UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA

Integrando Arqueas, Bactérias e Fungos do Microbioma Intestinal Humano com a Dieta

Tese apresentada ao Programa de Pós-Graduação em Biologia, da Universidade Federal de Goiás, como requisito parcial para a obtenção do Título de Doutor em Biologia. Área de concentração: Biologia Celular e Molecular.

Orientador: Prof. Dr. Frederic D. Bushman

Co-orientador: Prof. Dr. Luiz Artur M. Bataus

Christian Hoffmann

2013

Dados Internacionais de Catalogação na Publicação (CIP)

Hoffmann, Christian.
H698i Integrando Arqueas, Bactérias e Fungos do Microbioma Intestinal Humano com a Dieta [manuscrito]: / Christian Hoffmann. - 2013.
127 f. : figs, tabs.

Orientador: Prof. Dr. Frederic D. Bushman.
Tese (Doutorado) – Universidade Federal de Goiás, Instituto de Ciências Biológicas, 2013.
Bibliografia.

1. Dieta – Bactérias 2. Dieta – Fungos 3. Microbioma intestinal humano – Dieta 4. Linhagens bacterianas I Título.

CDU: 613.24:616-008.87

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA

Integrando Arqueas, Bactérias e Fungos do Microbioma Intestinal Humano com a Dieta

Tese apresentada ao Programa de Pós-Graduação em Biologia, da Universidade Federal de Goiás, como requisito parcial para a obtenção do Título de Doutor em Biologia. Área de concentração: Biologia Celular e Molecular.

Orientador: Prof. Dr. Frederic D. Bushman

Co-orientador: Prof. Dr. Luiz Artur M. Bataus

Christian Hoffmann

2013

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR AS TESES E DISSERTAÇÕES ELETRÔNICAS (TEDE) NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **[ ] Dissertação** **[X] Tese**

2. Identificação da Tese ou Dissertação

Autor (a):	Christian Hoffmann				
E-mail:	ecohoffmann@gmail.com				
Seu e-mail pode ser disponibilizado na página? [ ]Sim [X] Não					
Vínculo empregatício do autor	Universidade da Pennsylvania				
Agência de fomento:				Sigla:	
País:	Brasil	UF:	GO	CNPJ:	
Título:	Integrando Arquêas, Bactérias e Fungos do Microbioma Intestinal Humano com a Dieta				
Palavras-chave:	Microbioma, Dieta, Fungos, Arquêas				
Título em outra língua:	Integrating the Archaea, Bacteria and Fungi of the Gut Microbiome with Human Diet				
Palavras-chave em outra língua:	Microbiome, Diet, Fungi, Archaea				
Área de concentração:	Biologia Molecular, Ecologia Molecular				
Data defesa: (dd/mm/aaaa)	15/agosto/2013				
Programa de Pós-Graduação:	Biologia				
Orientador (a):	Prof. Dr. Frederic D. Bushman				
E-mail:	bushman@upenn.edu				
Co-orientador (a):*	Prof. Dr. Luis Artur M. Bataus				
E-mail:	bataus@icb.ufg.br				

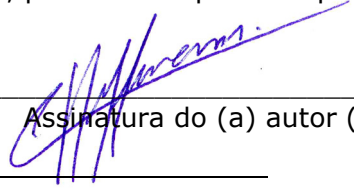
*Necessita do CPF quando não constar no SisPG

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC da tese ou dissertação.

O sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses e ou dissertações, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.



Assinatura do (a) autor (a)

Data: _12_ / _12_ / _2013_

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

DEDICATION

To my parents, for giving me the freedom to do whatever I wished with my life, and the support whenever possible, and sometimes even when not possible, to get me as far as they could on my way. They were directly responsible for my interest in all things scientific.

ACKNOWLEDGEMENTS

I thank my advisor, Rick Bushman, for his ideas, guidance and support, pre and during graduate work, in all the years I have spent in his lab. I have learned much, not only on topics related to my graduate work, but also related to lab life in general. I also thank Gary Wu, James Lewis and Hongzhee Li for incredible discussions throughout the years we have collaborated. Their experience has truly enriched my academic life.

People who have influenced me are too numerous to mention all, but some are kept very close to my heart: Kátia Fernandes, Kenya Silva Cunha, Luis Artur Bataus, you have helped shape the person I am today.

A special thanks to Carol di Medeiros and Ivan Nicolau de Campos. You have always found time for me, be that to answer my questions in the most obscure points of chemistry (whether you liked it or not), or to share in laughter and commiseration as times required.

I am very thankful to all Bushman lab members that have been co-workers and friends, in and out of the lab: Angela Ciuffi for being this incredibly larger than life friend, and with whom I have spent several hours of heated scientific discussion over the years. A special shout out also to all my other Bushman lab friends, with whom I have shared so many fond moments: Stephanie Grunberg, Rebecca Custer-Allen, Karen Ocwieja, Alexandra Bryson, Jacque Young, Jennifer Hwang, Nana Minkah, Anna Kline, Mali Skotheim, Frances Male, Scott Sherrill-Mix and Caitlin Greig. And Young Hwang: after all these years I still learn from his incredible amount of experience and technical expertise.

I also thank my friends, old and new, Nirav Malani, Keshet Ronen, Erik Clarke and Rohini Sinha, for being there when I needed, or just for being there. I cannot begin to enumerate how precious your friendships are to me. You bring warmth, zest and color to my life, and I earnestly hope you will continue to do so.

And I wish to thank John, for being the part of my soul, who was always there, and whose love and butt-kicking skills kept me from procrastinating too much.

ABSTRACT

Diet influences health as a source of nutrients and toxins, and by shaping the composition of resident microbial populations. Previous studies have begun to map out associations between diet and the bacteria and viruses of the human gut microbiome. This work investigates associations of diet with fungal and archaeal populations, taking advantage of samples from 98 well-characterized individuals, and integrates this novel data with the current knowledge regarding the bacteria of the human gut microbiome. Diet was quantified using inventories scoring both long-term and recent diet, and archaea and fungi were characterized by deep sequencing of marker genes in DNA purified from stool. For fungi, we found 66 genera, with generally mutually exclusive presence of either the phyla Ascomycota or Basiodiomycota. For archaea, *Methanobrevibacter*was the most prevalent genus, present in 30% of samples. Several other archaeal genera were detected in lower abundance and frequency. Myriad associations were detected for fungi and archaea with diet, with each other, and with bacterial lineages. *Methanobrevibacter* and*Candida* were positively associated with diets high in carbohydrates, but negatively with diets high in amino acids, protein, and fatty acids. Previously published data emphasized that bacterial population structure was associated primarily with long-term diet, but high *Candida* abundance was most strongly associated with the recent consumption of carbohydrates. *Methobrevibacter* abundance was associated with both long term and recent consumption of carbohydrates. These results confirm earlier targeted studies and provide a host of new associations to consider in modeling the effects of diet on the gut microbiome and human health.

RESUMO

A dieta influencia a saúde sendo uma fonte de nutrientes e toxinas, e por moldar a composição de populações microbianas residentes no corpo humano. Estudos prévios começaram a mapear as associações entre a dieta e bactérias e vírus do microbioma intestinal humano. Este trabalho investiga as associações entre a dieta e populações arqueanas e fúngicas, tomando vantagem de amostras oriundas de 98 indivíduos bem caracterizados, e integra esses dados novos com o conhecimento corrente relacionado a bactérias e o microbioma intestinal humano. A dieta foi quantificada utilizando questionários que acessam a dieta usual e recente, e arquêas e fungos foram caracterizados usando genes marcadores obtidos de amostras fecais através e sequenciamento de DNA em larga escala de última geração. Foram encontrados 66 gêneros de fungos, geralmente com uma presença mutuamente exclusiva dos filos Ascomycota e Basidiomycota. Quanto as arquêas, *Methanobrevibacter* foi o gênero mais prevalente, presente em 30% das amostras. Diversas outras arquêas foram detectadas com abundancia e frequência mais baixa. Uma miríade de associações foi observada entre fungos e arquêas e a dieta, entre fungos e arquêas, e entre estes e linhagens bacterianas. *Metanobrevibacter* e *Candida* foram positivamente associados com uma dieta rica em carboidratos, e negativamente com dietas ricas em amino acidos, proteínas e ácidos graxos. Dados publicados previamente enfatizam que a estrutura das populações bacterianas no intestino são primariamente com hábitos alimentares de longo prazo, porém, uma abundancia alta de *Candida* foi fortemente associada com a ingestao recente de carboidratos. A abundância de Methanobrevibacter foi associada tanto com a ingestão usual ou recente de carboidratos. Estes resultados confirmam estudos direcionados anteriores e provém varias novas associações a serem consideradas quando modelando os efeitos da dieta no microbioma intestinal e a na saúde humana.

TABLE OF CONTENTS

List of Tables and Figures

Figures

Tables

CHAPTER 1. INTRODUCTION

1.1 The human microbiome

The human body, composed of approximately 10 trillion cells, hosts anywhere from 10 to 100 times more microbial cell on its surface and interior. The combined genome of these communities, surpasses the human genome by at least two orders of magnitude(Hooper e Gordon, 2001; Xu, Mahowald *et al.*, 2007). These microorganisms are not evenly distributed in the host's body, several specialized communities exist, each adapted to specific "environmental" factors. Examples of these environmental factors include exposure to oxygen on the surface of the skin, in contrast to the anoxic conditions present in the colon, or variable amount of sweat and sebum across distinct regions of the human tegument (Figure 1). Human babies are born "germ-free", and the several sites which harbor these communities are colonized within the first year of life. The intestinal microbiome, for example, is colonized in a series of waves, which are heavily influenced by external factors, such as the mother's donor microbial community and the baby's diet (Gonzalez, Clemente *et al.*, 2011; Koenig, Spor *et al.*, 2011; Morowitz, Denef *et al.*, 2011), and a balance must exist between all these species living together and the host organism.

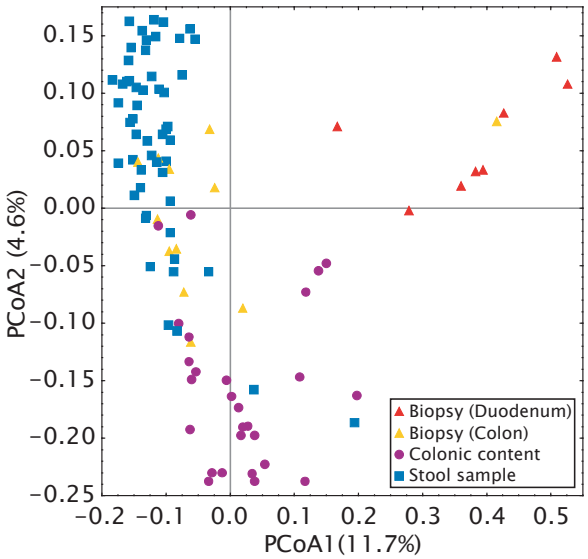


Figure 1-1. Differences in the microbiome across body sites. Beta diversity between samples from distinct body sites was measured using unweighted unifrac, and the unifrac distances were used as input for a principal coordinate analysis. Samples are color coded as the site or origin (Mckenna, Hoffmann *et al.*, 2008).

Extensive characterizations of the human microbiome have usually focused on bacterial groups, with little or no attention paid to other members of this complex ecosystem (Cantarel, Lombard *et al.*, 2012; Faust, Sathirapongsasuti *et al.*, 2012; Thehmpconsortium, 2012). Archaeal species, for example, have been known to be present in the human intestinal tract for decades, having been first detected in individuals who exhaled methane. The production of methane gas is a hallmark of Archaeal metabolism and no other living creature is known to posses this metabolic capacity (Miller, Wolin *et al.*, 1982).

Archaea present in the intestine are usually methanogenic, converting byproducts of bacterial fermentation into methane and preventing the accumulation of H^+. This process is well studied in other anaerobic microbial ecosystems, and Bacteria and Archaea can in fact control each other's growth rates: when Bacteria degradation results in accumulation of H^+, Archaeal growth is stimulated and bacterial growth inhibited; when archaeal methanogenesis removes the excess H^+ produced by bacterial fermentation, its growth rate

decreases and the bacterial growth rate increases again (Stams e Plugge, 2009; Matarazzo, Ribeiro *et al.*, 2012).

