivary community is unhappy in its own way".

Alcohol and stimulant effect. Even though regular alcohol consumption is known to reduce salivary flow and damages mucosal cells, a direct effect on the oral microbiota has yet to be shown (Riedel et al. 2003). While methamphetamine as a representative for stimulants has a very dramatic effect on oral health, commonly referred to as "Meth mouth" (Hamamoto and Rhodus 2009), my study failed to show a clear effect due to underlying dominant effects of age and smoking. A different analysis approach, which stresses recent as well as chronic exposure and an accounting for age and smoking are necessary to examine the effects of alcohol and stimulants independently.

Resilience and past exposure. This dataset demonstrated a smoking effect independent of age group, where nonsmokers are significantly more similar to each other than smokers. Interestingly, the irregular/former smokers were not significantly different from nonsmokers. This suggests that the recentness of tobacco consumption is crucial and the oral microbiota is resilient to perturbations. A study on smoking cessation has shown that the subgingival biofilm approaches health compatible levels 12 months after quitting (Delima et al. 2010). The high resilience of the oral microbiota might be an additional explanation why a stimulant or alcohol effect, if it exists, is difficult to detect. The definition of stimulants does not include the recentness requirement as the smoking definition does. However, in a preliminary study of

seven recent stimulant users, effects were low. In addition, alcohol consumption of the past hours or even shorter timeframes might differ from longer periods of time.

Other effects. In the multiple linear regression on the first three PCs of the PCoA plot, race, based on a split into Non-Hispanic Whites and others, was marginally significant for PC3. This suggests that genetics or culture including diet might have an effect on the oral microbiota, which is not directly observable with commonly applied metrics. Developments in analysis techniques are necessary to distinguish effects buried under other more dominant effects. For example a study by Nasidze et al. failed to show a cluster by geography, however no linear regression on PC3 was applied (Nasidze et al. 2009a).

Improvements in study design. The previously unreported age effect revealed by this study needs to be investigated further. Until then, oral microbiological studies should be restricted to narrower age ranges to minimize this effect. Studies, which aim to resolve other effects such as the effect of alcohol, need to control for age as well as smoking. Furthermore, the study of subtle effects, such as alcohol, requires adequately statistical power with a higher number of subjects and sequences to establish correlations. My analyses have shown that results basing on different subsampling can vary tremendously (data not shown). A deeper sequencing depth and an increase in sample size will helps to resolve this issue.

CHAPTER V
OUTLOOK AND SUMMARY

5.1 Future of oral microbiology and microbiology in general

16S rRNA gene based studies of human and non-human associated environments opened a whole new universe to explore, just as Hooke predicted in 1665 (Gest 2004, pg. 8). The diversity of microbes is vast compared to the diversity of multicellular organisms visible to the unaided eye (Pace 1997). The ultimate goal of the microbiome research is to understand the intricate interactions between the human host and its microbial inhabitants and apply this knowledge to improve human health (NIH Human Mirobiome Project 2013). With this knowledge, modification of bacterial communities in different human habitats to maintain and restore health will be feasible. In addition, microbes or their metabolic products can serve as diagnostic markers for human or bacterial diseases. While bacteria comprise the majority of microbes in the oral cavity, archaea, for example *Methanobrevibacter oralis* is commonly found in the oral cavity (Matarazzo et al. 2011). Furthermore, fungi including Candida species are frequently found in the oral cavity of healthy individuals (Ghannoum et al. 2010), but the significance of these findings is not yet understood.

5.1.1 Understanding interactions

In order to understand the interactions between host and microbe and among microbes, an understanding of the members is necessary. This goal can be started to be addressed with 16S rRNA gene based studies such as the one described in this dissertation. In addition, understanding what genetic and environmental forces shape the composition will aid in elucidating the interrelationships, which was the goal of this dissertation. There is a correlation

of taxonomy and functional properties, which validates the purpose of 16S rRNA gene based studies (Jansson et al. 2009; Human Microbiome Project Consortium 2012). However, taxonomically relatively unrelated organisms can have converged to evolve into similar functions (Lozupone et al. 2012).

Analysis of co-occurrence and exclusions of microbes in an environment provides evidence about shared environmental preference and symbiotic relationship as well as different environmental preference and competition (Lozupone et al. 2012). In addition, metabolic predictions from organisms might be possible to a certain degree, especially if the genome sequences are available. Current 16S rRNA gene based studies are not yet capable of identifying microorganisms on a species level in a high throughput fashion (Bizzarro et al. 2013). But if current trends in sequencing technology development continue, read length improvements will eventually provide full SSU sequences.

However, 16S rRNA gene based studies are not sufficient to understand the interactions and mechanisms completely. Data collection has outpaced analytical capacities and therefore many organisms remain unnamed and undescribed (Jenkinson 2011). It cannot be the main goal to accumulate sequences without biological meaning (MacLean et al. 2009). Further understanding will be obtained from a combination of whole genome shotgun sequencing (metagenomics), single cell sequencing, transcriptomics, proteomics, metabolomics, culture and animal studies. Science will need to combine and analyze all of them to get a complete picture of the interactions, the interactome, und ultimately functions of different biomes, which will require interdisciplinary collaborations of bioinformaticians, microbiologists, human geneticists, biochemists, and molecular biologists (Jenkinson 2011).

Metagenomics, the whole genome shotgun sequencing of the complete DNA content in a particular environment, provides information on the sum of functional gene groups available and supposedly necessary in the particular environment. This allows us to assess which organisms are present and which biological functions they are able to perform. However, it is

usually not possible to assign the functions to a particular organism (Tringe and Rubin 2005). Furthermore, human associated environments have the disadvantage of being "contaminated" with human DNA, which can be up to 90% in subgingival plaque in the oral cavity (Liu et al. 2012). Current problems of metagenomic studies are high costs of sample preparation and sequencing as well as limitation in gene annotations. It is possible to assess the 16S rRNA gene content from metagenomic studies if it has been sufficiently sequenced and the sequence reads are long enough to provide the necessary taxonomic information. Furthermore, the assembly of whole genomes of abundant organisms from metagenomic studies has been shown in the oral cavity (Liu et al. 2012).

Single cell sequencing. Advances in isolation and sequencing technologies allow single cell sequencing, as successfully demonstrated in a member of the TM7 phylum obtained from the human mouth (Marcy et al. 2007). Single cell sequencing is necessary, as protein content might be only 40% similar in bacteria with 97% identity of 16S rRNA gene (Welch et al. 2002). This effect can be exaggerated for pathogenicity islands which are often transferred via horizontal gene transfer (Nakamura et al. 2004). Horizontal gene transfer is thought to be facilitated in the oral biofilm due to close spatial proximity. In addition, DNA is found in the matrix of the biofilm. It has been proposed that a large degree of differences in the genomes of different *Porphyromonas gingivalis* strains is due to horizontal gene transfer. Oral streptococci are naturally competent and able to take up DNA from their surrounding (Kolenbrander et al. 2010). The actual gene content is crucial because the presence of various virulence factors, including adhesion molecules, quorum sensing, antibiotic resistance, antibacterial molecule production, can make an essential difference in the fitness of an organism.

Metagenomic studies have shown that while the microbial composition of a given habitat differs, the function likely remains the same in a healthy cohort. A study by Turnbaugh et al. has shown that while the bacterial composition was very diverse in a group of obese and lean twins, the functional gene groups obtained by shotgun sequencing were almost identical when

grouped into broad COG categories (Turnbaugh et al. 2009). Similar findings have been obtained for buccal mucosa, supragingival plaque, and tongue dorsum in the HMP cohort (Human Microbiome Project Consortium 2012). Lozupone et al. use a rainforest analogy in a recent review. Global rainforests, which look similar and have the same functions, are not necessarily composed of the same species. Instead, similar individual niches are occupied by habitants with similar functions, which independently evolved (Lozupone et al. 2012).

While metagenomic studies have shown relative consistent gene distribution, transcriptome studies are more likely to reveal differences because transcription can react rapidly to changing environmental conditions and provide a fast changing picture of the gene activation of a habitat (Bagaitkar et al. 2009). Ultimately, proteomics will provide a picture of changing functions in the habitat, as has been shown in saliva (Rudney et al. 2010; Jagtap et al. 2012). Metabolomics provides another important angle on the function of a habitat. NMR-based metabolomics compared the metabolic profiles of men and women, as well as smokers and nonsmokers and found significant differences (Takeda et al. 2009).

Culture studies are necessary to provide a detailed characterization of the biochemistry and potential pathogenicity of a microbe. While a number of human associated bacteria can be cultured with standard methods, many fail to propagate in vitro. Efforts to culture the hard to culture bacteria has been underway with modest success. Strategies include providing the natural environment or coculture of dependent organisms (Stewart 2012). To copy the natural environment of the oral cavity, in vitro systems have been developed in various stages of complexity, many of which rely on human saliva or artificial saliva as nutrient source. Examples are:

- Simple batch culture systems for example the Zürich Biofilm Model (Guggenheim et al. 2004)
- Saliva-conditioned flow cell (Foster and Kolenbrander 2004)

- Artificial mouths (Tang et al. 2003)
- Retrievable enamel chips for in vivo generation of plaque biofilm (Wood et al. 2000)

In order to understand the interactions of complex natural biofilms, Jo Handelsman suggested applying an approach analogous to molecular biology to “knock-out” (remove selectively) or "overexpress" (oversupply) certain members of the biofilm and observe the effect in the animal model (Handelsman 2005). Germ-free animals have been used for multiple studies of the gut microbiota, including cross species gut transplants between mice and zebra fish (Rawls et al. 2006) or as recipients of human gut samples (Ridaura et al. 2013). Germ-free animals are underused in oral microbiology, but their value starts to be recognized (Hajishengallis et al. 2011). Critics of this approach question the applicability of these studies as germfree animals have an underdeveloped immune system (Robinson et al. 2010) and the studies are expensive and time consuming (MacLean et al. 2009).

5.1.2 Diagnostics

The information obtained by the research described above can be used for the development of salivary diagnostics. Saliva is a body fluid that can be obtained in an easy and non-invasive way and it is therefore optimal for screening and consumer diagnostics. Development of “early detection of dozens of diseases from a saliva sample” has been one of the Grand Challenges of the 21st century mentioned by US president Obama in 2009 in the “Strategy for American Innovation” (http://www.whitehouse.gov/administration/eop/nec/StrategyforAmericanInnovation, accessed October 1, 2013). Future research to test host and microbial biomarkers for a variety of disorders including infectious and uncommunicable diseases such as oral cancer is currently conducted (Mager et al. 2005; Fabian et al. 2008; Pink et al. 2009; Ahn et al. 2012a). While the majority of tests are still under development, saliva based HIV screening tests have been FDA approved

(http://www.fda.gov/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/BloodDonorScreening/InfectiousDisease/ucm080466.htm, accessed October 1, 2013).

At present, the analysis of complete oral microbial communities lacks interpretable information, except in the case of overgrowth of one pathogen. Therefore, a culture-independent open approach is not routine in the medical practice, even though it would be feasible from a monetary standpoint (MacLean et al. 2009).

5.1.3 Therapeutics

From eradication to natural habitat management. Thus far, the "war" analogy is predominant in human medicine (Robinson et al. 2010), which states that bacteria are regarded as "enemies" which should be eradicated and antibiotics are prescribed readily and often unnecessarily, up to 50% (Centers for Disease Control and Prevention 2013, pg. 11). At any given point in time, 1-3% of individuals in different European countries take antibiotics (Goossens et al. 2005). In addition, antibiotics are used routinely in animal husbandry to prevent, control, and cure diseases. Subtherapeutic antibiotic treatment in meat production for example in pigs or chickens increases growth rate and a shift in microbial composition in the animal gut (Jukes 1972; Knarreborg et al. 2002; Collier et al. 2003).

Western civilization has already eradicated a large part of the microbial variety as studies from remote village people in the Amazon and elsewhere have demonstrate by an increased alpha diversity compared to Western samples (Contreras et al. 2010; Blaser et al. 2013). It has been suggested to establish a microbe bank, similar to plant seed banks in the case that the existing population will be susceptible to some external factor. Monoculture is fragile as plant production has shown. An infection of the main papaya producing strain with papaya ringspot virus threatened the whole papaya producing industry (Gonsalves 1998). In addition, changing environmental conditions might make it necessary to revert to old strains, to meet the

upcoming challenges. Cho and Blaser have stated (2012), that "[g]iven the ongoing extinction of our ancient commensal organisms, the future of a healthy human microbiome may include the restoration of our ancestral microbial ecology".

Current dental therapy focuses on the removal of the complete biofilm in order to eliminate a small number of pathogens. However, complete eradication might open up niches for opportunistic pathogens such as *Candida* in a non-competitive environment (He and Shi 2009; Boonanantansarn and Gill 2011, pg. 28). While less dramatic, this situation can be compared to *Clostridium difficile* overgrowth after antibiotic treatment in the gut. If a clinical approach is too aggressive and eradicates most commensal organisms, the most resilient members of the biofilm survive, which could be the most pathogenic (Filoche et al. 2010). To clarify, normal oral hygiene, tooth brushing and flossing is necessary and will not eradicate the oral microbiota, but reduce the growing biofilm to minimize obligate pathogenic species. Surviving members on other oral surfaces and in saliva initiate rapid recolonization after oral hygiene (Kolenbrander 2011b, pg. 5).

It has been suggested that instead of a war analogy, one should adopt a management of the microbial communities, more similar to a national park ranger than burning down the forest (Robinson et al. 2010). While this approach is well suited for the gut microbiota, it is currently unclear if this approach is applicable to the oral microbiome. A clinical study has introduced several streptococci strains in a probiotic mouthwash to reduce the presence of *Streptococcus mutants* and *Porphyromonas gingivalis* successfully (Zahradnik et al. 2009). However, before a targeted approach in the oral cavity can be applied, we need a much better understanding of the intricate details of microbial communities.

5.2 Significance of my findings

My research extends the understanding of factors that modify salivary microbiota, which is crucial for further developments in diagnostic and therapy. I have shown that host dependent and independent factors influence bacterial composition. I demonstrated that host genetics has little effect on the salivary microbiota of adolescents and young adults based on twin studies, similar to that demonstrated in studies of the gut by Turnbaugh et al. with comparable methodology (Turnbaugh et al. 2009). I have shown intra- and inter-individual variation, but also self-similarity for up to five years. I provided additional evidence that cohabitation is an important influence on microbial communities, as also proposed by others, which suggests that effective dental treatment should, perhaps, involve the whole family (Asikainen et al. 1997; Yatsunenko et al. 2012; Song et al. 2013). A novel finding was the observation of an increase in beta diversity with increasing age in a population, aged 12 to 65, enriched in smokers and other substance users. I confirmed that smoking has a significant effect on the oral microbiota (Barbour et al. 1997; Haffajee and Socransky 2001; Kumar et al. 2011) and further research should examine the oral microbiology of smokers and nonsmokers independently.

Given the findings of the presented research, it is critical to consider intrinsic (age) and extrinsic (habitation structures, smoking) factors in further research and development.