Relationships between humans and fungi, although less well described, have been observed for over 100 years, and yeasts, for instance, have been detected within the human intestinal system for nearly as long (Anderson, 1917). However, only a few of the over 160000 currently described fungal species are known to cause disease in humans (www.mycobank.org), and little is understood regarding commensal or possible mutualistic relationships between human or other mammals and their resident fungi(Wuthrich, Deepe *et al.*, 2012). This data is likely to be incomplete, as it relies almost solely on culture based methods, and recent reports using molecular and next generation sequencing techniques are shedding new insight onto mammal/fungi relationships. For example, 27 distinct fungal species were detected in the intestinal microbiome of a single healthy human individual using molecular techniques (Hamad, Sokhna *et al.*, 2012), illustrating the potential diversity of fungi in this ecosystem. Fungi have also been described in several other mammalian gut microbiomes. Over 14 genera been reported to be present within the mucus layer lining the intestinal epithelium in the healthy murine gut (Scupham, Presley *et al.*, 2006), and a recent study using a mouse model and next-generation sequencing demonstrate that normal commensal gut fungi can promote mucosal inflammation in the intestine of mice with sensitized genetic backgrounds (Iliev, Funari *et al.*, 2012). In chapter 3 we address the question of which Fungi and Archaea are present in the human gut, using a cross-sectional survey study of approximately 100 individuals, and explore the inter-domain relationships existing therein.

1.2 Microbial ecology and the human microbiome

The adult human intestinal microbiome is considered one of the places with the highest microbial densities known: we harbor approximately 10^{11} bacterial cells per cm^3 within the colon (Whitman, Coleman *et al.*, 1998). Estimations indicate that the gut microbiome has over 1800 bacterial genera and between 1500 and 3600 bacterial species, the vast majority of which has so far not been cultivated (Frank e Pace, 2008). Such estimations do not take under consideration other microbial groups which are present within the intestinal system, such as Archaea and Fungi (Dridi, Henry *et al.*, 2009; Dridi, Raoult *et al.*, 2011; Dridi, 2012; Iliev, Funari *et al.*, 2012).

Diversity estimation can aid in the characterization of any given environment. Alpha, beta and gamma diversity (Whittaker, 1972) can help us understand not only the species composition of a locale(alpha diversity), ideally placing each within a niche, but also to characterize differences between locales (beta diversity), eventually attributing such differences to intrinsic biotic and abiotic factors controlling the species distribution.

Unfortunately, the concept of species within Prokaryotes is not a trivial one. Uncultured microorganisms can only be detected using DNA based methods, such as the sequencing of marker genes. Even when cultured isolates exist, distinct physiological capabilities can exist in genetically similar organisms. Nevertheless, general patterns can be discerned from such DNA based surveys. For the human host bacterial surveys across many individuals (Arumugam, Raes *et al.*, 2011), show a very high level of inter-personal variation. Thus, the structure of the intestinal microbiome in each individual can be thought of as being composed by representative species from a larger pool of species, all of which fulfill defined roles in their ecosystem. Different individuals will often have a different subset of species, but they are likely to fulfill similar roles within those individuals (Cho e Blaser, 2012).

Perturbations within the gut ecosystem can lead to altered community states, which could in turn become detrimental to the human host (Figure 2). Such changes within the gut microbial network can be accomplished in many different ways (Lemon, Armitage *et al.*, 2012). A classical example would be the use of antibiotics, which eliminates a detrimental pathogen, but also eliminates essential parts of the microbiome. The equilibrium in the gut microbiome can then be reestablished by the targeted re-introduction of species that can fill in the niche which was made vacant. Another attractive idea would be the targeted removal of species which are only detrimental when the host has specific genetic backgrounds. This can be accomplished by the use of bacteriophages, natural bacterial predators which are extremely specific in their they choice of prey (Biswas, Adhya *et al.*, 2002; Chhibber, Kaur *et al.*, 2013). Finally, the fact that several alternate equivalent states exist among different individuals, allows the complete repopulation of the gut microbiome, even when most species are completely removed. Such cases are common in intestinal infections by *Clostridium dificille*, where it outcompetes autochthonous species, invading all niches available within the gut microbiome. The most promising treatment, so called fecal transplantation, re-introduces the entire microbial ecosystem. The newly introduced species are able to correctly colonize the empty niches and re-establish the normal gut microbiome functions (Van Nood, Vrieze *et al.*, 2013).

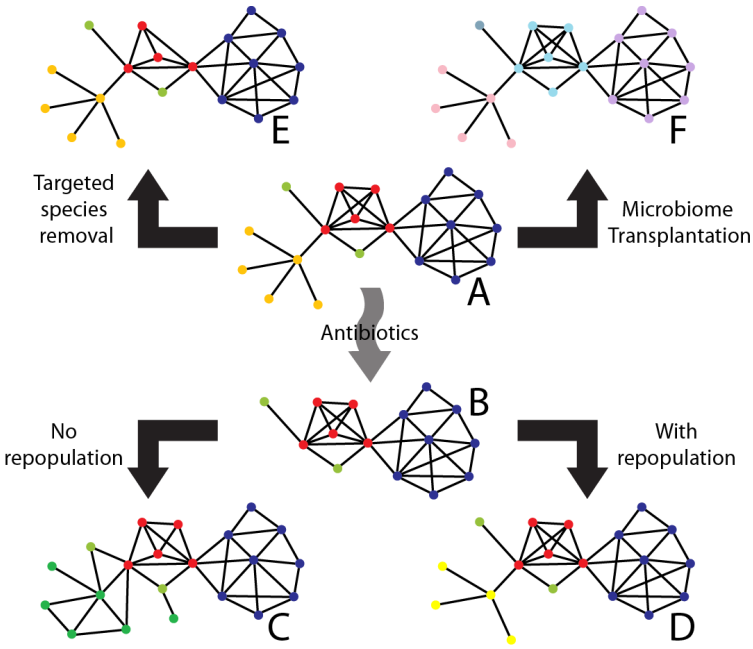


Figure 1-2: The intestinal microbial community network. Microbial interaction are represented as a network and several alternate states are represented. Nodes may be either species or genetic functions. The normal intestinal microbiota (A) can be altered in different ways: Antibiotics when used lead to the depletion of groups (B) which can be can be replenished by the targeted introduction of the bacterial groups (D), replacing the ones removed by the antibiotic therapy (D), and avoiding the insertion of detrimental bacteria to the system (C). Alternatively, a targeted approach could remove only one member of the community(E), or replace the entire community(F) (Figure adapted from Lemon et al 2012. (Lemon, Armitage *et al.*, 2012)).

Although targeted modifications of the gut microbiome by the introduction of selected species is an attractive idea, little is known about the consequences of such introduction over time. Even mild pathogens, for instance, which are naturally cleared by the hosts immune defenses, have profound effects on the overall composition of the gut microbial community (Hoffmann, Hill *et al.*, 2009). Such effects are noticeable even after the pathogen has been removed from the system (Figure 3). There are three main explanations for this: 1) the microbiome has acquired a novel, but equivalent steady state as the one preceding the pathogen introduction: all functions are already replaced, but the community has shifted to a slightly distinct composition; 2) even though the allochthonous species has been removed, and no other signs are present in the environment of its presence, not enough time has elapsed for the microbial community to completely recover from the

disturbance, and 3) key species were permanently removed from the system by the insertion of the pathogen, and the microbiome can no longer return to its previous state, that is, a set of functions has been removed from this system and can only be replaced by the active insertion of species which can fulfill the abrogated functions.

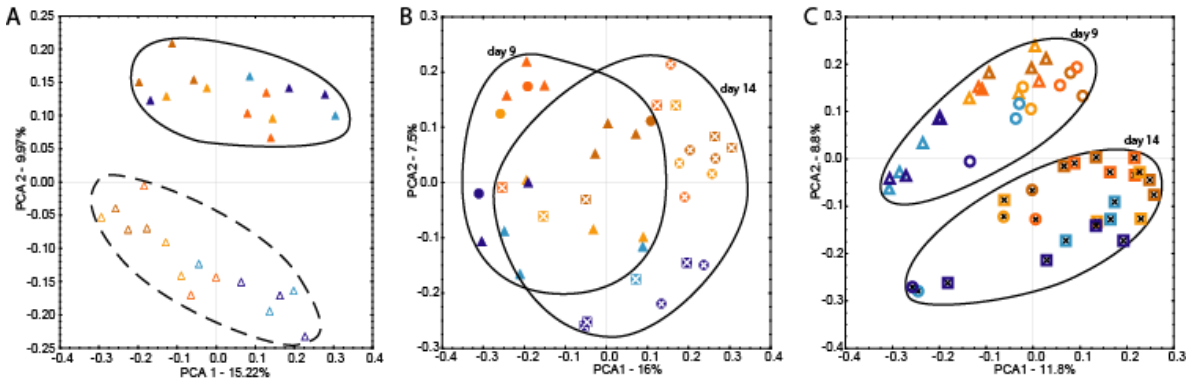


Figure 1-3. Disturbances in the gut microbiome by the insertion of a single allochtonous species. Plots shows a PCoA analysis of Unifrac distance between samples. Samples are color coded by specific intestinal sites: warm colors represent the luminal intestinal content and the cold colors represent the mucosal samples. A) differences between samples from infected (open symbols) and uninfected mice (closed symbols). B) and C) show the temporal effects caused by the introduction of a single species. B) control mice and C) infected mice at day 9 and day 14. The infection is cleared by day 11 and the pathogen itself is not detected any longer at day 14 (Hoffmann, Hill *et al.*, 2009).

1.3 The intestinal microbiome and diet

Relationships between the intestinal microbiome and the host's diet have been explored for a long period of time. Several examples exist for the relationship between the host's diet and intestinal microbiome, but perhaps the most well known is lactose intolerance. In such cases, lactose ingested is never absorbed by the host, and reaches the colonic microbial community, where it is fermented and used as a source of energy (He, Venema *et al.*, 2008). Recently, however, general relationships between mammals and their intestinal microbiome were shown to reflect each group's dietary needs (Figure 4), with the mammalian groups being separated into carnivores, omnivores and herbivores using only the bacterial data present in their intestine (Ley, Hamady *et al.*, 2008; Ley, Lozupone *et al.*, 2008). Considering the large breadth in diet across human populations, this raises the

question of whether such relationships exist in human as well, whether such relationships are related to health or disease states, or whether they can be harnessed to modify the gut microbiome.

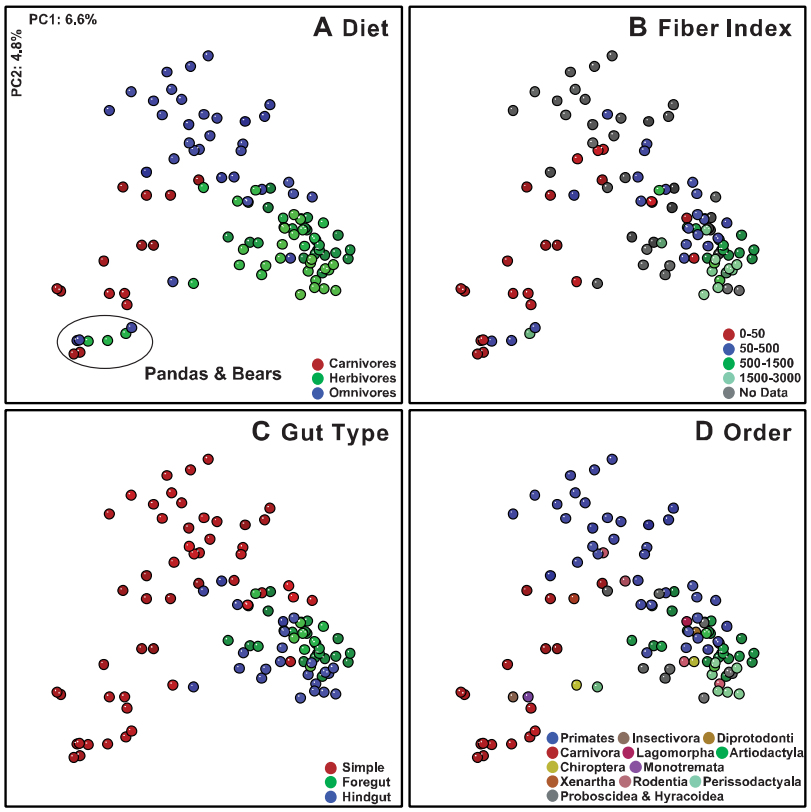


Figure 1-4. Diet separates mammals across several phylogenetic orders. Principal coordinate analysis using Unifrac as a measure of beta diversity across gut microbiome samples. A) samples are colored according to diet type, B) samples are colored according the amount of fiber ingested, C) samples are colored by the type of digestion performed by the host, and, D) samples are colored according to the host's taxonomic order (Figure adapted from Ley et al. 2008 (Ley, Hamady *et al.*, 2008)).