Bibliography

Aagaard K, Riehle K, Ma J, Segata N, Mistretta TA, Coarfa C, Raza S, Rosenbaum S, Van den Veyver I, Milosavljevic A et al. 2012. A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PLoS One* **7**(6): e36466.

Aas JA, Griffen AL, Dardis SR, Lee AM, Olsen I, Dewhirst FE, Leys EJ, Paster BJ. 2008. Bacteria of dental caries in primary and permanent teeth in children and young adults. *J Clin Microbiol* **46**(4): 1407-1417.

Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**(11): 5721-5732.

Abu Al-Soud W, Radstrom P. 1998. Capacity of nine thermostable DNA polymerases To mediate DNA amplification in the presence of PCR-inhibiting samples. *Appl Environ Microbiol* **64**(10): 3748-3753.

Adhikari BK, J; Malarcher, A; Pechacek, T; Tong, V. 2008. Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. *MMWR Morb Mortal Wkly Rep* **57**(45): 1226-1228.

Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, Bradshaw CJ, Townsend G, Soltysiak A, Alt KW et al. 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat Genet* **45**(4): 450-455, 455e451.

Ahn J, Chen CY, Hayes RB. 2012a. Oral microbiome and oral and gastrointestinal cancer risk. *Cancer Causes Control* **23**(3): 399-404.

Ahn JH, Kim BY, Song J, Weon HY. 2012b. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J Microbiol* **50**(6): 1071-1074.

Al-Hebshi NN, Al-Sharabi AK, Shuga-Aldin HM, Al-Haroni M, Ghandour I. 2010. Effect of khat chewing on periodontal pathogens in subgingival biofilm from chronic periodontitis patients. *J Ethnopharmacol* **132**(3): 564-569.

Al-Hebshi NN, Skaug N. 2005. Khat (Catha edulis)-an updated review. *Addict Biol* **10**(4): 299-307.

Alan G, Sarah JP. 2012. Microbes as forensic indicators. *Trop Biomed* **29**(3): 311-330.

Alberts B, Wilson JH, Hunt T. 2008. *Molecular biology of the cell*. Garland Science, New York.

Andersson AF, Lindberg M, Jakobsson H, Backhed F, Nyren P, Engstrand L. 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS One* **3**(7): e2836.

Arbes SJ, Jr., Agustsdottir H, Slade GD. 2001. Environmental tobacco smoke and periodontal disease in the United States. *Am J Public Health* **91**(2): 253-257.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**(12): 7724-7736.

Asikainen S, Chen C, Alaluusua S, Slots J. 1997. Can one acquire periodontal bacteria and periodontitis from a family member? *J Am Dent Assoc* **128**(9): 1263-1271.

Asikainen S, Karched M. 2008. Molecular Techniques in Oral Microbial Taxonomy, Identification and Typing. In *Molecular Oral Microbiology*, (ed. AH Rogers), pp. 1-27. Caister Academic Press, Norfolk, UK.

Atkinson MA, Chervonsky A. 2012. Does the gut microbiota have a role in type 1 diabetes? Early evidence from humans and animal models of the disease. *Diabetologia* **55**(11): 2868-2877.

Bagaitkar J, Demuth DR, Scott DA. 2008. Tobacco use increases susceptibility to bacterial infection. *Tob Induc Dis* **4**: 12.

Bagaitkar J, Williams LR, Renaud DE, Bemakanakere MR, Martin M, Scott DA, Demuth DR. 2009. Tobacco-induced alterations to Porphyromonas gingivalis-host interactions. *Environ Microbiol* **11**(5): 1242-1253.

Bahekar AA, Singh S, Saha S, Molnar J, Arora R. 2007. The prevalence and incidence of coronary heart disease is significantly increased in periodontitis: a meta-analysis. *Am Heart J* **154**(5): 830-837.

Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**(3): 541-555.

Baldwin BG. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the compositae. *Mol Phylogenet Evol* **1**(1): 3-16.

Barbour SE, Nakashima K, Zhang JB, Tangada S, Hahn CL, Schenkein HA, Tew JG. 1997. Tobacco and smoking: environmental factors that modify the host response (immune system) and have an impact on periodontal health. *Crit Rev Oral Biol Med* **8**(4): 437-460.

Bardell D. 1981. Viability of six species of normal oropharyngeal bacteria after exposure to cigarette smoke in vitro. *Microbios* **32**(127): 7-13.

-. 1983. The roles of the sense of taste and clean teeth in the discovery of bacteria by Antoni van Leeuwenhoek. *Microbiol Rev* **47**(1): 121-126.

Benedetti G, Campus G, Strohmenger L, Lingstrom P. 2013. Tobacco and dental caries: a systematic review. *Acta Odontol Scand* **71**(3-4): 363-371.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D, Hua K et al. 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**(44): 18933-18938.

Bercik P, Collins SM, Verdu EF. 2012. Microbes and the gut-brain axis. *Neurogastroenterol Motil* **24**(5): 405-413.

Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, Nelson KE, Gill SR, Fraser-Liggett CM, Relman DA. 2010. Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J* **4**(8): 962-974.

Bizzarro S, Loos BG, Laine ML, Crielaard W, Zaura E. 2013. Subgingival microbiome in smokers and non-smokers in periodontitis: an exploratory study using traditional targeted techniques and a next-generation sequencing. *J Clin Periodontol* **40**(5): 483-492.

Blaser MJ, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Estrada I, Gao Z, Clemente JC, Costello EK, Knight R. 2013. Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and US residents. *ISME J* **7**(1): 85-95.

Bobetsis YA, Barros SP, Offenbacher S. 2006. Exploring the relationship between periodontal disease and pregnancy complications. *J Am Dent Assoc* **137 Suppl**: 7S-13S.

Boomsma D, Busjahn A, Peltonen L. 2002. Classical twin studies and beyond. *Nat Rev Genet* **3**(11): 872-882.

Boonanantansarn K, Gill SR. 2011. The Oral Microbiome. In *Oral microbial communities : genomic inquiry and interspecies communication*, (ed. PE Kolenbrander, American Society for Microbiology.), pp. 21-34. ASM Press, Washington, DC.

Boraas JC, Messer LB, Till MJ. 1988. A genetic contribution to dental caries, occlusion, and morphology as demonstrated by twins reared apart. *J Dent Res* **67**(9): 1150-1155.

Borneman J, Triplett EW. 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* **63**(7): 2647-2653.

Brook I. 2011. The impact of smoking on oral and nasopharyngeal bacterial flora. *J Dent Res* **90**(6): 704-710.

Burnett GW, Scherp HW. 1957. *Oral microbiology and infectious disease; a textbook for students and practitioners of dentistry*. Williams & Wilkins, Baltimore,.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**(2): 266-267.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI et al. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**(5): 335-336.

Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N et al. 2011. Moving pictures of the human microbiome. *Genome Biol* **12**(5): R50.

Carpenter GH. 2013. The secretion, components, and properties of saliva. *Annu Rev Food Sci Technol* **4**: 267-276.

Centers for Disease Control and Prevention. Smoking-Attributable Mortality, Morbidity, and Economic Costs (SAMMEC), Five-Year Report: Average Annual Smoking-Attributable Mortality (United States, 2000-2004). U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta.

-. 2013. Antibiotic resistance threats in the United States. Vol http://www.cdc.gov/drugresistance/threat-report-2013/. CDC, Atlanta.

Chae JJ, Aksentijevich I, Kastner DL. 2009. Advances in the understanding of familial Mediterranean fever and possibilities for targeted therapy. *Br J Haematol* **146**(5): 467-478.

Chao A. 1984. Nonparametric-Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics* **11**(4): 265-270.

Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)* **2010**: baq013.

Cho AK, Melega WP. 2002. Patterns of methamphetamine abuse and their consequences. *J Addict Dis* **21**(1): 21-34.

Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. *Nat Rev Genet* **13**(4): 260-270.

Clarke JK. 1924. On the bacterial factor in the ae tiology of dental caries. *Brit J Exp Pathol* **5**(3): 141-147.

Clemente JC, Ursell LK, Parfrey LW, Knight R. 2012. The impact of the gut microbiota on human health: an integrative view. *Cell* **148**(6): 1258-1270.

Cohen AL, Shuler C, McAllister S, Fosheim GE, Brown MG, Abercrombie D, Anderson K, McDougal LK, Drenzek C, Arnold K et al. 2007. Methamphetamine use and methicillin-resistant Staphylococcus aureus skin infections. *Emerg Infect Dis* **13**(11): 1707-1713.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**(Database issue): D169-172.

Collier CT, Smiricky-Tjardes MR, Albin DM, Wubben JE, Gabert VM, Deplancke B, Bane D, Anderson DB, Gaskins HR. 2003. Molecular ecological analysis of porcine ileal microbiota responses to antimicrobial growth promoters. *J Anim Sci* **81**(12): 3035-3045.

Colman G, Beighton D, Chalk AJ, Wake S. 1976. Cigarette smoking and the microbial flora of the mouth. *Aust Dent J* **21**(2): 111-118.

Colombo AP, Bennet S, Cotton SL, Goodson JM, Kent R, Haffajee AD, Socransky SS, Hasturk H, Van Dyke TE, Dewhirst FE et al. 2012. Impact of periodontal therapy on the subgingival microbiota of severe periodontitis: comparison between good responders and individuals with refractory periodontitis using the human oral microbe identification microarray. *J Periodontol* **83**(10): 1279-1287.

Colombo AP, Boches SK, Cotton SL, Goodson JM, Kent R, Haffajee AD, Socransky SS, Hasturk H, Van Dyke TE, Dewhirst F et al. 2009. Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J Periodontol* **80**(9): 1421-1432.

Contreras M, Costello EK, Hidalgo G, Magris M, Knight R, Dominguez-Bello MG. 2010. The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology* **156**(Pt 11): 3282-3287.

Cook RT. 1998. Alcohol abuse, alcoholism, and damage to the immune system--a review. *Alcohol Clin Exp Res* **22**(9): 1927-1942.

Corby PM, Bretz WA, Hart TC, Schork NJ, Wessel J, Lyons-Weiler J, Paster BJ. 2007. Heritability of oral microbial species in caries-active and caries-free twins. *Twin Res Hum Genet* **10**(6): 821-828.

Corby PM, Lyons-Weiler J, Bretz WA, Hart TC, Aas JA, Boumenna T, Goss J, Corby AL, Junior HM, Weyant RJ et al. 2005. Microbial risk indicators of early childhood caries. *J Clin Microbiol* **43**(11): 5753-5759.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. *Science* **326**(5960): 1694-1697.

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. 2012. The application of ecological theory toward an understanding of the human microbiome. *Science* **336**(6086): 1255-1262.

Cottler LB. 2000. The reliability of the CIDI-SAM: A comprehensive substance abuse interview. Department of Psychiatry, Washington, University School of Medicine, St. Louis, MO

Cottler LB, Robins LN, Helzer JE. 1989. The reliability of the CIDI-SAM: a comprehensive substance abuse interview. *Br J Addict* **84**(7): 801-814.

Coventry J, Griffiths G, Scully C, Tonetti M. 2000. ABC of oral health: periodontal disease. *BMJ* **321**(7252): 36-39.

Crielaard W, Zaura E, Schuller AA, Huse SM, Montijn RC, Keijser BJ. 2011. Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Med Genomics* **4**: 22.

Cryan JF, Dinan TG. 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci* **13**(10): 701-712.

Curtis MA, Zenobia C, Darveau RP. 2011. The relationship of the oral microbiotia to periodontal health and disease. *Cell Host Microbe* **10**(4): 302-306.

Daglia M, Papetti A, Grisoli P, Aceti C, Dacarro C, Gazzani G. 2007. Antibacterial activity of red and white wine against oral streptococci. *J Agric Food Chem* **55**(13): 5038-5042.

Daglia M, Papetti A, Mascherpa D, Grisoli P, Giusto G, Lingstrom P, Pratten J, Signoretto C, Spratt DA, Wilson M et al. 2011. Plant and fungal food components with potential activity on the development of microbial oral diseases. *J Biomed Biotechnol* **2011**: 274578.

Darveau RP. 2010. Periodontitis: a polymicrobial disruption of host homeostasis. *Nat Rev Microbiol* **8**(7): 481-490.

Delima SL, McBride RK, Preshaw PM, Heasman PA, Kumar PS. 2010. Response of subgingival bacteria to smoking cessation. *J Clin Microbiol* **48**(7): 2344-2349.

DeLong EF, Wickham GS, Pace NR. 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243**(4896): 1360-1363.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**(7): 5069-5072.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. 2010. The human oral microbiome. *J Bacteriol* **192**(19): 5002-5017.

Diaz PI, Chalmers NI, Rickard AH, Kong C, Milburn CL, Palmer RJ, Jr., Kolenbrander PE. 2006. Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Appl Environ Microbiol* **72**(4): 2837-2848.

Dobell C, Leeuwenhoek Av. 1932. *Antony van Leeuwenhoek and his "Little animals" : being some account of the father of protozoology and bacteriology and his multifarious discoveries in these disciplines*. Harcourt, Brace and Company, New York.

Domejean S, Zhan L, DenBesten PK, Stamper J, Boyce WT, Featherstone JD. 2010. Horizontal transmission of mutans streptococci in children. *J Dent Res* **89**(1): 51-55.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* **107**(26): 11971-11975.

Dong YJ, Peng TK, Yin SJ. 1996. Expression and activities of class IV alcohol dehydrogenase and class III aldehyde dehydrogenase in human mouth. *Alcohol* **13**(3): 257-262.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**(16): 2194-2200.

Ertel A, Eng R, Smith SM. 1991. The differential effect of cigarette smoke on the growth of bacteria found in humans. *Chest* **100**(3): 628-630.

Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB. 2012. Microbiology. Animal behavior and the microbiome. *Science* **338**(6104): 198-199.

Fabian TK, Fejerdy P, Csermely P. 2008. Salivary Genomics, Transcriptomics and Proteomics: The Emerging Concept of the Oral Ecosystem and their Use in the Early Diagnosis of Cancer and other Diseases. *Curr Genomics* **9**(1): 11-21.

Faith DP. 1992. Conservation Evaluation and Phylogenetic Diversity. *Biol Conserv* **61**(1): 1-10.

Fardini Y, Chung P, Dumm R, Joshi N, Han YW. 2010. Transmission of diverse oral bacteria to murine placenta: evidence for the oral microbiome as a potential source of intrauterine infection. *Infect Immun* **78**(4): 1789-1796.

Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. 2009. Opportunistic pathogens enriched in showerhead biofilms. *Proc Natl Acad Sci U S A* **106**(38): 16393-16399.

Ferrazzano GF, Amato I, Ingenito A, De Natale A, Pollio A. 2009. Anti-cariogenic effects of polyphenols from plant stimulant beverages (cocoa, coffee, tea). *Fitoterapia* **80**(5): 255-262.

Fierer N, Hamady M, Lauber CL, Knight R. 2008. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci U S A* **105**(46): 17994-17999.

Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. 2010. Forensic identification using skin bacterial communities. *Proc Natl Acad Sci U S A* **107**(14): 6477-6481.

Filoche S, Wong L, Sissons CH. 2010. Oral biofilms: emerging concepts in microbial ecology. *J Dent Res* **89**(1): 8-18.

Fitzgerald RJ, Keyes PH. 1960. Demonstration of the etiologic role of streptococci in experimental caries in the hamster. *J Am Dent Assoc* **61**: 9-19.

Fitzpatrick SG, Katz J. 2010. The association between periodontal disease and cancer: a review of the literature. *J Dent* **38**(2): 83-95.

Foster JS, Kolenbrander PE. 2004. Development of a multispecies oral bacterial community in a saliva-conditioned flow cell. *Appl Environ Microbiol* **70**(7): 4340-4348.

Frank DN, Feazel LM, Bessesen MT, Price CS, Janoff EN, Pace NR. 2010. The human nasal microbiota and Staphylococcus aureus carriage. *PLoS One* **5**(5): e10598.

Freye E, Levy JV. 2009. *Pharmacology and abuse of cocaine, amphetamines, ecstasy and related designer drugs : a comprehensive review on their mode of action, treatment of abuse and intoxication*. Springer, Dordrecht.

Fullmer SC, Preshaw PM, Heasman PA, Kumar PS. 2009. Smoking cessation alters subgingival microbial recolonization. *J Dent Res* **88**(6): 524-528.

Gazzani G, Daglia M, Papetti A. 2012. Food components with anticaries activity. *Curr Opin Biotechnol* **23**(2): 153-159.

Gest H. 2004. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec R Soc Lond* **58**(2): 187-201.

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL et al. 2005. Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**(9): 733-739.

Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog* **6**(1): e1000713.

Gibbons A. 2012. Evolutionary biology. An evolutionary theory of dentistry. *Science* **336**(6084): 973-975.

Gift HC, Redford M. 1992. Oral health and the quality of life. *Clin Geriatr Med* **8**(3): 673-683.

Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, Casella G, Drew JC, Ilonen J, Knip M, Hyoty H et al. 2011. Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* **5**(1): 82-91.

Glenn TC. 2011. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* **11**(5): 759-769.

Gonsalves D. 1998. Control of papaya ringspot virus in papaya: a case study. *Annu Rev Phytopathol* **36**: 415-437.

Goodchild JH, Donaldson M. 2007. Methamphetamine abuse and dentistry: a review of the literature and presentation of a clinical case. *Quintessence Int* **38**(7): 583-590.

Goodson JM, Groppo D, Halem S, Carpino E. 2009. Is obesity an oral bacterial disease? *J Dent Res* **88**(6): 519-523.

Goossens H, Ferech M, Vander Stichele R, Elseviers M. 2005. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* **365**(9459): 579-587.

Gray H, Lewis WH. 1918. *Anatomy of the human body*. Lea & Febiger, Philadelphia and New York,.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW, Murray PR, Green ED et al. 2009. Topographical and temporal diversity of the human skin microbiome. *Science* **324**(5931): 1190-1192.

Griffen AL, Beall CJ, Firestone ND, Gross EL, Difranco JM, Hardman JH, Vriesendorp B, Faust RA, Janies DA, Leys EJ. 2011. CORE: a phylogenetically-curated 16S rDNA database of the core oral microbiome. *PLoS One* **6**(4): e19051.

Grobler SR, Chikte U, Westraat J. 2011. The pH Levels of Different Methamphetamine Drug Samples on the Street Market in Cape Town. *ISRN Dent* **2011**: 974768.

Gross EL, Leys EJ, Gasparovich SR, Firestone ND, Schwartzbaum JA, Janies DA, Asnani K, Griffen AL. 2010. Bacterial 16S sequence analysis of severe caries in young permanent teeth. *J Clin Microbiol* **48**(11): 4121-4128.

Guggenheim B, Guggenheim M, Gmur R, Giertsen E, Thurnheer T. 2004. Application of the Zurich biofilm model to problems of cariology. *Caries Res* **38**(3): 212-222.

Guncu GN, Tozum TF, Caglayan F. 2005. Effects of endogenous sex hormones on the periodontium--review of literature. *Aust Dent J* **50**(3): 138-145.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E et al. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**(3): 494-504.

Haffajee AD, Socransky SS. 2001. Relationship of cigarette smoking to the subgingival microbiota. *J Clin Periodontol* **28**(5): 377-388.

-. 2009. Relation of body mass index, periodontitis and Tannerella forsythia. *J Clin Periodontol* **36**(2): 89-99.

Haffajee AD, Teles RP, Patel MR, Song X, Yaskell T, Socransky SS. 2009. Factors affecting human supragingival biofilm composition. II. Tooth position. *J Periodontal Res* **44**(4): 520-528.

Hajishengallis G. 2010. Too old to fight? Aging and its toll on innate immunity. *Mol Oral Microbiol* **25**(1): 25-37.

Hajishengallis G, Liang S, Payne MA, Hashim A, Jotwani R, Eskan MA, McIntosh ML, Alsam A, Kirkwood KL, Lambris JD et al. 2011. Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe* **10**(5): 497-506.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**(3): 235-237.

Hamamoto DT, Rhodus NL. 2009. Methamphetamine abuse and dentistry. *Oral Dis* **15**(1): 27-37.

Handelsman J, Robinson, C.J., and Raffa, K. . 2005. Microbial communities in lepidopteran guts: from models to metagenomic. In *The influence of cooperative bacteria on animal host biology*, (ed. MJ McFall-Ngai, B Henderson, EG Ruby), pp. 143-168. Cambridge University Press, Cambridge ; New York, NY.

Hansen EE, Lozupone CA, Rey FE, Wu M, Guruge JL, Narra A, Goodfellow J, Zaneveld JR, McDonald DT, Goodrich JA et al. 2011. Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4599-4606.

He XS, Shi WY. 2009. Oral microbiology: past, present and future. *Int J Oral Sci* **1**(2): 47-58.

Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**(7290): 908-912.

Holgerson PL, Vestman NR, Claesson R, Ohman C, Domellof M, Tanner AC, Hernell O, Johansson I. 2013. Oral microbial profile discriminates breast-fed from formula-fed infants. *J Pediatr Gastroenterol Nutr* **56**(2): 127-136.

Homann N, Tillonen J, Meurman JH, Rintamaki H, Lindqvist C, Rautio M, Jousimies-Somer H, Salaspuro M. 2000. Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. *Carcinogenesis* **21**(4): 663-668.

Howell A, Jr., Jordan HV, Georg LK, Pine L. 1965. Odontomyces viscosus, gen. nov., spec. nov., a filamentous microorganism isolated from periodontal plaque in hamsters. *Sabouraudia* **4**(2): 65-68.

Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* **486**(7402): 207-214.

Hunt KM, Foster JA, Forney LJ, Schutte UM, Beck DL, Abdo Z, Fox LK, Williams JE, McGuire MK, McGuire MA. 2011. Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One* **6**(6): e21313.

Huse SM, Welch DM, Morrison HG, Sogin ML. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**(7): 1889-1898.

International Center for Alcohol Policies. 2012. International Drinking Guidelines. Vol 2013. http://www.icap.org/PolicyIssues/DrinkingGuidelines/.

Jagtap P, McGowan T, Bandhakavi S, Tu ZJ, Seymour S, Griffin TJ, Rudney JD. 2012. Deep metaproteomic analysis of human salivary supernatant. *Proteomics* **12**(7): 992-1001.

Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, Tysk C, Schmitt-Kopplin P. 2009. Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS One* **4**(7): e6386.

Jansson L, Lavstedt S, Frithiof L. 2002. Relationship between oral health and mortality rate. *J Clin Periodontol* **29**(11): 1029-1034.

Jenkinson HF. 2011. Beyond the oral microbiome. *Environ Microbiol* **13**(12): 3077-3087.

Jenkinson HF, Lamont RJ. 2005. Oral microbial communities in sickness and in health. *Trends Microbiol* **13**(12): 589-595.

Jorde LB, Wooding SP. 2004. Genetic variation, classification and 'race'. *Nat Genet* **36**(11 Suppl): S28-33.

Jukes TH. 1972. Antibiotics in Animal Feeds and Animal Production. *Bioscience* **22**(9): 526-&.

Kamens HM, Corley RP, McQueen MB, Stallings MC, Hopfer CJ, Crowley TJ, Brown SA, Hewitt JK, Ehringer MA. 2013. Nominal association with CHRNA4 variants and nicotine dependence. *Genes Brain Behav* **12**(3): 297-304.

Kamma JJ, Nakou M. 1997. Subgingival microflora in smokers with early onset periodontitis. *Anaerobe* **3**(2-3): 153-157.

Kanasi E, Dewhirst FE, Chalmers NI, Kent R, Jr., Moore A, Hughes CV, Pradhan N, Loo CY, Tanner AC. 2010. Clonal analysis of the microbiota of severe early childhood caries. *Caries Res* **44**(5): 485-497.

Karimbux NY, Saraiya VM, Elangovan S, Allareddy V, Kinnunen T, Kornman KS, Duff GW. 2012. Interleukin-1 gene polymorphisms and chronic periodontitis in adult whites: a systematic review and meta-analysis. *J Periodontol* **83**(11): 1407-1419.

Kato I, Nomura AM. 1994. Alcohol in the aetiology of upper aerodigestive tract cancer. *Eur J Cancer B Oral Oncol* **30B**(2): 75-81.

Kauppila JH, Mattila AE, Karttunen TJ, Salo T. 2013. Toll-like receptor 5 and the emerging role of bacteria in carcinogenesis. *Oncoimmunology* **2**(4): e23620.

Kawamura Y, Hou XG, Sultana F, Miura H, Ezaki T. 1995. Determination of 16S rRNA sequences of Streptococcus mitis and Streptococcus gordonii and phylogenetic relationships among members of the genus Streptococcus. *Int J Syst Bacteriol* **45**(2): 406-408.

Keijser BJ, Zaura E, Huse SM, van der Vossen JM, Schuren FH, Montijn RC, ten Cate JM, Crielaard W. 2008. Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* **87**(11): 1016-1020.

Keyes PH, Jordan HV. 1964. Periodontal Lesions in the Syrian Hamster. Iii. Findings Related to an Infectious and Transmissible Component. *Arch Oral Biol* **9**: 377-400.

Khachatryan ZA, Ktsoyan ZA, Manukyan GP, Kelly D, Ghazaryan KA, Aminov RI. 2008. Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One* **3**(8): e3064.

Kinane DF, Chestnutt IG. 2000. Smoking and periodontal disease. *Crit Rev Oral Biol Med* **11**(3): 356-365.

Klappenbach JA, Saxman PR, Cole JR, Schmidt TM. 2001. rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res* **29**(1): 181-184.

Knarreborg A, Simon MA, Engberg RM, Jensen BB, Tannock GW. 2002. Effects of dietary fat source and subtherapeutic levels of antibiotic on the bacterial community in the ileum of broiler chickens at various ages. *Appl Environ Microbiol* **68**(12): 5918-5924.

Kochanek KD, Xu J, Murphy SL, Minino AM, Kung HC. 2011. Deaths: final data for 2009. *Natl Vital Stat Rep* **60**(3): 1-117.

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2011. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4578-4585.

Kolenbrander PE. 2011a. Multispecies communities: interspecies interactions influence growth on saliva as sole nutritional source. *Int J Oral Sci* **3**(2): 49-54.

Kolenbrander PE, Palmer RJ, Jr., Periasamy S, Jakubovics NS. 2010. Oral multispecies biofilm development and the key role of cell-cell distance. *Nat Rev Microbiol* **8**(7): 471-480.

Kolenbrander PE, Periasamy S. . 2011b. Interspecies Interactions Promote Multispecies Communities. In *Oral microbial communities : genomic inquiry and interspecies communication*, (ed. PE Kolenbrander), pp. 3-19. ASM Press, Washington, DC.

Kononen E. 2000. Development of oral bacterial flora in young children. *Ann Med* **32**(2): 107-112.

Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE et al. 2011. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4592-4598.

Kraneveld EA, Buijs MJ, Bonder MJ, Visser M, Keijser BJ, Crielaard W, Zaura E. 2012. The relation between oral Candida load and bacterial microbiome profiles in Dutch older adults. *PLoS One* **7**(8): e42770.

Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, Mei Z, Curtin LR, Roche AF, Johnson CL. 2000. CDC growth charts: United States. *Adv Data*(314): 1-27.

Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE et al. 2010. Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol* **11**(5): 210.

Kumar PS. 2012. Smoking and the subgingival ecosystem: a pathogen-enriched community. *Future Microbiol* **7**(8): 917-919.

Kumar PS, Griffen AL, Barton JA, Paster BJ, Moeschberger ML, Leys EJ. 2003. New bacterial species associated with chronic periodontitis. *J Dent Res* **82**(5): 338-344.

Kumar PS, Griffen AL, Moeschberger ML, Leys EJ. 2005. Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol* **43**(8): 3944-3955.

Kumar PS, Matthews CR, Joshi V, de Jager M, Aspiras M. 2011. Tobacco smoking affects bacterial acquisition and colonization in oral biofilms. *Infect Immun* **79**(11): 4730-4738.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**(1): 118-123.

Lamont RJ, Jenkinson HF. 2010. *Oral microbiology at a glance*. Wiley-Blackwell, Chichester, West Sussex, U.K. ; Ames, Iowa.

Lauber CL, Hamady M, Knight R, Fierer N. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75**(15): 5111-5120.

Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N. 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol Lett* **307**(1): 80-86.

Lazarevic V, Whiteson K, Hernandez D, Francois P, Schrenzel J. 2010. Study of inter- and intra-individual variations in the salivary microbiota. *BMC Genomics* **11**: 523.

Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Osteras M, Schrenzel J, Francois P. 2009. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* **79**(3): 266-271.

Lemos LN, Fulthorpe RR, Triplett EW, Roesch LF. 2011. Rethinking microbial diversity analysis in the high throughput sequencing era. *J Microbiol Methods* **86**(1): 42-51.

Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. 2005. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* **102**(31): 11070-11075.

Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, Maresca JA, Bryant DA, Sogin ML, Pace NR. 2006a. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72**(5): 3685-3695.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006b. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**(7122): 1022-1023.

Li CL, Liang JP, Jiang YT. 2006. Association of uncultivated oral phylotypes AU126 and X112 with periodontitis. *Oral Dis* **12**(4): 371-374.

Li Y, Ismail AI, Ge Y, Tellez M, Sohn W. 2007. Similarity of bacterial populations in saliva from African-American mother-child dyads. *J Clin Microbiol* **45**(9): 3082-3085.

Li Y, Ku CY, Xu J, Saxena D, Caufield PW. 2005. Survey of oral microbial diversity using PCR-based denaturing gradient gel electrophoresis. *J Dent Res* **84**(6): 559-564.

Lie MA, van der Weijden GA, Timmerman MF, Loos BG, van Steenbergen TJ, van der Velden U. 1998. Oral microbiota in smokers and non-smokers in natural and experimentally-induced gingivitis. *J Clin Periodontol* **25**(8): 677-686.