In industrialized countries, diet has changed dramatically in a relatively short period of time. It has shifted from a diet rich in carbohydrates and poor in protein and saturated fats, to the opposite composition, and several studies have linked changes in diet to intestinal inflammatory diseases(Bernstein e Blanchard, 1999; Blanchard, Bernstein *et al.*, 2001; Yang, Loftus *et al.*, 2001). Obesity has also been linked to altered gut microbiome states (Turnbaugh, Hamady *et al.*, 2009) and studies using mouse models have shown that

the composition of the gut microbiome can greatly affected by diet (Figure 5) (Hildebrandt, Hoffmann *et al.*, 2009). Nevertheless, patterns of dietary intake and microbiome composition have only very recently started to be described (Wu, Chen *et al.*, 2011) and have so far focused only on bacterial species. Chapters 4 soughs to complete this picture and presents data on the relationships between the human diet and the fungal and archaeal groups detected in gut microbiome. These relationships are probed using dietary questionnaires to establish both the usual and recent dietary intake in a cohort of healthy human volunteers. In Chapter 5 we continue to explore the relationships between the gut microbiome and the human diet, taking advantage of interventional studies when healthy volunteer subjects are placed onto defined dietary interventions.

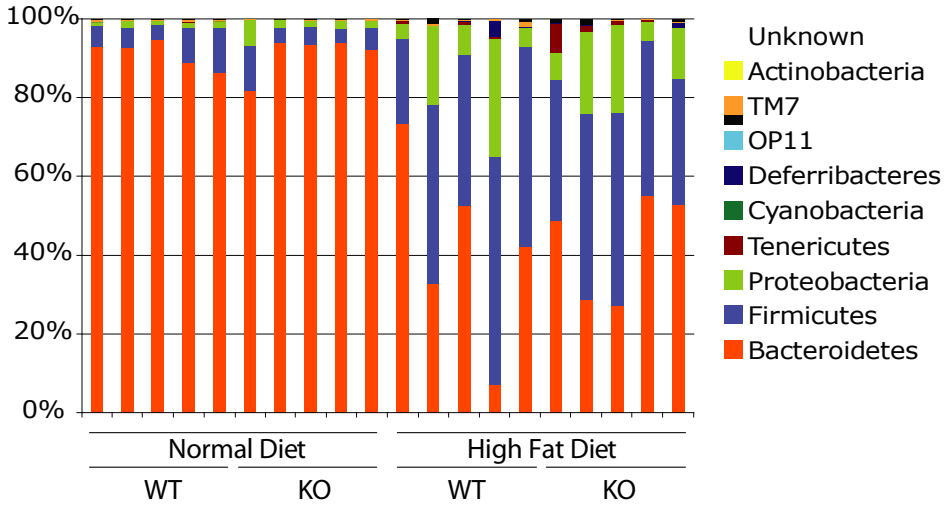


Figure 1-5. Differences in the intestinal microbiome due to diet in mice. Mice were fed a normal diet or a diet rich in fatty acids. Two genetic backgrounds were analyzed (wild type and a Relmβ gene knock out). Phylum level classification for sequences obtained is shown and each stacked bar represents one individual (Hildebrandt, Hoffmann *et al.*, 2009).

1.4 Tools to study the human microbiome

Carl Woese's seminal work (Woese e Fox, 1977) using the 16S rDNA for molecular phylogeny opened a new array of possibilities, such as comparative microbial surveys based only on DNA sequences as a marker of the species present in an environment. The 16S rRNA

gene can be amplified using polymerase chain reaction taking advantage of conserved regions within its sequence (Figure 6). The sequences obtained can be compared to sequences previously established for known microbial groups. Due to conserved physiological functions within microbial groups, even some metabolism can be inferred from such comparisons.

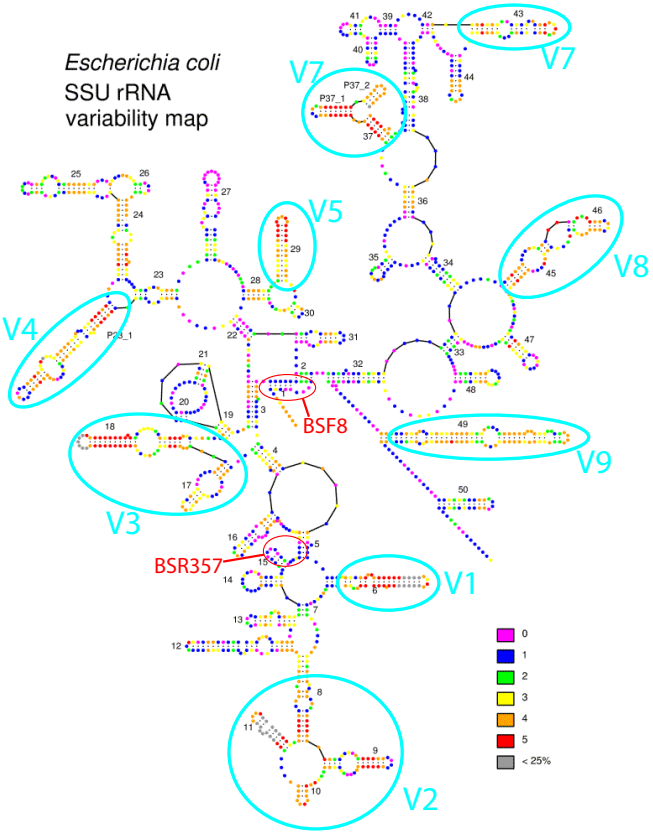


Figure 1-6. Secondary structure of *Escherichia coli*'s 16S rRNA. Colors represent the variability per position within the gene, across a multiple alignment of bacterial sequences. Hyper variable regions are marked from V1 to V9, according to Neefs et al 1993(Neefs, Van De Peer *et al.*, 1993). The annealing site for two commonly utilized pan-bacterial primers, BSF8 and BSR357, are marked in red.

Diversity studies can use the same principles: 16S rDNA sequences, representing individuals in a given community, are obtained and used to characterize the community species composition. Such studies, however, were costly and only a limited number of sequences could be reasonably obtained, limiting the number of samples which could be

studied. Recently, these methods have been coupled with next generation sequencing technologies, greatly decreasing the relative cost of such studies and allowing the characterization of hundreds of samples simultaneously, at a depth of thousands of sequences per sample(Mardis, 2008).

With the species determined, diversity and other traditional ecological methods can be readily used on, or adapted to microbial ecology studies. One such method developed for use in microbial ecology, called Unifrac (Lozupone e Knight, 2005; Lozupone, Hamady *et al.*, 2007; Hamady, Lozupone *et al.*, 2010; Chen, Bittinger *et al.*, 2012), is used to measure beta diversity between samples. Unifrac takes advantage of the phylogenetic relationships existing between species to weight the impact each species has in a comparison (Figure 7). The idea stems from the fact that microbial groups which are closely related tend to have similar metabolic capabilities, while groups distantly related do not. As such, samples containing distinct phylogenetic groups are less similar to each other than samples containing phylogenetically similar groups, even though both absolute richness and evenness may be the same. Beta diversity distance matrices obtained using Unifrac, or other method, can then be subjected to other statistical procedures to try to determine co-variation of the underlining microbial population and environmental factors(Chen, Bittinger *et al.*, 2012).

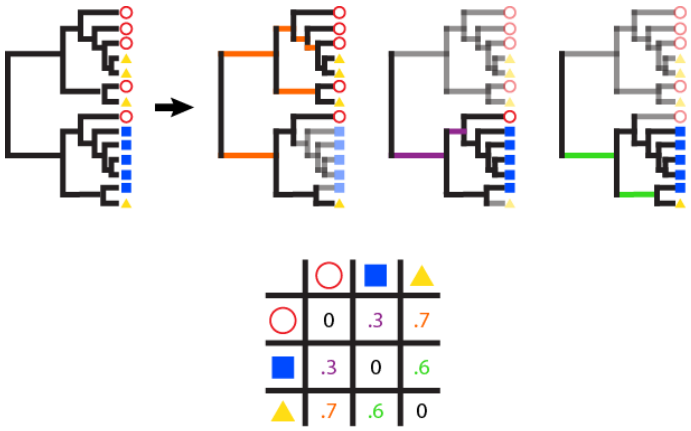


Figure 1-7. Unifrac method. Sequences from all samples being compared are added to a common tree. The shared branch length between each pair of sequences is summed and used to determine a measure of beta diversity. The method can be used with present/absence calls, or be weighted with abundance information for each taxon (Figure adapted from Lozupone, Hamady *et al.*, 2007 and Hamady, Lozupone *et al.*, 2010).

Although methods are available in the published literature for the study of gut microbiome samples, very little information is available that could be used in studies sampling larger cohorts of human subjects. We address these issues in Chapter 2, determining the effects sample storage and the methods employed to obtain 16S rDNA sequences have on the results obtained in gut microbial surveys.

1.5 References

ANDERSON, H. W. Yeast-Like Fungi of the Human Intestinal Tract. *The Journal of Infectious Diseases*, v. 21, n. 4, p. 13, 1917.

ARUMUGAM, M. *et al.* Enterotypes of the human gut microbiome. *Nature*, v. 473, n. 7346, p. 174-80, May 12 2011.

BERNSTEIN, C. N.; BLANCHARD, J. F. The epidemiology of Crohn's disease. *Gastroenterology*, v. 116, n. 6, p. 1503-1504, Jun 1999.

BISWAS, B. *et al.* Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant Enterococcus faecium. *Infect Immun*, v. 70, n. 1, p. 204-10, Jan 2002.

BLANCHARD, J. F. *et al.* Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis. *American Journal of Epidemiology*, v. 154, n. 4, p. 328-335, Aug 15 2001.

CANTAREL, B. L. *et al.* Complex carbohydrate utilization by the healthy human microbiome. *PLoS One*, v. 7, n. 6, p. e28742, 2012.

CHEN, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, v. 28, n. 16, p. 2106-13, Aug 15 2012.

CHHIBBER, S. *et al.* Co-Therapy Using Lytic Bacteriophage and Linezolid: Effective Treatment in Eliminating Methicillin Resistant Staphylococcus aureus (MRSA) from Diabetic Foot Infections. *PLoS One*, v. 8, n. 2, p. e56022, 2013.

CHO, I.; BLASER, M. J. The human microbiome: at the interface of health and disease. *Nat Rev Genet*, v. 13, n. 4, p. 260-70, Apr 2012.

DRIDI, B. Laboratory tools for detection of archaea in humans. *Clin Microbiol Infect*, v. 18, n. 9, p. 825-33, Sep 2012.

DRIDI, B. *et al.* High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol. *PLoS ONE*, v. 4, n. 9, p. e7063, 2009.

DRIDI, B. *et al.* Archaea as emerging organisms in complex human microbiomes. *Anaerobe*, v. 17, n. 2, p. 56-63, Apr 2011.

FAUST, K. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol*, v. 8, n. 7, p. e1002606, 2012.

FRANK, D. N.; PACE, N. R. Gastrointestinal microbiology enters the metagenomics era. *Current opinion in gastroenterology*, v. 24, n. 1, p. 4-10, Jan 2008.

GONZALEZ, A. *et al.* Our microbial selves: what ecology can teach us. *EMBO Rep*, v. 12, n. 8, p. 775-84, Aug 2011.

HAMAD, I. *et al.* Molecular detection of eukaryotes in a single human stool sample from Senegal. *PLoS One*, v. 7, n. 7, p. e40888, 2012.

HAMADY, M. *et al.* Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J*, v. 4, n. 1, p. 17-27, Jan 2010.

HE, T. *et al.* The role of colonic metabolism in lactose intolerance. *European journal of clinical investigation*, v. 38, n. 8, p. 541-547, Aug 2008.

HILDEBRANDT, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*, v. 137, n. 5, p. 1716-24 e1-2, Nov 2009.

HOFFMANN, C. *et al.* Community-wide response of gut microbiota to enteropathogenic Citrobacter infection revealed by deep sequencing. *Infection and immunity*, Jul 27 2009.

HOOPER, L. V.; GORDON, J. I. Commensal host-bacterial relationships in the gut. *Science (New York, N.Y.)*, v. 292, n. 5519, p. 1115-1118, May 11 2001.

ILIEV, I. D. *et al.* Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science*, v. 336, n. 6086, p. 1314-7, Jun 8 2012.

KOENIG, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*, v. 108 Suppl 1, p. 4578-85, Mar 15 2011.

LEMON, K. P. *et al.* Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med*, v. 4, n. 137, p. 137rv5, Jun 6 2012.

LEY, R. E. *et al.* Evolution of mammals and their gut microbes. *Science*, v. 320, n. 5883, p. 1647-51, Jun 20 2008.

LEY, R. E. *et al.* Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature reviews.Microbiology*, v. 6, n. 10, p. 776-788, Oct 2008.

LOZUPONE, C.; KNIGHT, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*, v. 71, n. 12, p. 8228-35, Dec 2005.

LOZUPONE, C. A. *et al.* Quantitative and Qualitative {beta} Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, v. 73, n. 5, p. 1576-1585, Mar 2007.

MARDIS, E. R. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, v. 9, p. 387-402, 2008.

MATARAZZO, F. *et al.* The domain Archaea in human mucosal surfaces. *Clin Microbiol Infect*, v. 18, n. 9, p. 834-40, Sep 2012.

MCKENNA, P. *et al.* The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS pathogens*, v. 4, n. 2, p. e20, Feb 8 2008.

MILLER, T. L. *et al.* Isolation of Methanobrevibacter smithii from human feces. *Appl Environ Microbiol*, v. 43, n. 1, p. 227-32, Jan 1982.

MOROWITZ, M. J. *et al.* Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc Natl Acad Sci U S A*, v. 108, n. 3, p. 1128-33, Jan 18 2011.

NEEFS, J. M. *et al.* Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Res*, v. 21, n. 13, p. 3025-49, Jul 1 1993.

SCUPHAM, A. J. *et al.* Abundant and diverse fungal microbiota in the murine intestine. *Appl Environ Microbiol*, v. 72, n. 1, p. 793-801, Jan 2006.

STAMS, A. J.; PLUGGE, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol*, v. 7, n. 8, p. 568-77, Aug 2009.

THEHMPCONSORTIUM. Structure, function and diversity of the healthy human microbiome. *Nature*, v. 486, n. 7402, p. 207-14, Jun 14 2012.

TURNBAUGH, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature*, v. 457, n. 7228, p. 480-484, Jan 22 2009.

VAN NOOD, E. *et al.* Duodenal infusion of donor feces for recurrent Clostridium difficile. *N Engl J Med*, v. 368, n. 5, p. 407-15, Jan 31 2013.

WHITMAN, W. B. *et al.* Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A*, v. 95, p. 6578-6583, 1998.

WHITTAKER, R. H. Evolution and Measurement of Species Diversity. *Taxon*, v. 21, n. 2/3, p. 213-251, may 1972 1972.

WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*, v. 74, n. 11, p. 5088-90, Nov 1977.

WU, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science*, v. 334, n. 6052, p. 105-8, Oct 7 2011.

WUTHRICH, M. *et al.* Adaptive immunity to fungi. *Annu Rev Immunol*, v. 30, p. 115-48, 2012.

XU, J. *et al.* Evolution of Symbiotic Bacteria in the Distal Human Intestine. *PLoS Biol.*, v. 5, n. 7, p. e156, Jun 19 2007.

YANG, S. K. *et al.* Epidemiology of inflammatory bowel disease in Asia. *Inflammatory bowel diseases*, v. 7, n. 3, p. 260-270, Aug 2001.

CHAPTER 2. METHODS FOR INTESTINAL MICROBIOME STUDIES

The contents of this chapter have been published as part of:

> Wu GD, Lewis JD, **Hoffmann C**, Chen YY, Knight R, et al. (2010) Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiology*. **10: 206.** doi:10.1186/1471-2180-10-206.

2.1 Introduction

Next Generation Sequencing technologies allow the characterization of complex microbial ecosystems in a parallelized, fast and relatively cheap manner. One such environment is the gut microbiome, and several studies have linked it, particularly its bacterial component, to an array of host physiological processes. These processes can be beneficial to the host, such as the degradation of indigestible carbohydrates (Lozupone, Hamady *et al.*, 2008) or lead to a disease state, such as obesity and intestinal inflammatory states (Ley, Backhed *et al.*, 2005).

Gut microbiome studies that utilize next-generation sequencing often utilize tissue biopsies of fecal samples to quantify the taxa present therein as well as the relationships between distinct sampling groups. While tissue samples are often difficult to obtain, especially for healthy individuals, feces are an easily accessible sample type, which has been widely studied and shown to offer a reasonable representation of the lower gastro-intestinal tract (Eckburg, Bik *et al.*, 2005; Gill, Pop *et al.*, 2006). However, a large sample size is necessary to correlate the gut microbiome composition and disease states or dietary intake, highlighting the need for developing sampling methods which are practical, consistent, reproducible a and that would introduce as little bias as possible in the results.

Although the number of sequences obtained using next-generation sequencing technologies is not a limiting factor in survey studies, their limited sequence length prompts the need for the careful selection of the amplicon being used. This can be easily accomplished for the 16S rRNA gene normally used for microbial survey studies. The 16S rRNA gene is comprised of highly conserved regions interspersed with more variable regions, allowing PCR primers to be designed to be complementary to universally conserved regions flanking variable regions. Amplification, sequencing, and comparison to databases using such primers allow the identification of various bacterial types and their proportions in a community (Ludwig, Strunk *et al.*, 2004; Wang, Garrity *et al.*, 2007), and these results mimic those obtained using the full length 16S rDNA sequence (Liu, Lozupone *et al.*, 2007).

Here we investigated 1) reproducibility of sampling within a specimen, 2) effects of storage time and temperature, and 3) effects of lysis and DNA purification methods on the bacterial communities detected. Additionally, 3 16S rDNA regions often used to characterize microbial communities were tested, as well as two distinct amplification conditions and two similar methods of deep sequencing. Once sampling and analysis were clarified, we demonstrated the drastic interpersonal variability in the gut microbiome.

2.2 Methods

2.2.1 Sample collection

Ten healthy adult volunteers (at least 18 years old) were recruited to provide a single stool sample within the Center for Clinical and Translational Research at the Hospital of the University of Pennsylvania. Exclusion criteria included having had diarrhea within one week prior to the sample collection, consumption of any antibiotics within four weeks prior to

sample collection, or any prior diagnosis with inflammatory bowel disease, irritable bowel syndrome, celiac sprue, or other chronic inflammatory diseases of the intestines.

After providing informed consent, each participant completed a brief survey describing their medical history and demographic characteristics. Each participant provided a single stool specimen. All specimens were collected using a collection hat that separated the fecal content from urine or the toilet water.

From the specimen provided, a research coordinator immediately removed six samples from the surface of the specimen. Samples 2 through 6 were obtained to be at least 1 cm away from the location of the first sample. All samples were collected in a Faeces Container with Screw Cap (Cat#80.734.001, Sarstedt, Newton, NC) and the sample was leveled with a disposable wooden spatula. The first three samples were placed in empty vials and immediately stored at -80° C. Two specimens were placed in empty tubes and stored in a Styrofoam cooler filled with ice packs. These specimens were transferred to a -80°C freezer after 24 hours and 48 hours, respectively. The final sample was placed in a vial filled with stool stabilizer from the PSP SPIN Stool DNA Plus kit (Invitek). The specimen was shaken but the specimen was not fully dissolved into the stabilizer solution. After 48 hours of storage at room temperature, the specimen was transferred to a -80°C freezer. Three patients had an extra sample collected and processed immediately. The consistency of the stool sample was characterized using the Bristol Stool Scale (Lewis e Heaton, 1997).

2.2.2 DNA isolation, PCR amplification, and amplicon purification

DNA was isolated from approximately 200 mg of stool using three different commercially-available kits: QIAamp DNA Stool Minikit (Cat#51504, Qiagen, Valencia, CA), PSP Spin Stool DNA Plus Kit (Cat#10381102, Invitek, Berlin, Germany), Mo Bio PowerSoil

DNA Isolation Kit (Cat#12888-05, Mo Bio Laboratories, Carlsbad, CA). DNA was isolated exactly as per the manufactures' instructions for both the QIAamp and PSP kits except for a 95˚C lysis incubation for 5 minutes, instead of the 70˚C recommended for the QIAamp kit. For isolation using the Mo Bio kit, the stool sample was vortexed to homogeneity in 1 ml of Mo Bio Lysis Buffer, centrifuged at 1500 rcf for 5 minutes at room temperature. The supernatant was then transferred to the Mo Bio PowerBead tube, incubated for 10 minutes at 65˚C, then 95˚C for an additional 10 minutes, followed by gentle vortexing to disperse the sample in the PowerBead solution. DNA was then isolated as per the manufacturer's instructions.

For the Phenol/Bead beating protocol, the protocol consisted of a re-suspension/disruption and lysis step that was performed prior to purification using the QIAamp Stool Kit. The frozen stool sample was placed within a MoBio 0.7 mm garnet bead tube (Cat# 13123-50 Mo Bio Laboratories, Carlsbad, CA), to which 0.5 mL of Tris equilibrated (pH 8.0) Phenol : Chloroform : IsoAmyl alcohol (25:24:1) (Cat# P3803, Sigma-Aldrich, St. Louis, MO) was added, and the remaining volume was filled up with buffer ASL from the QIAamp Stool Kit (approximately 0.9 mL). The sample was mechanically disrupted by bead beating using a MiniBeadBeater-16 (Cat# 607, Biospec, Bartlesville, OK) for 1 minute. The resulting homogenate was incubated at 95^{O}C for 5 minutes and centrifuged at 13000G for 1 minute to separate the aqueous and phenolic phases. The aqueous phase was transferred to a new 2 mL microcentrifure tube and the volume was completed to 1.2 mL with buffer ASL. One QIAamp Stool Kit inhibitX tablet was added to this lysate and homogenized according to manufacturer specifications. The remaining of the procedure was followed according to the QIAamp Stool Kit pathogen detection protocol.

After quantification by spectrophotometry, 100 ng of DNA was amplified with barcoded primers using 2.5 units of AmpliTaq (Cat# N8080161, ABI, Foster City, CA) in a reaction buffer containing 25 mM $MgCl_2$, 1% Triton, 10 mM dNTPs, and 10 mg/ml BSA (Cat #B90015, New England Biolabs, Ipswich, MA) (Mckenna, Hoffmann *et al.*, 2008). PCR was performed on an ABI 2720 Thermocycler using the following conditions: Initial denaturing at 95˚C for 5 minutes followed by 20 cycles of 95˚C X 30 seconds, 56˚C X 30 seconds, and 72˚C X 1 minute 30 seconds. The reaction was terminated after an 8 minute extension at 72˚C. The amplicons from each DNA sample, which was amplified in quadruplicate, were pooled and gel purified using an 0.8% agarose gel and a QIAquick Gel Extraction Kit (Cat# 28704, Qiagen) per the manufacturer's instructions.

2.2.3 Defined DNA community composition

Two defined DNA mixture were created using 10 different plasmids, each containing a near full length 16S rDNA amplicon, obtained using primers BSF8 and BSR1541. One mixture had an equal amount of each plasmid and one was staggered to contain different proportions of each clone. The strains and proportions on the Staggered mix are: *Clostridium dificile* (ATCC#: BAA-1382) – 39.99%, *Bacteroides fragilis* (ATCC#: 25285) – 32.01%, *Streptococcus pneumoniae* (ATCC#: BAA_334) – 4.92%, *Desulfovibrio vulgaris* (ATCC#: 29579) – 1.95%, *Campylobacter jejunii* (ATCC#: 700819) – 2.03%, *Rhizobium vitis* (ATCC#: BAA_846) - 2.00%, *Lactobacillus delbruekii* (ATCC#: BAA-365) – 5.06%, *Escherichia coli* HB101 – 2.01%, *Treponema sp.* (macaque stool clone) – 7.97%, and *Nitrosomonas sp.* (environmental clone) – 2.04% . Clones were made using the Topo-XL kit (Cat# K4700-20, Invitrogen, Carlsbad, CA). Two polymerases were tested for the Staggered mix, AmpliTaq (as used for stool DNA samples) and GreenTaq (Promega, Madison, WI) as per manufacturer

instructions. The PCR cycling conditions were the same as described for the stool sample DNA.

2.2.4 454/Roche pysequencing methods

Purified amplicon DNAs were quantified using Quant-iT PicoGreen kit (cat# P7589, Invitrogen, Carlsbad, CA) and pooled in equal amounts for pyrosequencing. Pyrosequencing using the 454/Roche GS FLX chemistry was carried out according to the manufacturer's instructions. Pyrosequencing using the Titanium method was carried out using the Titanium genomic kit. Primers for use with each method are in Table 1.

Table 2.1: Sequences of oligonuceotide primers used in this study.