Lif Holgerson P, Harnevik L, Hernell O, Tanner AC, Johansson I. 2011. Mode of birth delivery affects oral microbiota in infants. *J Dent Res* **90**(10): 1183-1188.

Lim SS Vos T Flaxman AD Danaei G Shibuya K Adair-Rohani H Amann M Anderson HR Andrews KG Aryee M et al. 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**(9859): 2224-2260.

Linden GJ, Lyons A, Scannapieco FA. 2013. Periodontal systemic associations: review of the evidence. *J Periodontol* **84**(4 Suppl): S8-S19.

Ling Z, Liu X, Wang Y, Li L, Xiang C. 2012. Pyrosequencing Analysis of the Salivary Microbiota of Healthy Chinese Children and Adults. *Microb Ecol*.

Liska K. 2004. *Drugs and the human body : with implications for society*. Prentice Hall, Upper Saddle River, N.J.

Listgarten MA. 1976. Structure of the microbial flora associated with periodontal health and disease in man. A light and electron microscopic study. *J Periodontol* **47**(1): 1-18.

Liu B, Faller LL, Klitgord N, Mazumdar V, Ghodsi M, Sommer DD, Gibbons TR, Treangen TJ, Chang YC, Li S et al. 2012. Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS One* **7**(6): e37919.

Liu WT, Marsh TL, Cheng H, Forney LJ. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**(11): 4516-4522.

Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* **35**(18): e120.

Loesche W. 2007. Dental caries and periodontitis: contrasting two infections that have medical implications. *Infect Dis Clin North Am* **21**(2): 471-502, vii.

Logares R, Haverkamp TH, Kumar S, Lanzen A, Nederbragt AJ, Quince C, Kauserud H. 2012. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *J Microbiol Methods* **91**(1): 106-113.

Lozupone C, Hamady M, Knight R. 2006. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.

Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**(12): 8228-8235.

Lozupone CA, Knight R. 2008. Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* **32**(4): 557-578.

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**(7415): 220-230.

Lukacs JR, Largaespada LL. 2006. Explaining sex differences in dental caries prevalence: saliva, hormones, and "life-history" etiologies. *Am J Hum Biol* **18**(4): 540-555.

MacLean D, Jones JD, Studholme DJ. 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7**(4): 287-296.

Madeira MF, Queiroz-Junior CM, Cisalpino D, Werneck SM, Kikuchi H, Fujise O, Ryffel B, Silva TA, Teixeira MM, Souza DG. 2013. MyD88 is essential for alveolar bone loss induced by Aggregatibacter actinomycetemcomitans lipopolysaccharide in mice. *Mol Oral Microbiol*.

Maes HH, Neale MC, Eaves LJ. 1997. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**(4): 325-351.

Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. 2005. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J Transl Med* **3**: 27.

Mager DL, Haffajee AD, Socransky SS. 2003a. Effects of periodontitis and smoking on the microbiota of oral mucous membranes and saliva in systemically healthy subjects. *J Clin Periodontol* **30**(12): 1031-1037.

Mager DL, Ximenez-Fyvie LA, Haffajee AD, Socransky SS. 2003b. Distribution of selected bacterial species on intraoral surfaces. *J Clin Periodontol* **30**(7): 644-654.

Magurran AE. 1988. *Ecological diversity and its measurement*. Princeton University Press, Princeton, N.J.

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**(6): 3140-3145.

Maier H, Zoller J, Herrmann A, Kreiss M, Heller WD. 1993. Dental status and oral hygiene in patients with head and neck cancer. *Otolaryngol Head Neck Surg* **108**(6): 655-661.

Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA et al. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci U S A* **104**(29): 11889-11894.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**(7057): 376-380.

Marsh PD, Percival RS. 2006. The oral microflora--friend or foe? Can we decide? *Int Dent J* **56**(4 Suppl 1): 233-239.

Matarazzo F, Ribeiro AC, Feres M, Faveri M, Mayer MP. 2011. Diversity and quantitative analysis of Archaea in aggressive periodontitis and periodontally healthy subjects. *J Clin Periodontol* **38**(7): 621-627.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**(3): 610-618.

Metzker ML. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* **11**(1): 31-46.

Meurman JH, Sanz M, Janket SJ. 2004. Oral health, atherosclerosis, and cardiovascular disease. *Crit Rev Oral Biol Med* **15**(6): 403-413.

Michalowicz BS, Diehl SR, Gunsolley JC, Sparks BS, Brooks CN, Koertge TE, Califano JV, Burmeister JA, Schenkein HA. 2000. Evidence of a substantial genetic basis for risk of adult periodontitis. *J Periodontol* **71**(11): 1699-1707.

Min KR, Rickard AH. 2009. Coaggregation by the freshwater bacterium Sphingomonas natatoria alters dual-species biofilm formation. *Appl Environ Microbiol* **75**(12): 3987-3997.

Mitchell DF, Johnson M. 1956. The nature of the gingival plaque in the hamster: production, prevention, and removal. *J Dent Res* **35**(4): 651-655.

Moore WE, Burmeister JA, Brooks CN, Ranney RR, Hinkelmann KH, Schieken RM, Moore LV. 1993. Investigation of the influences of puberty, genetics, and environment on the composition of subgingival periodontal floras. *Infect Immun* **61**(7): 2891-2898.

Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB et al. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* **13**(9): R79.

Musher DM. 1996. Haemophilus Species. In *Medical Microbiology*, (ed. S Baron). University of Texas Medical Branch at Galveston, Galveston, TX.

Muto M, Hitomi Y, Ohtsu A, Shimada H, Kashiwase Y, Sasaki H, Yoshida S, Esumi H. 2000. Acetaldehyde production by non-pathogenic Neisseria in human oral microflora: implications for carcinogenesis in upper aerodigestive tract. *Int J Cancer* **88**(3): 342-350.

Muyzer G, de Waal EC, Uitterlinden AG. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**(3): 695-700.

Nakagawa S. 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behav Ecol* **15**(6): 1044-1045.

Nakamura Y, Itoh T, Matsuda H, Gojobori T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**(7): 760-766.

Nakano K, Nemoto H, Nomura R, Inaba H, Yoshioka H, Taniguchi K, Amano A, Ooshima T. 2009. Detection of oral bacteria in cardiovascular specimens. *Oral Microbiol Immunol* **24**(1): 64-68.

Nasidze I, Li J, Quinque D, Tang K, Stoneking M. 2009a. Global diversity in the human salivary microbiome. *Genome Res* **19**(4): 636-643.

Nasidze I, Quinque D, Li J, Li M, Tang K, Stoneking M. 2009b. Comparative analysis of human saliva microbiome diversity by barcoded pyrosequencing and cloning approaches. *Anal Biochem* **391**(1): 64-68.

NIH Human Mirobiome Project. 2013. NIH Human Mirobiome Project. http://commonfund.nih.gov/hmp/.

Nishida N, Yamamoto Y, Tanaka M, Kataoka K, Kuboniwa M, Nakayama K, Morimoto K, Shizukuishi S. 2008. Association between involuntary smoking and salivary markers related to periodontitis: a 2-year longitudinal study. *J Periodontol* **79**(12): 2233-2240.

Northrup E. 1957. *Science looks at smoking; a new inquiry into the effects of smoking on your health*. Coward-McCann, New York,.

Ogunseitan O. 2005. *Microbial diversity : form and function in prokaryotes*. Blackwell Pub., Malden, MA.

Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**(5313): 734-740.

-. 2009. Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* **73**(4): 565-576.

Palmer RJ, Jr., Kazmerzak K, Hansen MC, Kolenbrander PE. 2001. Mutualism versus independence: strategies of mixed-species oral biofilms in vitro using saliva as the sole nutrient source. *Infect Immun* **69**(9): 5794-5804.

Palmer RM, Wilson RF, Hasan AS, Scott DA. 2005. Mechanisms of action of environmental factors--tobacco smoking. *J Clin Periodontol* **32 Suppl 6**: 180-195.

Papaioannou W, Gizani S, Haffajee AD, Quirynen M, Mamai-Homata E, Papagiannoulis L. 2009. The microbiota on different oral surfaces in healthy children. *Oral Microbiol Immunol* **24**(3): 183-189.

Papapostolou A, Kroffke B, Tatakis DN, Nagaraja HN, Kumar PS. 2011. Contribution of host genotype to the composition of health-associated supragingival and subgingival microbiomes. *J Clin Periodontol* **38**(6): 517-524.

Parahitiyawa NB, Jin LJ, Leung WK, Yam WC, Samaranayake LP. 2009. Microbiology of odontogenic bacteremia: beyond endocarditis. *Clin Microbiol Rev* **22**(1): 46-64, Table of Contents.

Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, Sahasrabudhe A, Dewhirst FE. 2001. Bacterial diversity in human subgingival plaque. *J Bacteriol* **183**(12): 3770-3783.

Paster BJ, Dewhirst FE. 2009. Molecular microbial diagnosis. *Periodontol 2000* **51**: 38-44.

Paster BJ, Dewhirst FE, Coleman BC, Lau CN, Ericson RL. 1998. Phylogenetic analysis of cultivable oral treponemes from the Smibert collection. *Int J Syst Bacteriol* **48 Pt 3**: 713-722.

Paster BJ, Olsen I, Aas JA, Dewhirst FE. 2006. The breadth of bacterial diversity in the human periodontal pocket and other oral sites. *Periodontol 2000* **42**: 80-87.

Pearson TA. 1996. Alcohol and heart disease. *Circulation* **94**(11): 3023-3025.

Percival RS, Challacombe SJ, Marsh PD. 1991. Age-related microbiological changes in the salivary and plaque microflora of healthy adults. *J Med Microbiol* **35**(1): 5-11.

Periasamy S, Kolenbrander PE. 2009. Mutualistic biofilm communities develop with Porphyromonas gingivalis and initial, early, and late colonizers of enamel. *J Bacteriol* **191**(22): 6804-6811.

Petersen PE. 2003. The World Oral Health Report 2003: continuous improvement of oral health in the 21st century--the approach of the WHO Global Oral Health Programme. *Community Dent Oral Epidemiol* **31 Suppl 1**: 3-23.

Peterson DA, Frank DN, Pace NR, Gordon JI. 2008. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe* **3**(6): 417-427.

Petersson J, Carlstrom M, Schreiber O, Phillipson M, Christoffersson G, Jagare A, Roos S, Jansson EA, Persson AE, Lundberg JO et al. 2009. Gastroprotective and blood pressure lowering effects of dietary nitrate are abolished by an antiseptic mouthwash. *Free Radic Biol Med* **46**(8): 1068-1075.

Pink R, Simek J, Vondrakova J, Faber E, Michl P, Pazdera J, Indrak K. 2009. Saliva as a diagnostic medium. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **153**(2): 103-110.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**(7285): 59-65.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**(Database issue): D590-596.

Quirynen M. 1994. The clinical meaning of the surface roughness and the surface free energy of intra-oral hard substrata on the microbiology of the supra- and subgingival plaque: results of in vitro and in vivo experiments. *J Dent* **22 Suppl 1**: S13-16.

Rappe MS, Giovannoni SJ. 2003. The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369-394.

Rasiah IA, Wong L, Anderson SA, Sissons CH. 2005. Variation in bacterial DGGE patterns from human saliva: over time, between individuals and in corresponding dental plaque microcosms. *Arch Oral Biol* **50**(9): 779-787.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO et al. 2011. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4680-4687.

Rawls JF, Mahowald MA, Ley RE, Gordon JI. 2006. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* **127**(2): 423-433.

Redford AJ, Bowers RM, Knight R, Linhart Y, Fierer N. 2010. The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ Microbiol* **12**(11): 2885-2893.

Reeder J, Knight R. 2010. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* **7**(9): 668-669.

Rehani K, Scott DA, Renaud D, Hamza H, Williams LR, Wang H, Martin M. 2008. Cotinine-induced convergence of the cholinergic and PI3 kinase-dependent anti-inflammatory pathways in innate immune cells. *Biochim Biophys Acta* **1783**(3): 375-382.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**(7304): 334-338.

Rhea SA, Bricker JB, Corley RP, Defries JC, Wadsworth SJ. 2013a. Design, Utility, and History of the Colorado Adoption Project: Examples Involving Adjustment Interactions. *Adopt Q* **16**(1): 17-39.

Rhea SA, Bricker JB, Wadsworth SJ, Corley RP. 2013b. The Colorado Adoption Project. *Twin Res Hum Genet* **16**(1): 358-365.

Rhea SA, Gross AA, Haberstick BC, Corley RP. 2006. Colorado Twin Registry. *Twin Res Hum Genet* **9**(6): 941-949.

-. 2013c. Colorado Twin Registry: an update. *Twin Res Hum Genet* **16**(1): 351-357.

Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR et al. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**(6150): 1241214.

Riedel F, Goessler U, Hormann K. 2003. Alcohol-related diseases of the mouth and throat. *Best Pract Res Clin Gastroenterol* **17**(4): 543-555.

Robertson CE, Baumgartner LK, Harris JK, Peterson KL, Stevens MJ, Frank DN, Pace NR. 2013a. Culture-independent analysis of aerosol microbiology in a metropolitan subway system. *Appl Environ Microbiol* **79**(11): 3485-3493.

Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, Feazel LM, Park K, Pace NR, Frank DN. 2013b. Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. *Bioinformatics*.

Robinson CJ, Bohannan BJ, Young VB. 2010. From structure to function: the ecology of host-associated microbial communities. *Microbiol Mol Biol Rev* **74**(3): 453-476.

Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS et al. 2012. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation* **125**(1): e2-e220.

Romeo J, Warnberg J, Nova E, Diaz LE, Gomez-Martinez S, Marcos A. 2007. Moderate alcohol consumption and the immune system: a review. *Br J Nutr* **98 Suppl 1**: S111-115.

Rostron B. 2013. Smoking-attributable mortality by cause in the United States: revising the CDC's data and estimates. *Nicotine Tob Res* **15**(1): 238-246.

Rudney JD, Xie H, Rhodus NL, Ondrey FG, Griffin TJ. 2010. A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. *Mol Oral Microbiol* **25**(1): 38-49.

Russell RB. 2008. The Molecular Biology of Cariogenic Bacteria. In *Molecular Oral Microbiology*, (ed. AH Rogers), pp. 109-133. Caister Academic Press, Norfolk, UK.

Saarela MH, Dogan B, Alaluusua S, Asikainen S. 1999. Persistence of oral colonization by the same Actinobacillus actinomycetemcomitans strain(s). *J Periodontol* **70**(5): 504-509.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**(4732): 1350-1354.

Salaspuro V, Salaspuro M. 2004. Synergistic effect of alcohol drinking and smoking on in vivo acetaldehyde concentration in saliva. *Int J Cancer* **111**(4): 480-483.

Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**(3): 441-448.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**(12): 5463-5467.