Primer pair	16S variable region queried	Sequencing chemistry	Sequence forward	Sequence reverse
27f(bsf8)/ 357R	v1-v2	GS FLX	5'-AGAGTTTGATCCTGGCTCAG-3'	5'-CTGCTGCCTYCCGTA-3'
27f / 534r	v1-v3	Titanium	5'-AGAGTTTGATCCTGGCTCAG-3'	5'-ATTACCGCGGCTGCTGG-3'
BSF8 / BSR534	v1-v3*	Titanium	5'-AGAGTTTGATCCTGGCTCAG-3'	5'-ATTACCGCGGCTGCTGGC-3'
357f / 926r	v3-v5	Titanium	5'-CCTACGGGAGGCAGCAG-3'	5'-CCGTCAATTCMTTTRAGT-3'
BSF343 / BSR926	v3-v5*	Titanium	5'-TACGGRAGGCAGCAG-3'	5'-CCGTCAATTYYTTTRAGTTT-3'
968f / 1492r	v6-v9	Titanium	5'-AACGCGAAGAACCTTAC-3'	5'-TACGGYTACCTTGTTAYGACTT-3'
BSF917 / BSR1492	v6-v9*	Titanium	5'-GAATTGACGGGGRCCC-3'	5'-GYTACCTTGTTAYGACTT-3'

Pyrosequence reads were uploaded into QIIME and processed as described (Caporaso, Kuczynski *et al.*, 2010). Briefly, QIIME accepts as input bar coded 16S rRNA gene sequences, classifies them using the RDPclassifier (Wang, Garrity *et al.*, 2007), aligns them using Pynast (Caporaso, Bittinger *et al.*, 2010), constructs phylogenetic trees using FastTree2 (Price, Dehal *et al.*, 2009), calculates Unifrac distances, and generates data summaries that proportions of taxa present and PCA plots of Unifrac distances. We used 97% OTUs in the analysis. For the RDP classifier, we required >50% confidence for all calls.

2.2.5 Statistical methods

Clinical characteristics were compared as median, range, counts and percentages.

Unifrac distances were used to access the global relationships between samples. Unifrac takes advantage of the phylogenetic relationships between distinct taxa within samples to quantify the similarity between samples (beta diversity). It quantifies the unique fraction of branch lengths within the phylogenetic tree that can be attributed to any given sample. Thus, samples with distinct, but closely related bacteria, are deemed more similar to each other than samples with the same number of bacteria present, but which are more phylogenetic divergent. Unifrac distances can be calculated using only the taxa's presence/absence information (Unweighted Unifrac), or their relative abundances (Weighted Unifrac).

For analysis in Figures 1 and 2, no corrections for multiple comparisons were applied. UniFrac (Lozupone, Hamady *et al.*, 2006; Lozupone, Hamady *et al.*, 2007) was used to generate distances between all pairs of communities; both weighted and unweighted

UniFrac were used in the analyses. Statistical analysis was carried out by comparing distances within groups to distances between groups (subjects, storage methods, and DNA extraction methods). Comparisons were summarized using the t-statistic and significance assessed using 10,000 label permutations. Clustering was visualized for weighted and unweighted UniFrac data using principal coordinates analysis.

We used the distance based Permutational Multivariate Analysis of Variance (Permanova) to perform an overall test of the difference between the two gold standards (samples taken 1 cm apart from the same piece of stool) and between gold standards and other sampling methods using both the weighted and unweighted UniFrac distance matrix. If the overall test gave significant results, then we used signed rank test on the proportion data to pinpoint the taxonomic groups that showed significant differences in abundance between the two sampling methods. For Permanova tests used to detect differences due to DNA extraction methods and storage methods permutations were restricted to within subject distances.

2.3 Results

2.3.1 Sample acquisition and storage

We collected stool samples from 10 healthy volunteers (described in table 2), each providing one single sample. Aliquots were made from each sample and processed immediately or after different methods of storage as described in table 3. Briefly, samples were either kept at 4^{o}C for 24 hours and frozen at -80^{o}C, kept at 4^{o}C for 48 hours and frozen at -80^{o}C, or re-suspended in the PSP® (Invitek) re-suspension buffer and kept at room temperature (~22^{o}C) and then frozen at -80^{o}C.

Table 2.2: : Characteristics of participants.

Total number of participants		**10**
Sex		
	Female	4
	Male	6
Race		
	Black/African-American	3
	White	7
Bristol stool category		
	1	0
	2	4
	3	1
	4	4
	5	0
	6	1
	7	0
Stool frequency 1-2 times/day		10
Median age (range)		26.5 years (20 - 61)
Median body mass index (range)		25.5 (19.2 - 37.4)
Current smoker		1

2.3.2 Cell lysis and DNA purification

Four methods were used for DNA isolation from the stool samples: the QIAamp DNA Stool Minikit, PSP Spin Stool DNA Plus Kit, and the MoBio Powersoil DNA Isolation Kit. The fourth method, based on the QIAamp DNA Stool Minikit, was also performed, where samples were pre-processed with hot phenol bead beating prior to extraction with commercial kit to improve cell lysis. No discernible differences in DNA yield were observed with the different types of storage tested.

2.3.3 454/Roche pyrosequence analysis

DNA was PCR amplified using standard 16S rDNA primers targeting the V1V2 region and amplicons were gel purified using the Qiagen Gel purification it according to the manufacturer's instructions. The V1V2 region was chosen based on published simulations (Liu, Lozupone *et al.*, 2007). Primers were barcoded using 8 nucleotides that indexes the specific aliquot being amplified (Binladen, Gilbert *et al.*, 2007; Hoffmann, Minkah *et al.*, 2007; Hamady, Walker *et al.*, 2008). Gel purified amplicons were quantified, pooled in equal amounts and the amplicon pool was sequenced using Roche's GS FLEX sequencer.

Subsequent analysis was carried out using the QIIME software (Caporaso, Kuczynski *et al.*, 2010). Raw sequences were decoded yielding a total of 473,169 sequence reads of average length 260 bases with correct bar codes and primer sequences were obtained for the 57 samples analyzed (Appendix 1). Taxonomic assignments were performed within QIIME using RDP classifier(Wang, Garrity *et al.*, 2007). The 16S rDNA sequences obtained were aligned using PyNast and a phylogenetic tree was created using the FastTree2 program (Price, Dehal *et al.*, 2009). Pairwise distances were created using weighted and unweighted Unifrac (Lozupone e Knight, 2005; Lozupone, Hamady *et al.*, 2007) and used in further analysis described below.

2.3.4 Bacterial taxa detected

Taxonomic assignments are shown as a heatmap in Figure 1 labeled with their phylum and genus classifications. Two aliquots taken 1cm apart from each subject and processed in the same way (henceforth referred as the Gold Standards) are shown. Taxa represented as at least 1% of the total sequences recovered are well represented in the

replicate extractions, indicating good reproducibility of the methods and demonstrating that a single 200 mg aliquot from a stool sample is a good representative of the sample as a whole. Low abundance taxa, as expected, were detected more sporadically. Although the overall proportion of each genus was the same, exact values were not exactly the same and statistically significantly differences are shown as asterisks in Figure 1.

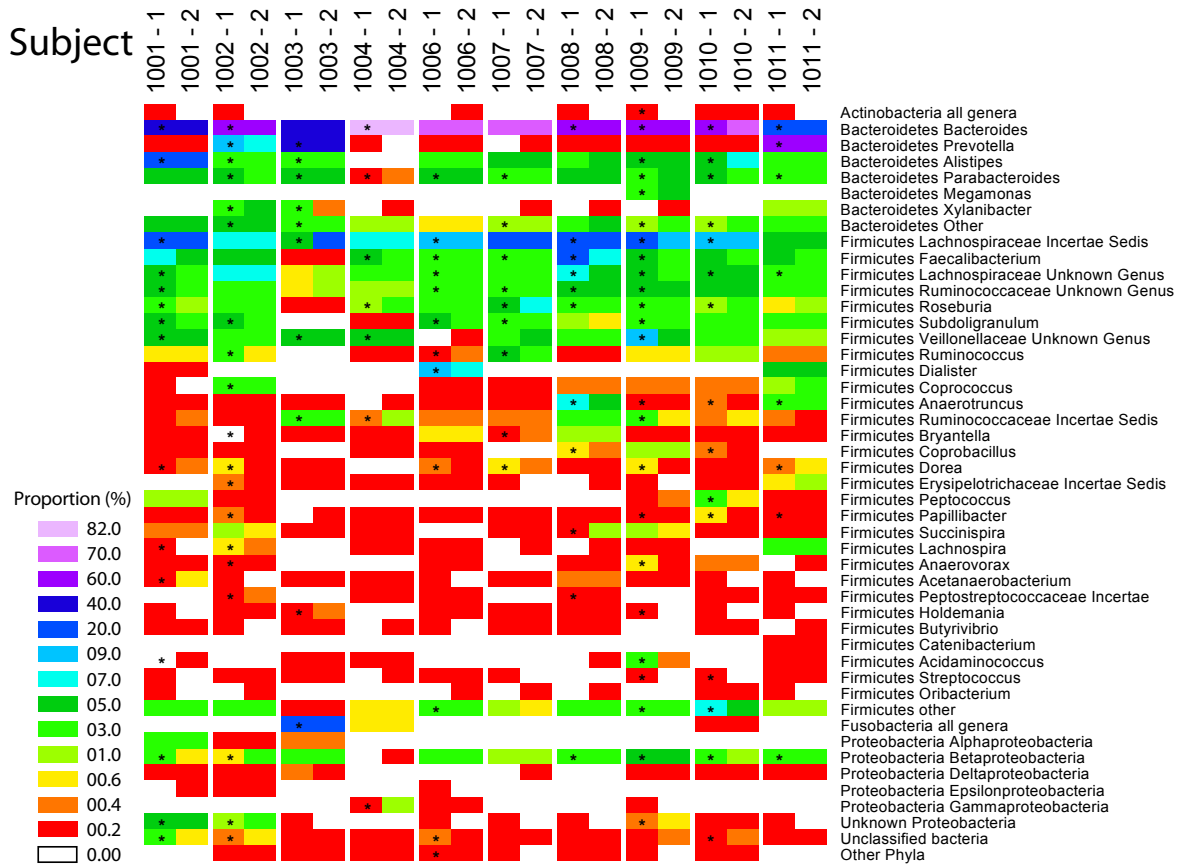


Figure 2-1: Composition of the gut microbiome in the ten subjects studied. Bacterial taxonomic assignments are indicated to the right of the heat map at the Phylum and Genus level except in cases where small numbers were detected (e. g. Proteobacteria), in which case taxa are summarized at higher levels. The relative abundance of each bacterial group is color coded as indicated by the key on the left (the number beside each colored tile indicates the lower bound for the indicated interval). Two samples were compared for each stool specimen, sampled on cm a part but otherwise worked up identically (conditions 1 and 2 in Table 3). The numbers of reads for the two samples from each subject were compared for significant differences using Fisher's exact test. The * indicates $P < 0.05$. Note that because each sequence read is treated as an individual measurement, the sample size is very large, with the result that many taxa with only modest differences nevertheless achieve significance.

Intersubject variability was evident, and subject 1003, for example, as the only one showing detectable levels of *Fusobacteria*. Overall, samples were dominated by taxa within the *Bacteriodetes* and *Firmicute* phyla, with a lower amount of *Proteobacteria* followed by

other Phyla. This results are in agreement with previously reported data available (Eckburg, Bik *et al.*, 2005; Gill, Pop *et al.*, 2006; Palmer, Bik *et al.*, 2006).

2.3.5 Bacterial taxa recovered using the different storage and DNA isolation procedures

Bacteria taxa detected using the different extraction and storage methods tested (Table 3) is shown in Figure 2. Samples were pooled within each treatment group with example of the gold standards to mark variability across aliquots. Storage methods yielded similar results for taxa recovery (Figure 2C), with exception of the samples stored in the PSP buffer for 48 hours, which presented an increased amount of *Firmicutes* (Figure 2D). The same trend in recovery bias was also observed to a greater extend when phenol bead beating was used (Figure 2A). Fewer differences were observed between the Qiagen and MoBio extraction kits, and although some *Firmicutes* presented a trend towards an increased relative recovery, not enough samples were available in this experimental comparison for statistical testing.

Table 2.3: Sampling methods compared in this study.

Method Identifier	Storage Method	DNA Purification Method	days at -80C	
			Min	Max
1	Immediately frozen (-80°C)	Qiagen Stool	2	14
2	Immediately frozen (-80°C, sampled 1 cm from sample 1)	Qiagen Stool	6	63
3	Immediately frozen (-80°C)	MoBio PowerSoil	58	72
4	4C for 24 h, then frozen (-80°C)	Qiagen Stool	1	21

5	4C for 48 h, then frozen (-80°C)	Qiagen Stool	0	12
6	PSP for 48 h, then frozen (-80°C)	PSP	0	12
7	Immediately frozen (-80°C)	Qiagen Stool (70°C)	7	7
8	Fresh	Qiagen Stool	0	0
9	Immediately frozen (-80°C)	Hot phenol with bead beating	118	137

Overall, the storage methods tested showed little effects on the taxa recovered, but an increased relative proportion of *Firmicutes* was observed using either hot phenol bead beating or the PSP kit.

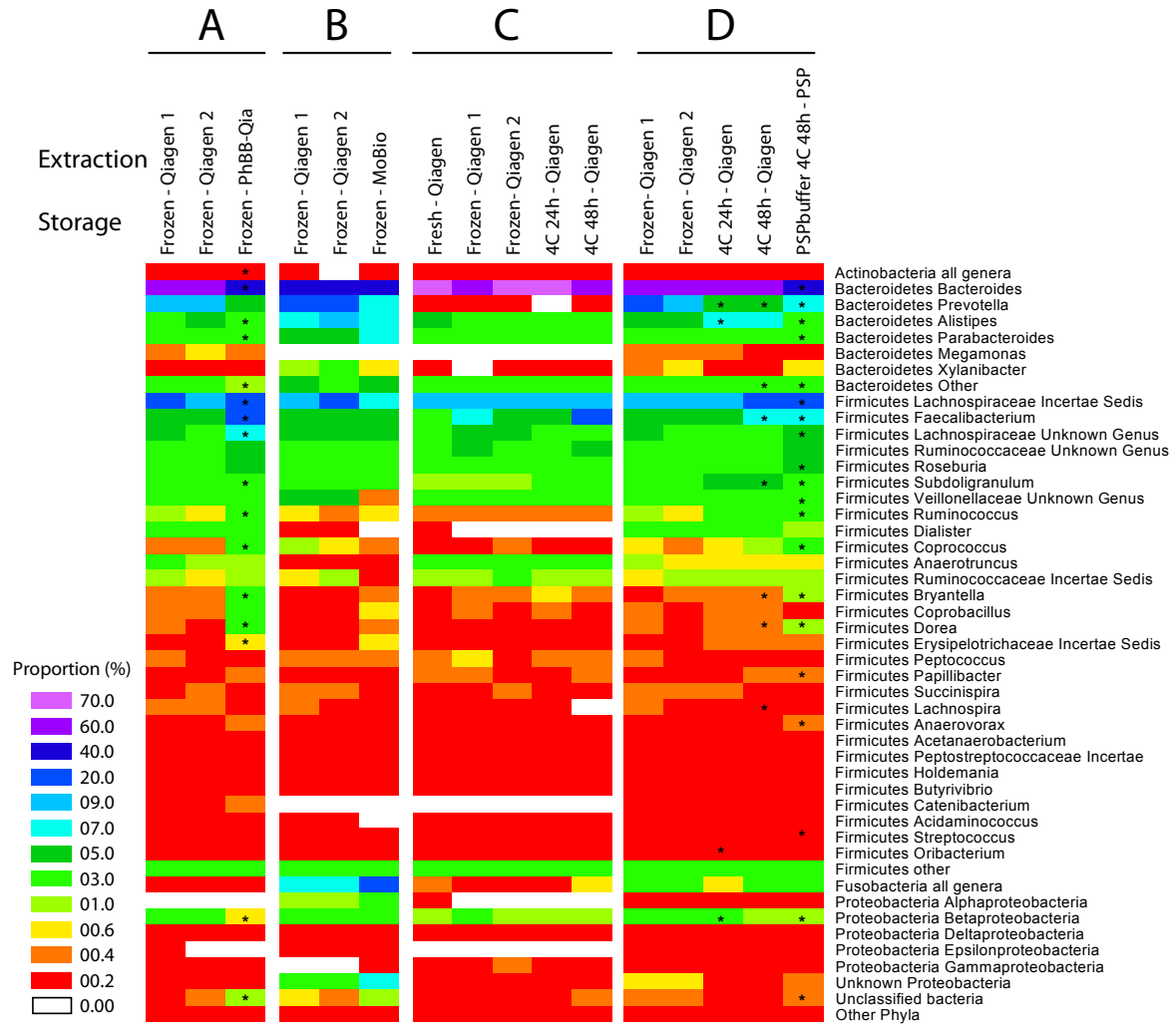


Figure 2-2. Identification of different bacteria using different stool storage and DNA isolation methods. 473,169 sequence reads were used to characterize the 57 communities analyzed. All subjects tested for each method were pooled for comparison (summarized in Additional File 1). Methods are numbered at the top of the heat map. For the heat map scale, the number beside each colored tile indicates the lower bound for the indicated interval. Taxa are mostly indicated at the genus level; rare taxa are pooled. A) Comparison of DNA isolation using the Qiagen stool kit (methods 1 and 2) to lysis by bead-beating in hot phenol (method 9). Six subjects were compared. B) Comparison of the Qiagen stool kit samples (methods 1 and 2) to the MoBio Powersoil kit (method 3). Three subjects were compared. C) Comparison of methods for storage of stool specimens. DNA was prepared from fresh samples (method 8), samples stored frozen at -80 for several days (methods 1 and 2), or samples stored at 4°C for 24 hr (method 4) or 48 hr (method 5). Three subjects were compared. D) Comparison of stool storage in PSP (method 6) to storage methods 1, 2, 4 and 5. All 10 subjects were compared. For A) and D), the methods were compared using the Wilcoxon signed rank test to identify bacterial groups that significantly changed in proportion. (* indicates $P < 0.05$). Numbers of samples were too low in B) and C) for statistical testing.

2.3.6 UniFrac cluster analysis

The high interpersonal variability in the human microbiome is a well known. However, little is known on whether sampling and storage methods of the intestinal microbiota affects this perceived variability. Here, we used Unifrac to investigate whether the methods tested had any effect on this variation between subjects. Principal coordinate

analysis was used to visualize the relationships captured by the Unifrac distances (Figures 3 and 4).

A Permanova test was used to determine if the distance between the gold standard samples and all other samples within each subject were significantly different. Unweighted Unifrac distances were not significantly different ($p=0.085$) indicating that the two gold standards were not distinct with regard to species composition. Moreover, Permanova using the weighted Unifrac distances was not significant ($p=0.197$), indicating that distances between the gold standards were as big as the distances between themselves and the other methods tested, indicating that the relative abundance of species detected were not biased by the methods tested. Figure 3 is color coded using subject ID's (Figure 3A), storage method (Figure 3B) and extraction method tested (Figure 3C), emphasizing that inter-subject variability was greater than the variability introduced due to the methods studied. The significance for the clustering observed was tested using a label permutation procedure (10000 permutations), asking if distance within the cluster were bigger than the distance between clusters. The cluster types tested were: subject, extraction and storage. While clustering by subject was highly significant ($p<0.0001$), clustering the data by storage and extraction were not significant ($p=0.16$ and $p=0.98$ respectively), indicating that the presence of absence of taxa detected is dominated by the intrinsic differences between individuals.

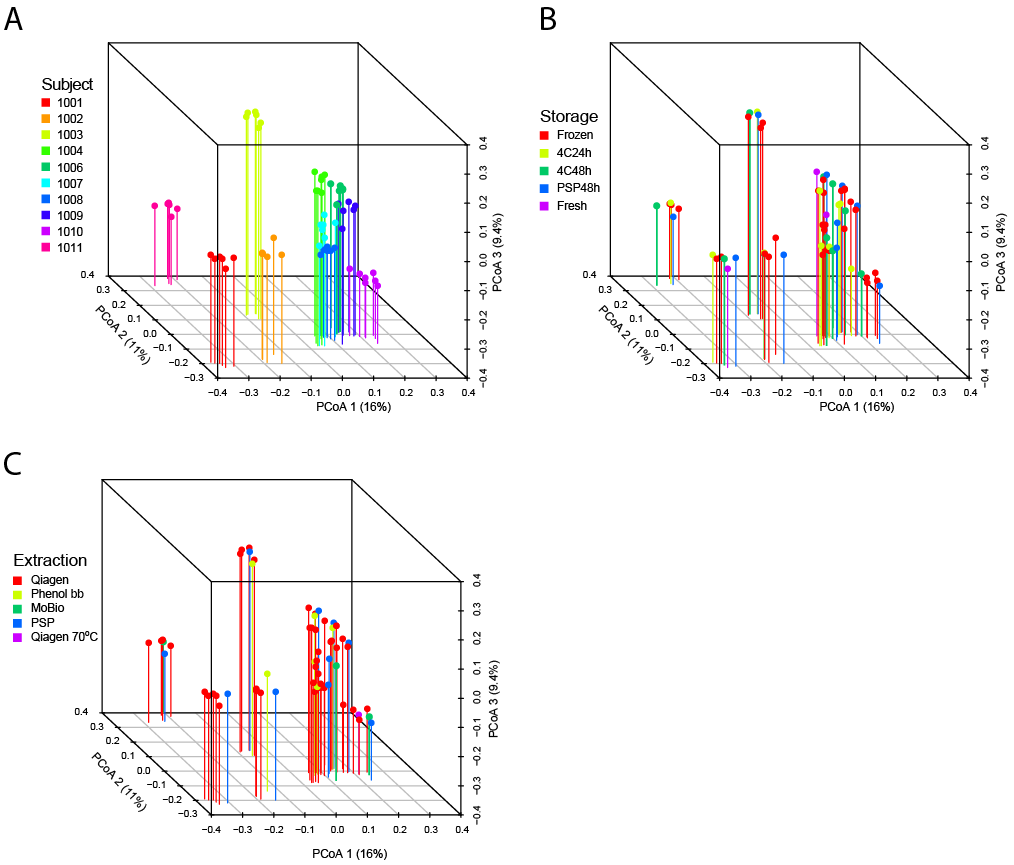


Figure 2-3. Effects of different storage conditions or DNA isolation methods on the presence or absence of bacterial taxa**.**Unweighted UniFrac was used to generate a matrix of pairwise distances between communities based on the presence/absence pattern of bacterial lineages detected. A scatterplot was generated from the matrix of distances using Principal Coordinate Analysis. The same scatterplot is shown in A)-C), but colored by subject A), storage method B), or extraction method C). The P-values cited in the text were generated using distances from the original UniFrac matrix. Percent of variation explained in the data represented by each axis, according to the Principal Coordinate Analysis, is shown in parenthesis.

Results using the weighted Unifrac distances is shown in Figure 4 and the same label permutation test applied to unweighted distances was applied. Again, distances between subjects were statistically significant when compared to distances within subjects (P<0.0001), but now differences between extraction methods were also statistically significant (P=0.001). No statistical significant difference was observed for the storage method used. This demonstrates that not only interindividual variability accounts for differences in taxa detected, but also that the extraction method used influences the relative proportion of taxa recovered.

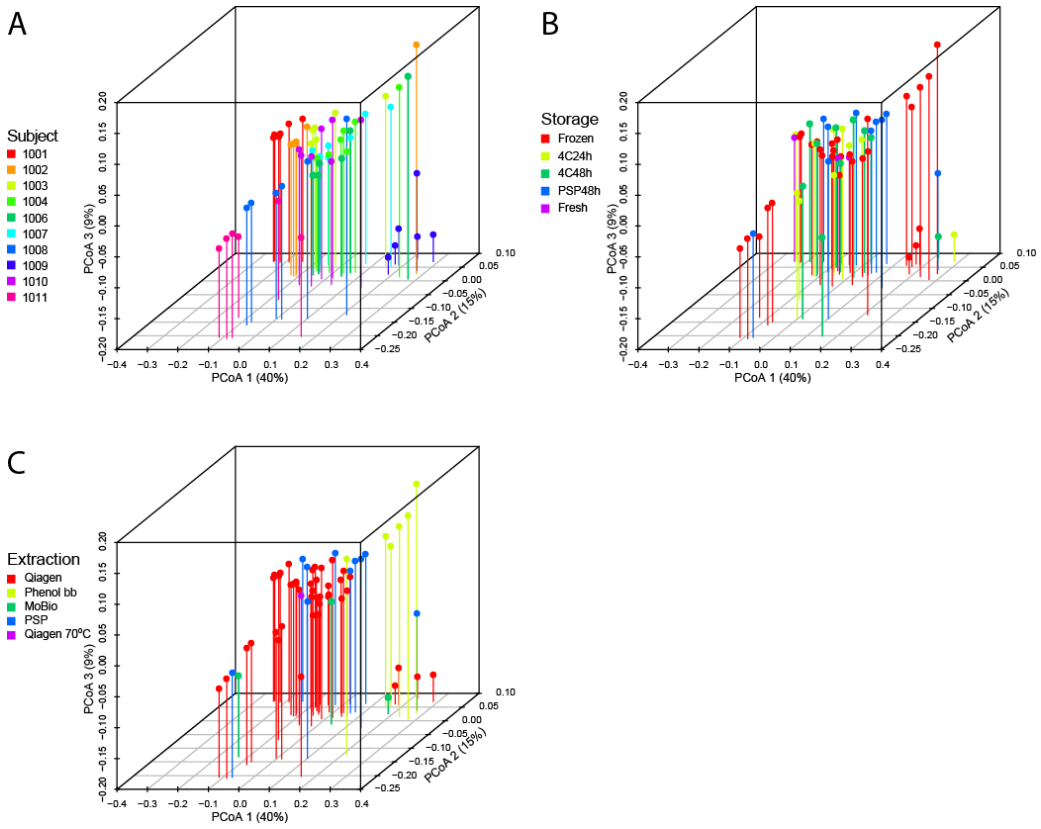


Figure 2-4. Effects of different storage conditions or DNA isolation methods on the relative abundance of bacterial taxa**.**Weighted UniFrac was used to generate a matrix of pairwise distances between communities, then a scatterplot was generated from the matrix of distances using Principal Coordinate Analysis, based on the relative proportion of the bacterial taxa detected. The same scatterplot is shown in A)-C), but colored by subject A), storage method B), or extraction method C). The P-values cited in the text were generated using distances from the original UniFrac matrix. Percent of variation explained in the data represented by each axis, according to the Principal Coordinate Analysis, is shown in parenthesis.