Sayers NM, Drucker DB, Telford DR, Morris JA. 1995. Effects of nicotine on bacterial toxins associated with cot death. *Arch Dis Child* **73**(6): 549-551.

Sayers NM, Gomes BP, Drucker DB, Blinkhorn AS. 1997. Possible lethal enhancement of toxins from putative periodontopathogens by nicotine: implications for periodontal disease. *J Clin Pathol* **50**(3): 245-249.

Schaefer AS, Richter GM, Groessner-Schreiber B, Noack B, Nothnagel M, El Mokhtari NE, Loos BG, Jepsen S, Schreiber S. 2009. Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *PLoS Genet* **5**(2): e1000378.

Schlafer S, Riep B, Griffen AL, Petrich A, Hubner J, Berning M, Friedmann A, Gobel UB, Moter A. 2010. Filifactor alocis--involvement in periodontal biofilms. *BMC Microbiol* **10**: 66.

Schloss PD. 2010. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6**(7): e1000844.

Scott DA, Martin M. 2006. Exploitation of the nicotinic anti-inflammatory pathway for the treatment of epithelial inflammatory diseases. *World J Gastroenterol* **12**(46): 7451-7459.

Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, Huttenhower C, Izard J. 2012. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* **13**(6): R42.

Seitz HK, Stickel F. 2007. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat Rev Cancer* **7**(8): 599-612.

Shaddox LM, Huang H, Lin T, Hou W, Harrison PL, Aukhil I, Walker CB, Klepac-Ceraj V, Paster BJ. 2012. Microbiological characterization in children with aggressive periodontitis. *J Dent Res* **91**(10): 927-933.

Shannon CE, Weaver W. 1949. *The mathematical theory of communication*. University of Illinois Press, Urbana,.

Shchipkova AY, Nagaraja HN, Kumar PS. 2010. Subgingival microbial profiles of smokers with periodontitis. *J Dent Res* **89**(11): 1247-1253.

Shetty V, Mooney LJ, Zigler CM, Belin TR, Murphy D, Rawson R. 2010. The relationship between methamphetamine use and increased dental disease. *J Am Dent Assoc* **141**(3): 307-318.

Signoretto C, Bianchi F, Burlacchini G, Sivieri F, Spratt D, Canepari P. 2010. Drinking habits are associated with changes in the dental plaque microbial community. *J Clin Microbiol* **48**(2): 347-356.

Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmborg JV et al. 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* **6**(5): 399-408.

Siqueira JF, Jr., Fouad AF, Rocas IN. 2012. Pyrosequencing as a tool for better understanding of human microbiomes. *J Oral Microbiol* **4**.

Slomiany BL, Murty VL, Piotrowski J, Slomiany A. 1996. Salivary mucins in oral mucosal defense. *Gen Pharmacol* **27**(5): 761-771.

Slots J, Slots H. 2011. Bacterial and viral pathogens in saliva: disease relationship and infectious risk. *Periodontol 2000* **55**(1): 48-69.

Smith CJ, Hansch C. 2000. The relative toxicity of compounds in mainstream cigarette smoke condensate. *Food Chem Toxicol* **38**(7): 637-646.

Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL, Jr. 1998. Microbial complexes in subgingival plaque. *J Clin Periodontol* **25**(2): 134-144.

Socransky SS, Haffajee AD, Smith C, Martin L, Haffajee JA, Uzel NG, Goodson JM. 2004. Use of checkerboard DNA-DNA hybridization to study complex microbial ecosystems. *Oral Microbiol Immunol* **19**(6): 352-362.

Socransky SS, Smith C, Martin L, Paster BJ, Dewhirst FE, Levin AE. 1994. "Checkerboard" DNA-DNA hybridization. *Biotechniques* **17**(4): 788-792.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* **103**(32): 12115-12120.

Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S et al. 2013. Cohabiting family members share microbiota with one another and with their dogs. *Elife* **2**: e00458.

Sopori M. 2002. Effects of cigarette smoke on the immune system. *Nat Rev Immunol* **2**(5): 372-377.

Spor A, Koren O, Ley R. 2011. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9**(4): 279-290.

Spratt DA, Daglia M, Papetti A, Stauder M, O'Donnell D, Ciric L, Tymon A, Repetto B, Signoretto C, Houri-Haddad Y et al. 2012. Evaluation of plant and fungal extracts for their potential antigingivitis and anticaries activity. *J Biomed Biotechnol* **2012**: 510198.

Stahringer SS, Clemente JC, Corley RP, Hewitt J, Knights D, Walters WA, Knight R, Krauter KS. 2012. Nurture trumps nature in a longitudinal survey of salivary bacterial communities in twins from early adolescence to early adulthood. *Genome Res* **22**(11): 2146-2152.

Staley JT, Konopka A. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**: 321-346.

Stallings MC, Corley RP, Dennehey B, Hewitt JK, Krauter KS, Lessem JM, Mikulich-Gilbertson SK, Rhee SH, Smolen A, Young SE et al. 2005. A genome-wide search for quantitative trait loci that influence antisocial drug dependence in adolescence. *Arch Gen Psychiat* **62**(9): 1042-1051.

Stallings MC, Corley RP, Hewitt JK, Krauter KS, Lessem JM, Mikulich SK, Rhee SH, Smolen A, Young SE, Crowley TJ. 2003. A genome-wide search for quantitative trait loci influencing substance dependence vulnerability in adolescence. *Drug Alcohol Depen* **70**(3): 295-307.

Stewart EJ. 2012. Growing unculturable bacteria. *J Bacteriol* **194**(16): 4151-4160.

Stewart JA, Chadwick VS, Murray A. 2005. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J Med Microbiol* **54**(Pt 12): 1239-1242.

Stranneheim H, Lundeberg J. 2012. Stepping stones in DNA sequencing. *Biotechnol J* **7**(9): 1063-1073.

Substance Abuse and Mental Health Services Administration. 2011. Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings. (ed. SAaMHS Administration), Rockville, MD.

Swan GE, Lessov-Schlaggar CN. 2007. The effects of tobacco smoke and nicotine on cognition and the brain. *Neuropsychol Rev* **17**(3): 259-273.

Takeda I, Stretch C, Barnaby P, Bhatnager K, Rankin K, Fu H, Weljie A, Jha N, Slupsky C. 2009. Understanding the human salivary metabolome. *NMR Biomed* **22**(6): 577-584.

Talloczy Z, Martinez J, Joset D, Ray Y, Gacser A, Toussi S, Mizushima N, Nosanchuk JD, Goldstein H, Loike J et al. 2008. Methamphetamine inhibits antigen processing, presentation, and phagocytosis. *PLoS Pathog* **4**(2): e28.

Tang G, Yip HK, Cutress TW, Samaranayake LP. 2003. Artificial mouth model systems and their contribution to caries research: a review. *J Dent* **31**(3): 161-171.

Tanner AC, Mathney JM, Kent RL, Chalmers NI, Hughes CV, Loo CY, Pradhan N, Kanasi E, Hwang J, Dahlan MA et al. 2011. Cultivable anaerobic microbiota of severe early childhood caries. *J Clin Microbiol* **49**(4): 1464-1474.

Teucher B, Skinner J, Skidmore PM, Cassidy A, Fairweather-Tait SJ, Hooper L, Roe MA, Foxall R, Oyston SL, Cherkas LF et al. 2007. Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res Hum Genet* **10**(5): 734-748.

Tipton DA, Legan ZT, Dabbous M. 2010. Methamphetamine cytotoxicity and effect on LPS-stimulated IL-1beta production by human monocytes. *Toxicol In Vitro* **24**(3): 921-927.

Tonetti MS. 2009. Periodontitis and risk for atherosclerosis: an update on intervention trials. *J Clin Periodontol* **36 Suppl 10**: 15-19.

Tremaroli V, Backhed F. 2012. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**(7415): 242-249.

Tringe SG, Hugenholtz P. 2008. A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* **11**(5): 442-446.

Tringe SG, Rubin EM. 2005. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* **6**(11): 805-814.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP et al. 2009. A core gut microbiome in obese and lean twins. *Nature* **457**(7228): 480-484.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. *Nature* **449**(7164): 804-810.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**(7122): 1027-1031.

U.S. Department of Health and Human Services. 2000. Oral health in America: a report of the Surgeon General. Rockville, MD.

Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R. 2012. The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol* **129**(5): 1204-1208.

US Department of Transportation NHTSA. Drugs and Human Performance Fact Sheets - Methamphetamine. In *Drugs and Human Performance Fact Sheets*, Vol 2012. US Department of Transportation, National Highway Traffic Safety Administration, Washington, DC

Valencia F, Bubar MJ, Milligan G, Cunningham KA, Bourne N. 2012. Influence of methamphetamine on genital herpes simplex virus type 2 infection in a mouse model. *Sex Transm Dis* **39**(9): 720-725.

Van Winkelhoff AJ, Boutaga K. 2005. Transmission of periodontal bacteria and models of infection. *J Clin Periodontol* **32 Suppl 6**: 16-27.

Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, Sitaraman SV, Knight R, Ley RE, Gewirtz AT. 2010. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**(5975): 228-231.

Vinkhuyzen AA, van der Sluis S, de Geus EJ, Boomsma DI, Posthuma D. 2010. Genetic influences on 'environmental' factors. *Genes Brain Behav* **9**(3): 276-287.

von Wintzingerode F, Gobel UB, Stackebrandt E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**(3): 213-229.

Wade WG. 2013. The oral microbiome in health and disease. *Pharmacol Res* **69**(1): 137-143.

Wagner BD, Robertson CE, Harris JK. 2011. Application of two-part statistics for comparison of sequence variant counts. *PLoS One* **6**(5): e20296.

Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. 2011. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* **27**(8): 1159-1161.

Wang GC, Wang Y. 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl Environ Microbiol* **63**(12): 4645-4650.

Wang GJ, Volkow ND, Chang L, Miller E, Sedler M, Hitzemann R, Zhu W, Logan J, Ma Y, Fowler JS. 2004. Partial recovery of brain metabolism in methamphetamine abusers after protracted abstinence. *Am J Psychiatry* **161**(2): 242-248.

Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**(16): 5261-5267.

Welch RA, Burland V, Plunkett G, 3rd, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J et al. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc Natl Acad Sci U S A* **99**(26): 17020-17024.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**(12): 6578-6583.

Wickstrom C, Herzberg MC, Beighton D, Svensater G. 2009. Proteolytic degradation of human salivary MUC5B by dental biofilms. *Microbiology* **155**(Pt 9): 2866-2872.

Wilson M. 2005. *Microbial inhabitants of humans : their ecology and role in health and disease*. Cambridge University Press, Cambridge, UK ; New York.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**(11): 5088-5090.

Wood SR, Kirkham J, Marsh PD, Shore RC, Nattress B, Robinson C. 2000. Architecture of intact natural human plaque biofilms studied by confocal laser scanning microscopy. *J Dent Res* **79**(1): 21-27.

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R et al. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**(6052): 105-108.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP et al. 2012. Human gut microbiome viewed across age and geography. *Nature* **486**(7402): 222-227.

Ye L, Peng JS, Wang X, Wang YJ, Luo GX, Ho WZ. 2008. Methamphetamine enhances Hepatitis C virus replication in human hepatocytes. *J Viral Hepat* **15**(4): 261-270.

Yokoyama A, Tsutsumi E, Imazeki H, Suwa Y, Nakamura C, Mizukami T, Yokoyama T. 2008. Salivary acetaldehyde concentration according to alcoholic beverage consumed and aldehyde dehydrogenase-2 genotype. *Alcohol Clin Exp Res* **32**(9): 1607-1614.

Young SE, Stallings MC, Corley RP, Krauter KS, Hewitt JK. 2000. Genetic and environmental influences on behavioral disinhibition. *Am J Med Genet* **96**(5): 684-695.

Zahradnik RT, Magnusson I, Walker C, McDonell E, Hillman CH, Hillman JD. 2009. Preliminary assessment of safety and effectiveness in humans of ProBiora3, a probiotic mouthwash. *J Appl Microbiol* **107**(2): 682-690.

Zaura E, Keijser BJ, Huse SM, Crielaard W. 2009. Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* **9**: 259.

Zhong Q, Ding C, Wang M, Sun Y, Xu Y. 2012. Interleukin-10 gene polymorphisms and chronic/aggressive periodontitis susceptibility: a meta-analysis based on 14 case-control studies. *Cytokine* **60**(1): 47-54.

Zhu B, Wang X, Li L. 2010. Human gut microbiome: the second genome of human body. *Protein Cell* **1**(8): 718-725.

Zonuz AT, Rahmati A, Mortazavi H, Khashabi E, Farahani RM. 2008. Effect of cigarette smoke exposure on the growth of Streptococcus mutans and Streptococcus sanguis: an in vitro study. *Nicotine Tob Res* **10**(1): 63-67.

Appendix A: Universal methods

A.1 Molecular methods and 454 sequencing

The variable regions 1 and 2 (V1 and V2) of the bacterial 16S rRNA gene were amplified via PCR in triplicates. The bacterial PCR primers are composed of 454 Life Science primers as well as the broadly conserved bacterial primer 27F and 338R respectively (27F: 5'-GCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG-3', 338R: 5'-GCCTCCCTCGCGCCATCAGNNNNNNNNNNNNCATGCTGCCTCCCGTAGGAGT-3'). To allow for pooling more than one sample on one sequencing run, samples had to be tagged with a nucleotide barcode to be able to discern the origin of a particular sequence read (Andersson et al. 2008; Hamady et al. 2008). Each 338R primer contained a unique 12-bp error corrected barcode, marked as NNNNNNNNNNNN specific for each individual sample to trace back the origin of each sequence obtained from the pooled parallel sequencing.

PCR amplifications were set up in triplicates (25µl each) with 0.6µM forward and reverse primer, 1µl of template DNA and 1X of Invitrogen Platinum PCR Super mix. The PCR conditions were: initial denaturation at 94°C for 5 min, denaturation at 94°C for 45 sec, annealing at 50°C for 30 sec and elongation at 72°C for 90 sec for a total of 32 cycles. The final elongation step was 10 min at 72°C followed by a 4°C hold. Extreme care was taken to avoid environmental contaminations. All PCR reactions were assembled in a UV irradiated tissue culture hood and all equipment was wiped down with DNAaway (Molecular BioProducts). 3µl of the PCR product was visualized on an Ethidium Bromide stained 1% agarose gel in 0.5X TBE. The triplicates were pooled and cleaned up with Mobio Ultraclean-htp with an elution volume of 100µl TE. The concentration was measured by a Quant-iT PicoGreen dsDNA Assay (Invitrogen) on a plate reader. 100ng of each sample was pooled in a 15ml Falcon tube. If less than 100ng of DNA was available, 80µl of this sample was used. The pooled DNA samples were precipitated in a 15ml Falcon tube with 2.2xVol 95% EtOH and 1/10xVol of 3M NaOAc, washed

with 70% EtOH and resuspended in 142.5µl H_20 for a final concentration of 80ng/µl, determined with a Nanodrop spectrophotometer (Fisher). Unidirectional sequencing with primer A was done on a 454 Life Sciences Genome Sequencer Ti (Roche) at the Memorial Sloan-Kettering Cancer Center, NY for the twin study and Engencore, University of South Carolina, Columbia, SC for the drug study.

A.2 Quality filtering, clustering and taxonomy assignment

All of the following steps were computed with QIIME, version 1.2.0 and 1.3.0 for the twin dataset and version 1.7.0 for the drug dataset.

The analysis started with generation of a *fasta* file with quality scores from the raw *sff*-files obtained directly from the 454 sequencing machine. The quality scores of the raw 454 sequences were examined with quality_scores_plot.py and the sequences were truncated at the suggested nucleotide position (position: 418 for the drug study), as indicated by a drop in quality. Sequences were demultiplexed and assigned to their original sample identities via the 12 base pair barcodes. Quality filters identified low quality or short length sequences that were excluded from further analysis. The thresholds were: lengths shorter than 200 bp, mean quality score below 25, any ambiguous bases, a stretch of homopolymers longer than 6, an uncorrectable barcode, or no identifiable primer sequence. The data was then clustered ("denoised") to remove sequences with potential sequencing errors and chimeras were removed according to the QIIME tutorial (http://qiime.org/tutorials).

For the twin dataset, clustering and chimera removal, I used OTUpipe (Robert Edgar, http://drive5.com/otupipe). For the drug dataset, the flowgrams were clustered (“denoised”) using Denoiser (Reeder and Knight 2010). Chimeras were identified with ChimeraSlayer (Haas et al. 2011) and removed. Prior to chimera detection, the reverse barcode sequence was identified, if present, and removed from analysis in the drug dataset. Similar sequences were clustered de novo, presorted by abundance, into OTUs with a sequence identity of a minimum

of 97% using uclust (Edgar 2010). Representative sequences were aligned with PyNAST (Caporaso et al. 2010a) against the Greengenes database, version 4feb11 for the twin dataset and version 13_5 (May 2013) for the drug dataset (DeSantis et al. 2006). The alignment parameters were default with a minimum of 150 bp length and a minimum identity of 75%. To assign taxonomy in the twin dataset, the RDP database was used to assign the genera (Cole et al. 2007) with a confidence of 0.8. For other taxonomic assignments, a blast against the Greengenes dataset (version: 4feb2011, (DeSantis et al. 2006)) was used. For the drug dataset, taxonomy was assigned with RDP classifier (Wang et al. 2007) implemented in QIIME 1.7.0 using the Greengenes 13_5 release with a confidence of 0.8 (DeSantis et al. 2006; McDonald et al. 2012). Although no mechanical disruption method was used in the extraction of the DNA from saliva, I readily detected known "hard to lyse" species including *Actinomyces* and *Streptococcus* in comparable abundances to previously published studies that used zirconia beads to disrupt bacterial cell walls (Keijser et al. 2008). However, I have not excluded a systematic bias of our DNA extraction method compared with other studies.

The sequencing of the drug study had to be repeated because a first 454 run yielded 120,967 reads of which only 54,748 (45.3% of total) passed quality control. One main issue, besides the low read number was a C homopolymer in the SSU complementary part of the reverse primer sequence 5'-CATGCTGCCTCCCGTAGGAGT-3'. The original sequence has a homopolymer of three Cs, it was frequently and incorrectly determined as four Cs (5'-CATGCTGCCT***CCC***(***C***)GTAGGAGT-3'). 23,761 reads were affected (19.6% of total). In order to eliminate further inconsistencies, I did not attempt to correct the error; instead, the whole 454 run was discarded. There was a suspected issue with reagent problems at this time. A second run of the same amplicon pool, which was spiked with low amplicon samples as identified from the first run, yielded an acceptable number of reads (430,099 raw reads, 254,396 reads after quality control, 59.2% of total reads).

A.3 Data analysis and statistical methods

A.3.1 Alpha and beta diversity calculations

Alpha and beta diversity was calculated using the QIIME platform. Richness was estimated with the following alpha diversity metrics: direct count of OTUs ("observed species"), estimated richness based on observed singletons and doubletons, Chao1 (Chao 1984), and a phylogenetic richness estimator, PD whole tree (Faith 1992). Rarefaction curves, based on the average of 10 rarefactions, were visually inspected for significant differences between groups. None were detected.

For between sample diversity, the UniFrac distance was used as the beta diversity metric of choice (Lozupone and Knight 2005; Lozupone et al. 2006). The UniFrac distance is a widely used qualitative (presence/absence) pairwise community comparison measure. It is based on the fraction of branch lengths on a common phylogenetic tree of two samples, which is unique to either one of the samples. UniFrac values range from 0 (identical) to 1 (maximally different, no shared branch lengths).

In order to avoid biases of the beta diversity analysis due to different sampling depths, sequence reads were rarefied at a sequencing depth of 800 or 482 reads/sample for the twin and drug dataset, respectively. PCoA for dimensionality reduction was applied to visualize the dataset and to analyze the data with multiple linear regression. The beta diversity values within and between individuals were extracted from the full distance matrix. The average and standard error of the mean (SEM) of all pairwise comparisons in each group were calculated. The statistical analysis was conducted using a Mann-Whitney U-test (see A.3.1). I repeated the rarefaction for a total of 10 times to exclude random effects, which could appear in the random subsampling on this comparatively low rarefaction depth. The p-value of each permutation was recorded and the lowest significance level that has occurred in at least nine out of 10

rarefactions is presented, which is a very conservative approach. I scored each p-value as one of the 4 significance categories, ns: >0.05, *: <0.05, **: <0.01, ***:<0.001.

A.3.2 Wilcoxon test

The two sample Wilcoxon Signed-Rank Test (also known as "Mann-Whitney" or "Mann-Whitney-U"-test) is applied to determine whether population mean ranks differ. It is non-parametric, that means the test examines the ranks instead of actual values. This statistical test is an alternative for the student's t-test, when normal distribution does not need to be assumed. The Wilcoxon test was used to determine whether the mean rank beta diversities between twin pairs, same individual and unrelated individuals differed. Even though the beta diversities were normally distributed, the Wilcoxon was a more conservative approach. The Mann-Whitney U test was performed in R (R programming environment, http://www.R-project.org) for each of the rarefactions.

A.3.3 ANOVA

The analysis of variance (ANOVA) test is used to test if population means between three or more groups differ. It was used to test whether OTU or higher taxonomic abundances are different between categories, for example, it was applied to age groups in the twin dataset. The built in ANOVA of the QIIME command *otu_category_significance.py* was applied (Caporaso et al. 2010b).

A.3.4 Pearson product-moment correlation

The Pearson product-moment correlation coefficient is used to measure the degree of linear dependence. It was used to test whether OTUs or higher taxonomic abundances are correlated with the continuous variable age in both datasets. The built in Pearson product moment correlation function of the QIIME command *otu_category_significance.py* was applied (Caporaso et al. 2010b).

A.3.5 Bonferroni correction

To address the problem of multiple statistical tests, the results of the ANOVA and Pearson's correlation were Bonferroni corrected. However, the usefulness of this correction has been questioned as it might cause a loss of significantly associated organisms and the false discovery rate (FDR) has been proposed instead (Nakagawa 2004).

A.3.6 Multiple linear regression

I used multiple linear regression in order to model the relationship of multiple explanatory variables simultaneously, for example, age, smoking behavior and gender, on an observed response variable (here: individual PC values).

Multiple linear regression fits a liner equation according to the general form

$$y = \alpha + \beta_1 * x_1 + \beta_2 * x_2 + ... + \beta_k * x_k + e$$

where x_1, x_2, …, x_k are the k explanatory variables for the response or dependent variable y and α is the intercept value when all explanatory variables would be zero.

Parameters β_1, β_2, …, β_k describe the individual partial regression coefficients for each explanatory variable x_1, x_2, …, x_k.

Nonsignificant explanatory variables were dropped from the model. The multiple linear regression was performed in R (R programming environment, http://www.R-project.org).

A.3.7 Two-part test

The two-part test was used to identify OTUs and genera, which differ between two groups, for example smokers and nonsmokers. The two-part test, which was developed by Wagner et al. at the University of Colorado (Wagner et al. 2011), allows for a simultaneous evaluation of differences in bacterial taxa abundance and presence. It takes the characteristics of high throughput next generation 16S rRNA sequencing (zero-inflated, non-negative skewed counts and limited number of samples) into account. It sums the test statistic of two independent statistical tests, one for presence on the proportion of non-zero counts (two-

proportion Z-test) and one for abundance based on the medians of non-zero counts (Wilcoxon test). It reaches significance if either the abundance or presence differs substantially between groups. The built in two-part of Explicet, version 2.8.6, was applied to the drug dataset (Robertson et al. 2013b).

A.3.8 UPGMA - hierarchical clustering

To explore relationships between samples, a hierarchical clustering method, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was applied on 10 jackknifed rarefied unweighted UniFrac distances matrices. Dendrograms with colored node support were visually inspected. The built-in jackknifed beta diversity and UPGMA clustering functions of the QIIME commands jackknifed_beta_diversity.py and make_bootstrapped_tree.py were applied (Caporaso et al. 2010b).

Appendix B: The effect of sampling and extraction methods on the oral microbiota

B.1 Introduction

Like many other institutions, the Institute of Behavioral Genetics at the University of Colorado has several thousand banked DNA samples derived from human saliva originally collected for human genetics studies. All these samples could potentially be analyzed on their microbial content with a wide array of phenotypic metadata available. It has been shown that studies with a shallower sequencing depth, but a larger number of samples yields greater statistical power to discriminate differences (Kuczynski et al. 2010), although the cost of sample collection and DNA extraction is often prohibitive. Banked samples could yield an extensive source of specimens.

However, as a study of the oral cavity has shown, different collection methods may yield different results (Contreras et al. 2010). In this study by Contreras et al., buccal swabs were either immediately frozen in liquid nitrogen or preserved in the Aware Messenger oral fluid collection device (Calypte Biomedical Corporation), which was not developed to preserve nucleic acids. Not surprisingly, the results were vastly different. The Institute for Behavioral Genetics at the University of Colorado has applied up to six different sampling and DNA extraction techniques to yield human DNA from saliva or oral mucosa samples over the past two decades. Often the exact collection method was not recorded. All saliva samples reported in this dissertation were collected with Scope mouthwash, except for the 8-11 year old children in the twin dataset. Testers recalled that they did not use Scope on 8-11 year old twins in the late 1990s, but likely used buccal swabs instead. However, there is recall bias and there was no documentation on the actual method used. Due to this uncertainty, subjects from the CAP project sampled prior to 2001 were also excluded from the analysis of the drug dataset because

these samples exhibited a decrease in alpha diversity. To avoid excluding samples based on weak suspicions, a preliminary experiment to compare different sampling and extraction methods was performed. The sample collection and DNA extraction was performed by Justin Eagles-Soukup, supervised by Simone Stahringer.

This section aims to analyze the differences of six different sampling methods and extraction techniques on four to six individuals (ID) each in order to assess the possibility of combining datasets derived from different sampling methods. The six different collection methods used at the Institute of Behavioral Genetics were:

1. Scope mouthwash – abbreviated with *Scope*
2. Mucosal swabs and subsequent swishing with water – *SwabSwish*
3. Mucosal swabs - *CheekSwab*
4. Commercial saliva collection kits for human DNA extraction (Oragene, DNA Genotek, Ontario, Canada) - *Oragene*
5. Unstimulated saliva - *PhenolSpit*
6. In house developed dehydrated DNA extraction buffer developed by Dr. Andy Smolen, University of Colorado - *Smolen*

B.2 Sample description

Samples were taken at least 2h after teeth brushing, preferably in the morning, from each individual by the same researcher and immediately processed (JES).

1. **Scope:** Approximately 10ml of Scope mouthwash (Original Mint, Proctor&Gamble, Cincinnati, OH) were swished vigorously in the mouth for 30 seconds and then released into a sterile Falcon tube. DNA was extracted according to manufacturer's specification with the Gentra Puregene kit (Qiagen, Hilden, Germany).

2. **SwabSwish:** Mucosal surfaces in the mouth (inner cheeks, gums and tongue) were rubbed with sterile cotton swabs. Immediately after, 10ml of bottled water were swished in the mouth for 30 seconds and then released into a sterile Falcon tube. The cotton swab was added to the water-saliva suspension. DNA was extracted using a common DNA extraction protocol including phenol/chloroform extraction as described below.
3. **CheekSwab:** Mucosal surfaces were sampled as described in SwabSwish, but without a water rinse. DNA was extracted with the traditional protocol described below.
4. **Oragene**: The commercial saliva collection kit (now Oragene Discover – OGR 500, DNA Genotek, Ontario, Canada) was developed for human DNA collection. The website states that the percentage of microbial DNA "contamination" is comparatively low. DNA was collected from unstimulated saliva and extracted according to manufacturer's specifications.
5. **PhenolSpit**: 3-5ml of unstimulated saliva were expectorated into a sterile Falcon tube and DNA was extracted with the traditional protocol described below.
6. **Smolen**: 3-5ml of unstimulated saliva were expectorated into a sterile Falcon tube and a cotton swab with dehydrated DNA extraction buffer developed by Dr. Andy Smolen was added. DNA was extracted according to manufacturer's specification with the Gentra Puregene kit (Qiagen, Hilden, Germany).

B.2.1 Traditional DNA extraction protocol

This protocol was used for SwabSwish, CheekSwab, and PhenolSpit.

For the SwabSwish method, the swished water is spun down at 3,000g and the supernatant is removed. The pellet is resuspended in 0.5 ml STE (100mM NaCl, 10mM TrisHCL [pH 8.0], and 10 mM EDTA), then the swab is placed in this solution and put in the water bath.

After collecting the sample, place cotton swab, sample end first, into a Falcon tube containing 0.5 ml of STE buffer with 0.5% SDS and 0.2 mg/ml proteinase K.
Vortex samples.
Place samples in 65º water bath for 2 hours.
To collect a maximum amount of buffer from the cotton swabs after lysis, place swabs in a syringe, then place the swab and syringe in a Falcon tube.
Centrifuge for 5 minutes at 1,000rpm.
Isolation of DNA using phenol:chloroform:isoamylalcohol (24:24:1) and chloroform:isoamylalcohol (24:1) followed by isopropanol precipitation. Add an equal volume of phenol:chloroform:isoamylalcohol (24:24:1) to the DNA solution.
Mix on rocking platform for 5 minutes and microcentrifuge for 10 minutes at 10,000 rpm at room temperature. Verify that phases are well separated.
Remove the aqueous phase and transfer to a new tube. Repeat the phenol:chloroform purification. If, after completing, a white precipitate is visible, repeat the phenol:chloroform purification again.
Add an equal volume of chloroform:isoamyl alcohol 24:1. Mix gently for 2 minutes and spin for 1 minute at 10,000 rpm. Carefully remove aqueous top phase and transfer to a new tube.
Precipitate DNA with ethanol or isopropanol, wash twice in 70% EtOH and resuspend in TE buffer with 1 ul RNase.