A further test to detect whether differences between specific extraction methods was also carried out using a label permutation test (10000 permutations). Unweighted Unifrac distances were not significantly different between extraction methods tested. Weighted Unifrac distances, however, were significantly different for hot phenol bead beating (p=0.041) and for the Qiagen method (p=0.0014). We attributed the strong effect detected for the Qiagen method to two factors: the relatively large sample size, as most samples were extracted with this method, and the fact that the "other" group contained the phenol bead-beating and the PSP kits. Comparison of the two gold standards versus each other method was carried out using Permanova and both the PSP and hot phenol bead beating methods were significantly distinct (for both, p=0.001). The individual taxa detected

for these two methods were further investigated indicating a higher recovery of *Firmicutes* and concomitant decreased in *Bacteroidetes,* potentially due to increased lysis of *Firmicutes*.

2.3.7 Comparison sequence length obtained and the effect of 16S rRNA gene region sequenced

Figure 5 shows the results we obtained comparing the microbiota of two subjects using sequences obtained from different 16S rDNA windows. A shorter V1V2 amplicon region was compared to longer V1V3, V3V5, and V6V9 regions within the 16S rDNA. Two primer pair were tested for the longer amplicons (V1V3, V3V5, and V6V9). A total of 295946 sequence reads were analyzed in these comparisons.

Communities detected using the V1V2, V1V3 and V3V5 regions yielded similar results while communites detected using the V6V9 region showed the most distinct taxomic signatures. (figure 5A). Sequences obtained with the V6V9 region were also the least well classified, with far fewer sequences classified down to genus level. The V6V9 amplicon used was sequenced from the V9 region and therefore had no good coverage of the V6 region (Figure 5C). The data suggests that while taxonomic analysis using sequences obtained from V1V2, V1V3 or V3V5 amplicons is likely to be reasonably consistent, care should be taken when combining data from the obtained from the V6V9 region (especially towards the 3' end of the region).

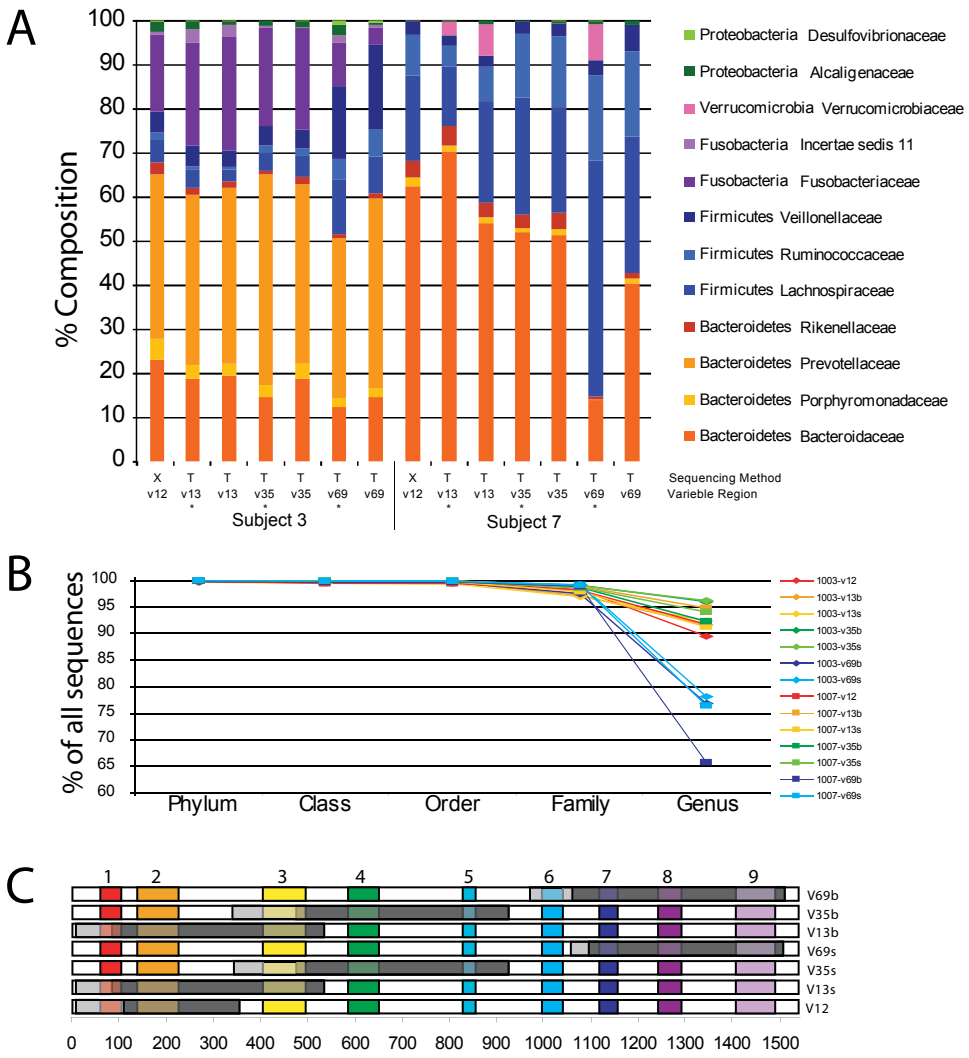


Figure 2-5: Analysis of community composition using different recovery and sequencing strategies. A) Results of analysis of Subjects 3 and 7 are shown comparing sequencing using 454/Roche GS FLX versus Titanium, and use of different variable region primers. The percentages of different bacterial families are compared in bar graphs. "Sequencing Method" indicates GS FLX ("X") or Titanium ("T"). The genera present are indicated in the key beside the graphs. "Variable Region" indicates the 16S rRNA gene region amplified by each primer set. The * indicates slightly different versions of the primers used. B) Percentages of sequences assigned for each primer set as a function of taxonomic level. C) Summary of regions amplified and regions sequenced for each primer set. Gray indicates the regions amplified, dark gray indicates the regions sequenced, light gray indicates regions amplified but not sequenced.

2.3.8 Comparison of mock community of 10 cloned 16S rRNA gene sequences

A central question in using any sequencing method is whether the data actually reflect the original community composition, or are a reflection of biases introduced during PCR and sequencing (Polz e Cavanaugh, 1998). To address this question, we created two mock communities composed of plasmid DNA containing near full length 16S genes from 10 distinct species. One DNA mixture was made with even amounts of each 16S gene, and the

second mixture was staggered to represent different bacterial abundances. PCR amplification of the V1V2 16S region was carried out for the mixtures and the staggered mixture was amplified using two enzyme systems (Amplitaq, from Applied Biosystems, and GreenTaq, from Promega). A total of 26161 sequence reads was obtained for this analysis.

Sequences were recovered in approximately the same proportions as present in the template DNA mixture (Figure 6). The distinct polymerase systems tested had no major effects on the recovered sequences. The results indicate that results obtained from pyrosequencing procedures are able to provide relative 16S gene proportions similar to those present in the input DNA.

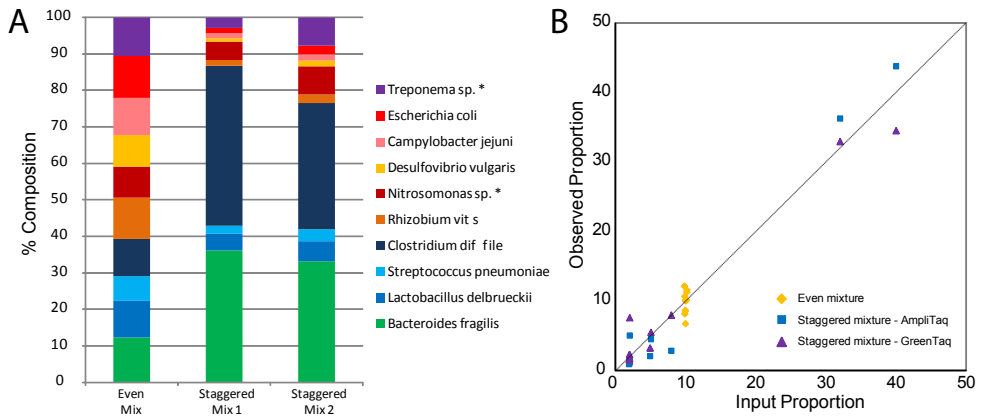


Figure 2-6: Analysis of recovery efficiency after pyrosequencing of a cloned DNA mock community. A) Bar graph illustrating proportional recovery of 16S rRNA gene pyrosequence reads from a plasmid DNA mock community. Each of the 10 templates consisted of a bacterial 16S rRNA gene sequence cloned in a bacterial plasmid. "Even mix" indicates that the same copy number for each of the 10 templates was used in the amplification reaction. "Staggered mix" indicates different amounts. The "Staggered mix 2" sample was amplified with a different polymerase mixture (Promega's GreenTaq Master Mix, Madison, WI) instead of AmpliTaq which was used in all other experiments, revealing that the two mixtures yielded similar results. The taxonomic assignments for the sequences obtained is shown on the right. B) Scatter plot comparing the theoretical proportion of each input sequences (x-axis) to the proportions inferred from 454 GS FLX sequence data (y-axis).

2.4 Discussion

One of the main goals when analyzing the human microbiome is establishing links between microbiota patterns observed and healthy or disease states (Eckburg, Bik *et al.*, 2005; Gill, Pop *et al.*, 2006; Dethlefsen, Huse *et al.*, 2008; Costello, Lauber *et al.*, 2009; Turnbaugh, Hamady *et al.*, 2009). Large surveys are being reported in the literature,

nevertheless, no standard method to perform such studies exist, and any possible biases introduced by the choice of method made are largely unknown. Here, we have performed a series of experiments to determine optimal methods for the study of the human gut microbiome, namely: 1) the effects of sample storage prior to analysis, 2) the effects of the DNA extraction method used, 3) the effects of the choice of 16S region being analyzed, and 4) the efficiency of recovery of distinct 16S sequences from mixed DNA populations. The results should provide guidelines for future gut microbiome studies, allowing us to make recommendation on the best choice of methods to use.

We obtained stool specimens from 10 individuals, from which aliquots were made and stored under different conditions prior to DNA extraction. Aliquots frozen immediately after sampling were considered our gold standard, to which all other comparisons were made. Slight differences were observed between the 2 replicate gold standards taken 1 cm apart. Such variation may indicate changes introduces during sample handling, but are more likely representative of slight variation in the composition of the whole sample. Storage of samples at $4^{o}C$ for up to 48 hours had little effect on the intrinsic composition of the samples analyzed. The PSP kit yielded a greater amount of DNA extracted from the samples, as compared to other methods studied, and have the added advantage of allowing samples to be stored at room temperature. No difference observed was statistically significant between storage methods tested, indicating that the choice of storage method can be dictated by convenience of sample collection.

The DNA extraction, however, did introduce bias on the bacterial communities being studied. This effect was observed when the relative proportion of organisms was taken under consideration, as measured by weighted Unifrac, and not when their presence or

absence as considered. This indicated that surveys concerned only with unweighted analysis are relatively unhindered by the choice of the DNA extracion used, while greater care needs to be taken when the abundance data is also desired. This bias was most strongly correlated with an increase detection of members of the *Firmicutes* phylum. This bacterial phylum is composed of Gram-positive taxa, which are usually harder to lyse (*Firmus* - strong, *cutis* - skin). Both the PSP method, and to a much greater extent, the hot phenol bead beating method, improved the recovery of *Firmicutes*. Although using hot phenol has several disadvantages (e.g. health hazards, toxic waste production), the use of the PSP kit, potentially with the addition of mechanical lysis by bead beating, as well as increased temperature incubations, may improve the recovery of harder to lyse taxa, and provide a more complete picture of intestinal microbiota samples.