B.3 Results

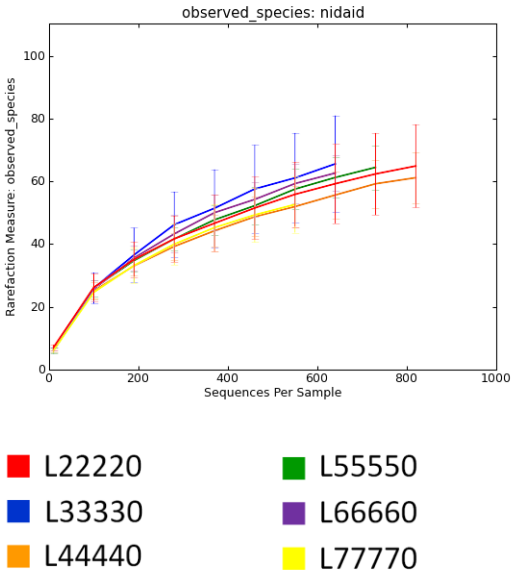
The samples were sequenced on the same 454 plate as the drug experiment. We obtained 31,149 sequence reads, which mapped to the barcodes used in this experiment. Samples with less than 580 reads were excluded from the following analysis. The mean sequencing depth was 974 reads/sample, with a range of 580 to 1631 reads. Alignment, binning, and taxonomy assignment were performed as described before (Appendix A). Table B.1 gives an overview of successful samples by ID and method.

Table B.1: Overview of samples utilized in this preliminary study. Up to six different sampling methods on six individuals were applied. Shaded areas did not yield sufficient reads or were not sampled.

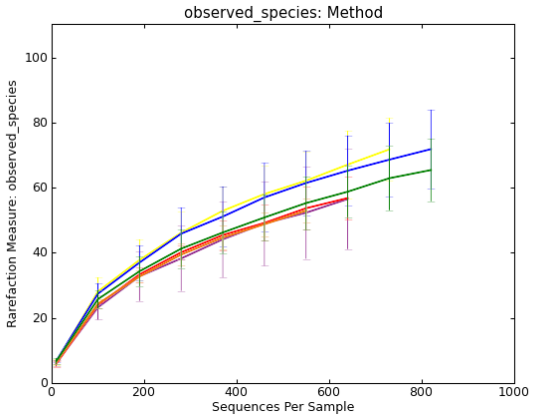
	L22220	**L33330**	**L44440**	**L55550**	**L66660**	**L77770**
Scope	available	available	available	available	available	available
SwabSwish	available	available	available	available	available	available
CheekSwabs	available		available	available	available	
Oragene	available	available	available	available	available	
PhenolSpit	available	available	available	available	available	available
Smolen	available	available	available		available	

The goal of this experiment was to discriminate whether samples cluster by sample method or individual IDs. Alpha diversities based on rarefaction of OTUs observed did not yield significant difference between IDs or sampling methods at the current sequencing depth. However, a trend of clustering into two groups by method can be inferred. Group A is comprised of Oragene and SwabSwish, group B with a lower alpha diversity are Scope, CheekSwab, PhenolSpit, and Smolen (Figure B.1). It is unclear if this result has any significant meaning based on the low number of samples.

A



B



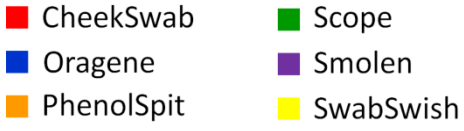
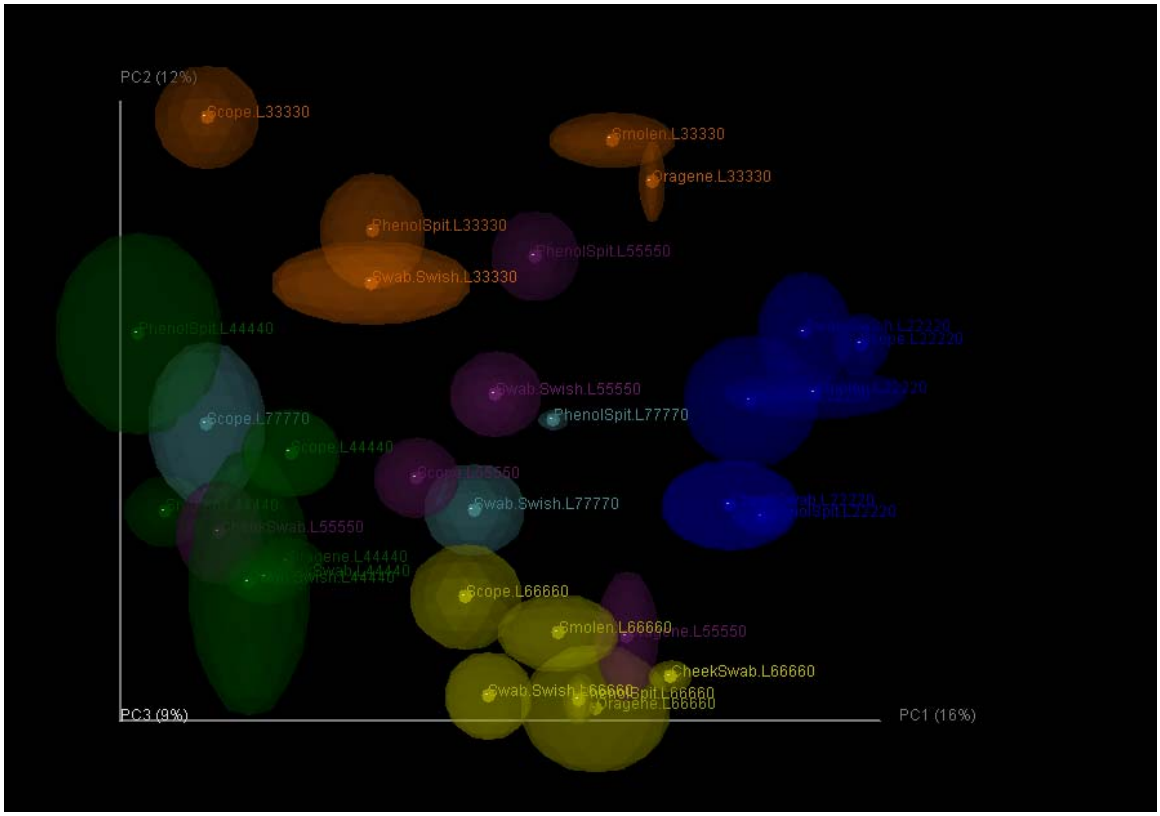


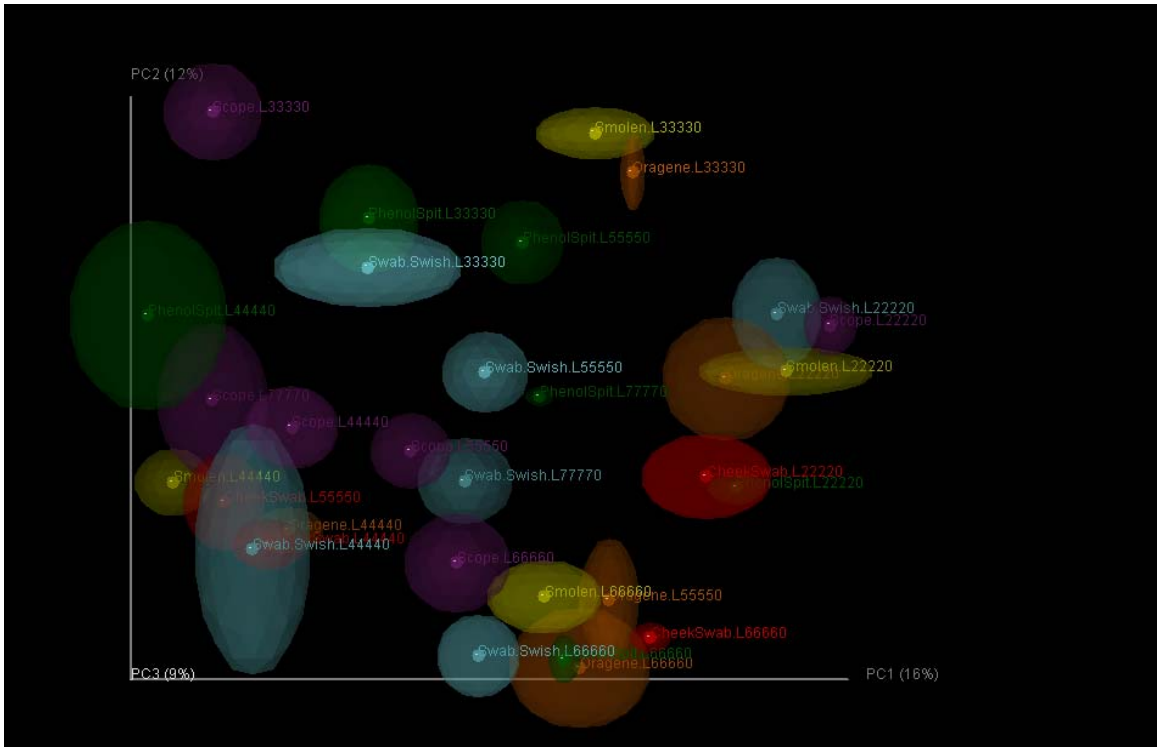
Figure B.1: Alpha diversity rarefaction curves based on OTUs observed are plotted. Results are grouped by A) ID or B) sampling and extraction method.

Similarities based on either ID or sampling and extraction method were assessed based on unweighted UniFrac values. Figure B.2 shows the PCoA plots and UPGMA trees to visualize beta diversity differences.

A


PC2 (12%)
Scope.L33330
Smolen.L33330
Oragene.L33330
PhenolSpit.L33330
PhenolSpit.L55550
Swab.Swish.L33330
PhenolSpit.L44440
Swab.Swish.L55550
PhenolSpit.L77770
Scope.L77770
Scope.L44440
Swab.Swish.L77770
Scope.L66660
Smolen.L66660
CheekSwab.L66660
PC3 (9%)
PC1 (16%)

B


PC2 (12%)
Scope.L33330
Smolen.L33330
Oragene.L33330
PhenolSpit.L33330
PhenolSpit.L55550
Swab.Swish.L33330
PhenolSpit.L44440
Swab.Swish.L22220
Scope.L22220
Swab.Swish.L55550
Smolen.L22220
PhenolSpit.L77770
Scope.L77770
Scope.L44440
Swab.Swish.L77770
CheekSwab.L22220
Smolen.L44440
Oragene.L44440
Swab.Swish.L44440
Scope.L66660
Smolen.L66660
CheekSwab.L66660
Swab.Swish.L66660
PC3 (9%)
PC1 (16%)

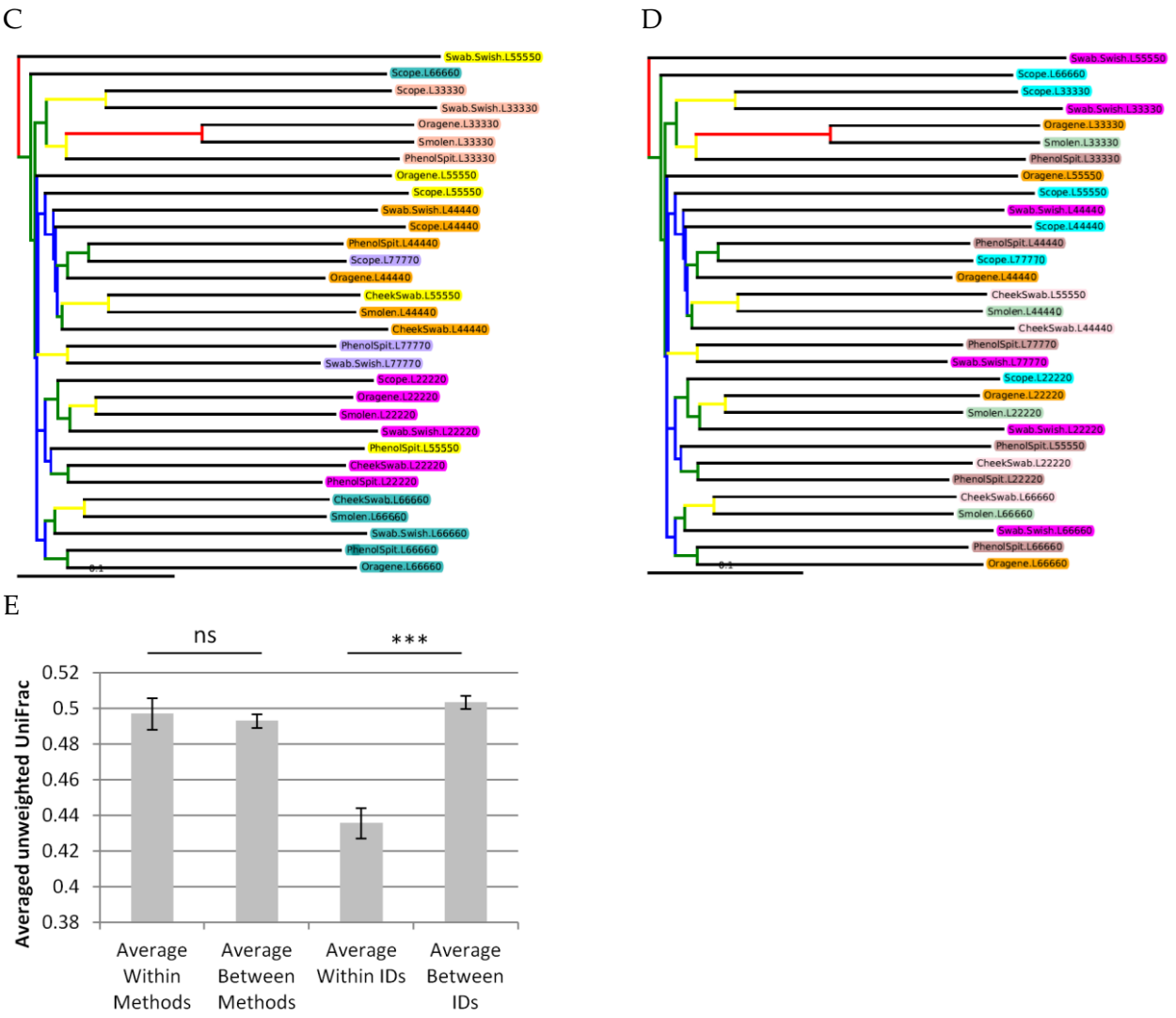


Figure B.2: Between sample differences (beta diversities) based on unweighted UniFrac. A,B: Jackknifed PCoA plots, colored by A) ID and B) sample collection and extraction method. C,D: UPGMA bootstrapped tree. Branch tips are colored by C) ID and D) sample collection and extraction method. The colors of the nodes represent the confidence levels in the tree with red for 75-100% support, yellow for 50-75%, green for 25-50%, and blue for < 25% support. E: Average beta diversities of within and between values of method and IDs. The difference between within method and between method is not significant (two-sided Student's two-sample t-test, $p = 0.69$), whereas the difference within ID and between ID is highly significant ($p = 1.9$ E-12), Error bars: ±SEM. The p-value outcomes are denoted as follows: (ns) nonsignificant, (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

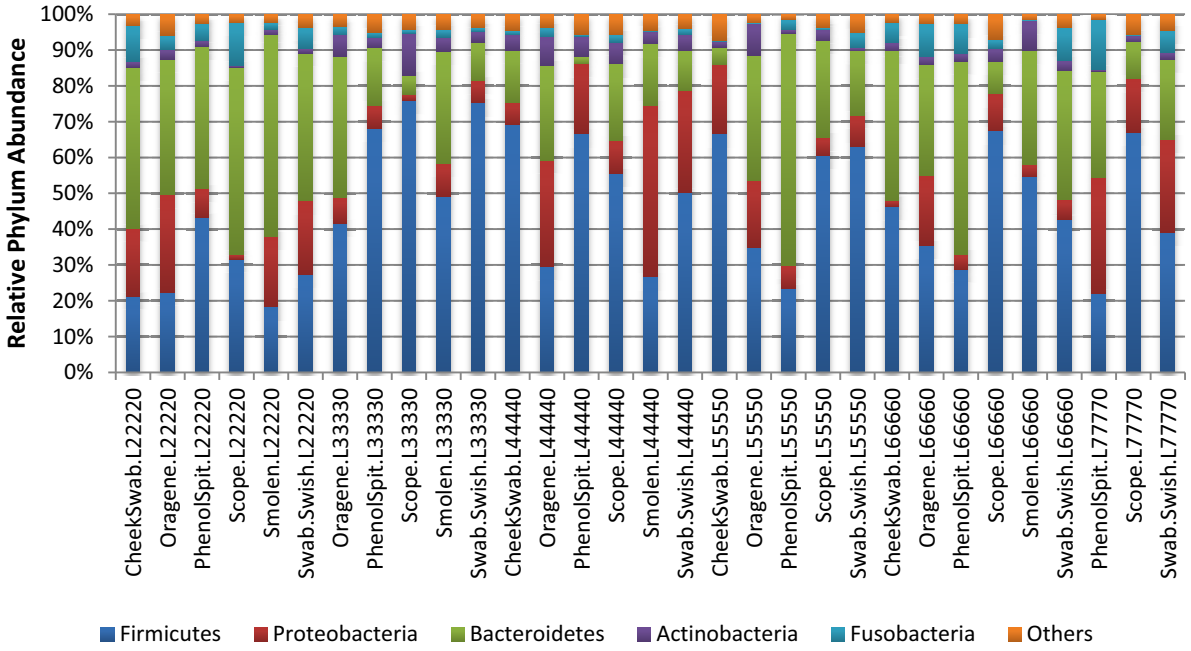
While the overall clustering in the PCoA plot is weak for both variables, there is a trend of clustering by ID, but not sampling method. Likewise, most UPGMA bootstrapped tree nodes

have weak support and there is no clear clustering by ID or sampling method. However, there seems to be an accumulation of the same IDs next to each other, whereas the sampling methods do not cluster.

Furthermore, the averaged beta diversities were calculated within and between methods as well as IDs. If method or ID has an effect on the average beta diversity, significant differences should be seen between the within and between values. As Figure B.2 E shows, there is no significant difference based on sample method, but the averaged beta diversity within IDs is significantly reduced compared to the averaged beta diversity between IDs. This result suggests that the method has no systematic effect on the averaged beta diversity, but identity has.

To assess the differences on a taxonomic level, phylum level distributions were plotted and sorted by ID (Figure B.3 A) and sampling method (Figure B.3 B). No clear trend can be observed. However, certain groups indicate similarities based on ID, as well as method. For example, individual L22220 has an increase in Bacteroidetes in all six sampling methods compared to other samples. L33330 has the highest Firmicute proportion compared to other individuals, which is detected by most sampling methods. Likewise, Oragene derived samples have a low proportion of Firmicutes and Scope derived samples a relatively high proportion.

A



B

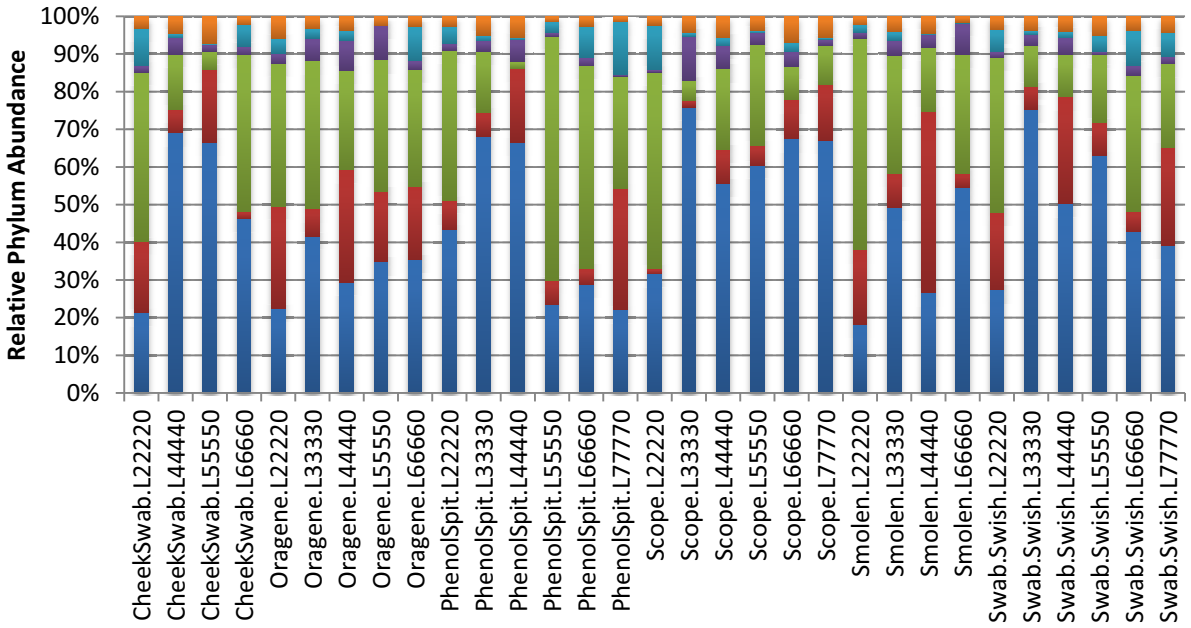


Figure B.3: Stacked Bar graph of main phylum distribution, sorted A) by ID and B) by sampling method.

To analyze these results further, the average phylum abundance of all individuals and all sample methods were grouped. Individuals were grouped by sample method in two ways;

one time all available samples were grouped, the second time only the three individuals where all six sampling methods were available were grouped (L22220, L44440, L66660). This excluded the possibility that unaccounted samples skewed the distribution. However, as Figure B.4 A and B shows, the results are almost identical. That means unaccounted samples are not driving this distribution.

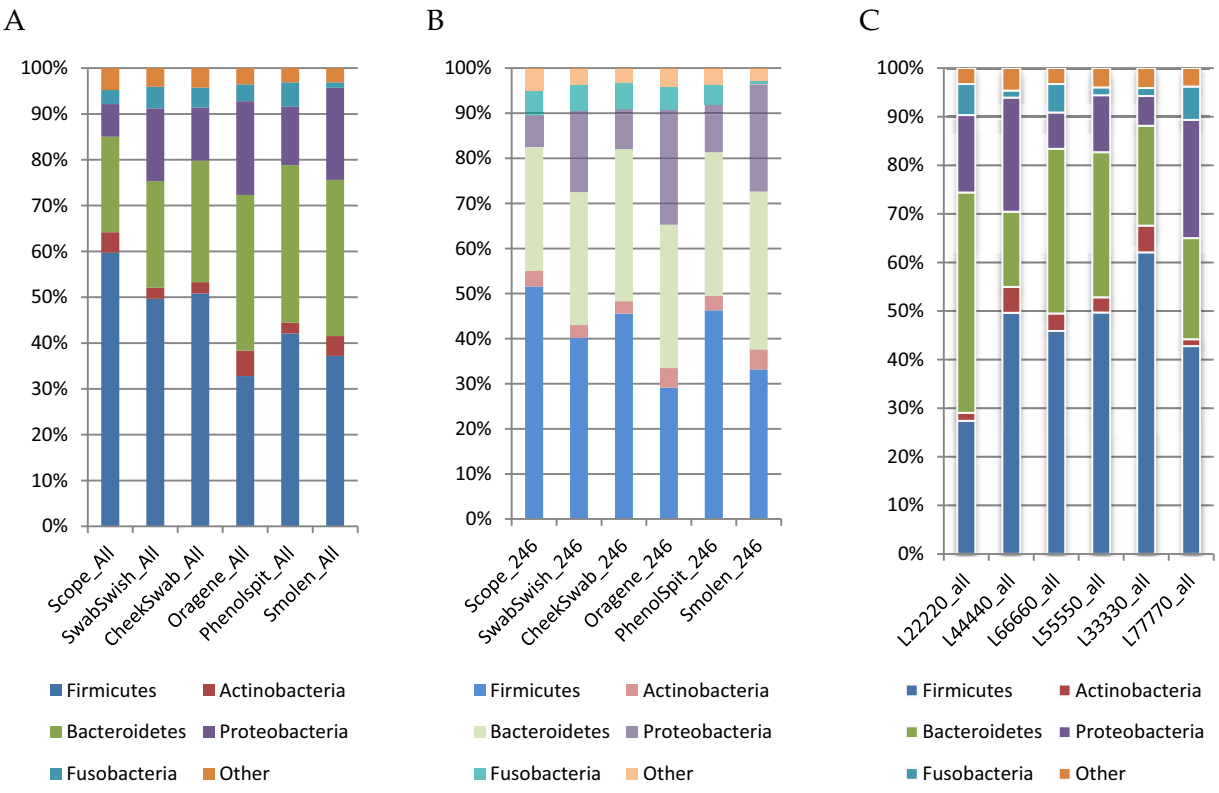


Figure B.4: Relative abundance of major phyla grouped A) by sampling method, B) by sampling method, but only individuals L22220, L44440, L66660 C) by individual.

Figure B.4 A shows differences in main phyla distribution. Of the six sampling methods examined, Scope shows the highest Firmicute content, which is similar to mucosal surfaces as has been shown by Segata et al. (2012) on the HMP dataset. The two sampling methods which are based on sampling the oral cavity with a cotton swab have a slightly lower Firmicute content. This can be explained by the additional sampling of the tongue, which was shown to

have lower Firmicute abundance (Segata et al. 2012). Interestingly, SwabSwich and CheekSwab yield almost identical profiles, which suggests that water samples the mucosal surfaces only superficially, whereas Scope might be more abrasive. Oragene, PhenolSpit, and Smolen are all based on whole, unstimulated saliva and they yield on average similar phylum distribution pattern. Pure saliva has been shown in the HMP to have the lowest Firmicute content compared to other habitats in the oral cavity. Oragene and Smolen which both have a longer incubation period with lysis buffer (from collection to extraction) show a very similar pattern.

As expected and previously shown (Costello et al. 2009; Segata et al. 2012; Stahringer et al. 2012), individuals exhibit different phylum distributions (Figure B.4 C).

B.4 Discussion

The result of this preliminary experiment suggests that while samples from the same individual do not cluster very closely and do not allow identification of an individual, the average unweighted UniFrac value is lower intra-individually than inter-individually. This suggests that different sample collection methods, while not ideal, might be able to be pooled if the overall sample number is high. However, this is a preliminary study with a limited sample number and the results do not allow any definite conclusions at present. Trends in phylum distribution are visible based on sampling method. Besides low sample numbers, another culprit of the study is temporal individual variation as the samples have been collected up to 6 months apart. Studies have shown that day to day variation can be high (Caporaso et al. 2011), even though, on average, intra-individual samples are more like each other than inter-individual samples up to 5 years in time (Costello et al. 2009; Human Microbiome Project Consortium 2012; Stahringer et al. 2012).

To address this question further, an experiment with increasing sample number would be warranted. My opinion at present is that unweighted UniFrac distances are more likely to group samples by individual, while phyla level comparisons might be more influenced by

sampling method and are not recommended. The main sampling methods used at the Institute of Behavioral Genetics are Scope and Oragene. Subsequent studies should take a possible effect of these different sampling methods into account.

Appendix C:
Supplemental information

C.1 Questionnaire wording

Table C.3.1: The questions in the computerized CIDI-SAM

Phenotype: Smoking - Prerequisite for these questions: Has smoked at least 20 cigarettes in their life		
Question Name	**Wording**	**Possible Outcomes**
b1b	How would you describe your usual pattern of cigarette smoking in the last 12 months?	• Every day • 5 or 6 days a week • 3 or 4 days a week • 1 or 2 days a week • 1 to 3 days a month • Less than once a month
b1c	In the past 12 months, when you were smoking cigarettes […], how many would you usually smoke in a day?	[Number of cigarettes]
b1d	When was the last time you had a cigarette?	• Today • Yesterday • 2 to 6 days ago • 7 to 13 days ago • 14 to 20 days ago • 21 to 30 days ago • More than a month ago

Phenotype: Alcohol No prerequisites		
Question Name	**Wording**	**Possible Outcomes**
ndays	Now I'm going to ask you some questions about your use of alcohol like beer, wine, wine coolers, or hard liquor like vodka, gin, or whiskey. Each can or bottle of beer, glass of wine or wine cooler, shot of hard liquor or mixed drink with liquor counts as one drink. […] The next questions are about your use of alcohol in the past week. What did you have to drink yesterday and how much did you drink of each type of alcohol? What about the day before that […]? This was repeated to complete the past 7 days. The number of days drinking was summarized.	[number of days drinking past week] Range: 0 to 7
c2tot4	As before. The total number of drinks was summarized	[number of drinks past week] Range: 0 to infinity

Phenotype: Stimulants Prerequisite for these questions: Has used stimulants more than 5 times in lifetime		
Question Name	**Wording**	**Possible Outcomes**
d2a2r	[W]hen was the last time you used […] stimulants (amphetamines, diet pills, ice, khat, methamphetamine, Ritalin, speed, uppers)?	• past 30 days • not past 30 days, but in past 12 months • more than 12 months ago

Table C.3.2: The question wordings in the CIDI-SAM supplement and paper version 2

Phenotype: Smoking /Tobacco No prerequisites		
Question Name	**Wording**	**Possible Outcomes**
samx1a/dsq1a	Have you ever used Tobacco?	• Yes • No
samx1f/dsq1e	How many days have you used tobacco in the past 6 months?	[number of days] Range: 0 to 180

Phenotype: Stimulants No prerequisites		
Question Name	**Wording**	**Possible Outcomes**
samx5a/dsq5a	Have you ever used amphetamine?	• Yes • No
samx5f/dsq5e	How many days have you used amphetamine in the past 6 months?	[number of days] Range: 0 to 180