The choice of 16S window used to perform microbial surveys was also tested. Previous literature has reported that the study of microbial communities by PCR amplification and sequencing of the of the 16S gene can be biased (Polz e Cavanaugh, 1998). Here, we have compared the use of 3 different amplicons targeting distinct 16S rDNA regions to evaluate their usefulness in stool microbiome surveys. The choice of amplicon did introduce bias on the results obtained. This was most notable for the V6V9 region of the 16S rRNA, however, our sample size was much more reduced than the one used for the storage methods testing. Nevertheless, the results indicate a trend for distinct patterns observed when the 3' end of the 16S gene is analyzed. Notably, there was reduced level of sequence classified down to lower level taxonomy, especially genus level. Care should be taken when conducting meta-analysis or combining data from the literature.

The results for the mock community experiment indicate that only a very small bias was introduced in the recovered sequences from PCR and sequencing. This supports the idea that PCR and DNA sequencing are good methods for general microbiological surveys. Our mock DNA community was limited in the range of relative DNA proportions and higher levels of bias may be detected using real samples, which have a far higher alpha diversity and large range of relative species proportions. Nevertheless, the results support the idea that DNA extraction methods, and not the PCR or Sequencing are the greatest sources of bias in this type of analysis.

2.5 Conclusions

With these results, we can make the following recommendations: 1) the storage of stool samples for microbiome studies can be chosen as convenience dictates, as little effects were detected across the storage methods tested; 2) the method for DNA extraction does have a strong effect on the results obtained, and the samples being compared should all be analyzed with the same extraction method. The preferred method is the PSP kit, potentially including a harsher bead beating step and a higher temperature incubation. In cases where it is important to correctly detect *Firmicutes* or other hard to lyse bacteria, a hot phenol bead beating method should be considered; 3) Analysis of multiple datasets must account of the choice of primers used, if possible avoiding the use the V6-V9 regions as it may yield less taxonomic information; 4) differences across subjects was the most pronounced source of variation in the data, and therefore larger sample sizes are likely to be needed to demonstrate the effect of other factors (e.g. disease state, diet, drug use) on the gut microbiome.

2.6 References

BINLADEN, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, v. 2, p. e197, Feb 14 2007.

CAPORASO, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, v. 26, n. 2, p. 266-7, Jan 15 2010.

CAPORASO, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, v. 7, n. 5, p. 335-6, May 2010.

COSTELLO, E. K. *et al.* Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science (New York, N.Y.)*, Nov 5 2009.

DETHLEFSEN, L. *et al.* The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS biology*, v. 6, n. 11, p. e280, Nov 18 2008.

ECKBURG, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science*, v. 308, n. 5728, p. 1635-8, Jun 10 2005.

GILL, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science (New York, N.Y.)*, v. 312, n. 5778, p. 1355-1359, Jun 2 2006.

HAMADY, M. *et al.* Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature methods*, v. 5, n. 3, p. 235-237, Mar 2008.

HOFFMANN, C. *et al.* DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic acids research*, v. 35, n. 13, p. e91, 2007.

LEWIS, S. J.; HEATON, K. W. Stool form scale as a useful guide to intestinal transit time. *Scandinavian journal of gastroenterology*, v. 32, n. 9, p. 920-924, Sep 1997.

LEY, R. E. *et al.* Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, v. 102, n. 31, p. 11070-5, Aug 2 2005.

LIU, Z. *et al.* Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic acids research*, Sep 18 2007.

LOZUPONE, C. *et al.* UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC bioinformatics*, v. 7, p. 371, Aug 7 2006.

LOZUPONE, C.; KNIGHT, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*, v. 71, n. 12, p. 8228-35, Dec 2005.

LOZUPONE, C. A. *et al.* The convergence of carbohydrate active gene repertoires in human gut microbes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 39, p. 15076-15081, Sep 30 2008.

LOZUPONE, C. A. *et al.* Quantitative and Qualitative {beta} Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, v. 73, n. 5, p. 1576-1585, Mar 2007.

LUDWIG, W. *et al.* ARB: a software environment for sequence data. *Nucleic acids research*, v. 32, n. 4, p. 1363-1371, Feb 25 2004.

MCKENNA, P. *et al.* The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS pathogens*, v. 4, n. 2, p. e20, Feb 8 2008.

PALMER, C. *et al.* Rapid quantitative profiling of complex microbial populations. *Nuc. Acids Res.*, v. 10, p. e5, 2006.

POLZ, M. F.; CAVANAUGH, C. M. Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, v. 64, n. 10, p. 3724-3730, Oct 1998.

PRICE, M. N. *et al.* FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*, v. 26, n. 7, p. 1641-1650, Jul 2009.

TURNBAUGH, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature*, v. 457, n. 7228, p. 480-4, Jan 22 2009.

WANG, Q. *et al.* Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, v. 73, n. 16, p. 5261-5267, Aug 2007.

CHAPTER 3. ARCHAEA AND FUNGI OF THE HUMAN GUT MICROBIOME: CORRELATIONS WITH BACTERIAL RESIDENTS

The contents of this chapter have been partially submitted for publication as:

> **Hoffmann C**, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, and Bushman FD. (2013) Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE.* 8(6): e66019. doi:10.1371/journal.pone.0066019**.**

3.1 Introduction

Humans live in association with immense populations of bacteria, viruses, fungi and archaea (Minot, Sinha *et al.*, 2011; Faust, Sathirapongsasuti *et al.*, 2012; Gevers, Knight *et al.*, 2012; Minot, Grunberg *et al.*, 2012; Minot, Wu *et al.*, 2012; Segata, Haake *et al.*, 2012; 2012; Ursell, Clemente *et al.*, 2012). Many groups have now contributed surveys using deep sequencing to characterize these populations, revealing that the human microbiome differs radically at different body sites, and between individuals (Gill, Pop *et al.*, 2006; Grice, Kong *et al.*, 2009; Kuczynski, Costello *et al.*, 2010; Wu, Chen *et al.*, 2011). Differences in body sites are caused at least in part by the availability of nutrients, water, oxygen, and other site-specific features, but the origin of differences between individuals are less clear. Such differences likely reflect, at least in part, distinct colonization early in life and different environmental exposures such as antibiotic use (Dethlefsen, Huse *et al.*, 2008; Koenig, Spor *et al.*, 2011; Costello, Stagaman *et al.*, 2012).

Bacteria are abundant members of the gut microbiome, but not the only residents. Bacteriophage particles within the intestinal tract are present in potentially greater numbers than bacterial cells (Reyes, Haynes *et al.*, 2010; Reyes, Semenkovich *et al.*, 2012). Recently

changes in bacteriophage communities in gut have been correlated with dietary interventions (Minot, Sinha *et al.*, 2011).

Archaea are also present in human gut, the most frequently occurring of which is *Methanobrevibacter smithii* (Miller, Wolin *et al.*, 1982; Eckburg, Bik *et al.*, 2005; Dridi, Raoult *et al.*, 2011; Matarazzo, Ribeiro *et al.*, 2012), which produces methane from byproducts of bacterial fermentation (Matarazzo, Ribeiro *et al.*, 2012). Reported colonization rates by methanogenic archaea range from 25% to 95% of humans (Stewart, Chadwick *et al.*, 2006; Dridi, Henry *et al.*, 2009). In microbial ecosystems such as the human gut, when H_2 accumulates due to bacterial catabolism, archaeal growth is stimulated (Matarazzo, Ribeiro *et al.*, 2012). Support for such syntrophy in the mammalian gut has been shown in a gnotobiotic mouse model, where co-colonization by *M. smithii* and *Bacteroides thetaiotamicron* promoted increased growth of both species compared to mono-colonization (Samuel, Hansen *et al.*, 2007).

Yeasts have been detected in human stool samples at least since 1917 (Anderson, 1917), and by the mid 20th century their presence in the human intestine had been proposed to have a saprotrophic role (Gumbo, Isada *et al.*, 1999). Gut fungi may also be involved in pathogenic processes. Anti-*Saccharomyces* antibodies are detected in inflammatory bowel disease cohorts and are used as a predictor of disease progression (Seibold, Stich *et al.*, 2001; Murdoch, Xu *et al.*, 2012). Recent work using a murine model has suggested that normally mutualistic or commensal fungi species may exacerbate intestinal inflammation in mice with sensitized genotypes (Iliev, Funari *et al.*, 2012). In mice, over 14 fungal genera have been reported to be present within the mucus layer lining the intestinal epithelium (Scupham, Presley *et al.*, 2006). Available data is likely incomplete, because of

reliance mostly on culture-based methods. Recent reports using next generation sequencing also suggest diverse fungal communities in humans (Hamad, Sokhna *et al.*, 2012),(Ott, Kuhbacher *et al.*, 2008; Dollive, Peterfreund *et al.*, 2012).

Here we investigate associations between the bacteria, the fungi and the archaea of the human gut in a cohort of 96 healthy individuals who were previously characterized for their bacteria/diet relationships (Wu, Chen *et al.*, 2011). Fungi were characterized by sequencing the Internal Transcribed Spacer region 1 (ITS1) of the rRNA locus and the archaea by sequencing a segment of the 16S rRNA gene. Analysis showed notable correlations of the three domains of life with each other, thus these data begin to specify potential multi-domain syntrophic assemblages in the human gut microbiota.

3.2 Methods

3.2.1 Samples

The sample set described in (Wu, Chen *et al.*, 2011) was used in this study. The bacterial 16S sequences data published therein were used in this analysis. Briefly, healthy volunteers were screened to be free from any chronic gastrointestinal disease, cardiac disease, diabetes mellitus or immunodeficiency diseases, to have a normal bowel frequency (minimum once every 2 days, maximum 3 times per day), and body mass index (BMI) between 18.5 and 35. All volunteers were on *ad lib* det. One stool sample was provided per subject and kept frozen at -80^{o}C until processed for DNA extraction (Wu, Chen *et al.*, 2011).

3.2.2 16S rDNA gene, ITS1 region and AmoA gene PCR

Pyrosequencing was carried out using barcoded composite primers constructed as described in (Mckenna, Hoffmann *et al.*, 2008). PCR reactions were set in triplicate using the Accuprime system (Invitrogen, Carlsbad, CA, USA), and each reaction contained 50

nanograms of DNA and 10 picoMol of each primer. Archaeal specific 16S rDNA primers and ITS1 fungal primers were adapted from the literature and used to amplify a rDNA 16S fragment; the final PCR cycling conditions were empirically optimized to maximize specificity (Tables 1 and 2) (Delong, 1992; Lepp, Brinig *et al.*, 2004; Ghannoum, Jurevic *et al.*, 2010).

A nested PCR for using specific primers for the AmoA gene was used to confirm the detection of *Nitrosophaera* sequences (Tourna, Stieglmeier *et al.*, 2011). PCR reactions were set using Invitrogen's Accuprime system and 1 μL of the total extracted DNA was used as template for the initial PCR. The nested PCR used 1 μL of PCR1 as template (Table 2). Blank extractions were used to control for environmental, reagent and consumables contamination. All PCR work was set up using a laminar flow hood and all consumables and equipment were UV irradiated for a minimum of 30 minutes prior to use.

Table 2.4: Primers used in this study.

primer	target	sequence
958aF-deg	archaea 16S	5'-AATTGGAKTCAACGCCKGR-3'
1378aR	archaea 16S	5'-TGTGTGCAAGGAGCAGGGAC-3'
ITS1Fungal	fungi ITS1	5'-CTTGGTCATTTAGAGGAAGTAA-3'
ITS2-rev	fungi ITS1	5'-GCTGCGTTCTTCATCGATGC-3'
CrenAmoA-23f	archaeal amoA	5-ATGGTCTGGCTWAGACG-3
CrenAmoA-616r	archaeal amoA	5-GCCATCCATCTGTATGTCCA-3
CrenAmoA-104f	archaeal amoA	5'-GCAGGAGACTACATMTTCTA-3'

Table 2.5: PCR amplification conditions used in this study. The "->" symbol indicates a touchdown annealing decreasing by 1oC per amplification cycles.

Archaeal amplification conditions:		
step	number of cycles	conditions
1	1	95°C / 5 min

2	12	95°C / 30sec; 67°C -> 55°C / 30 sec; 72°C / 35 sec
3	10	95°C / 30sec; 55°C / 30 sec; 72°C / 35 sec
4	5	95°C / 30sec; 53°C / 30 sec; 72°C / 35 sec
5	15	95°C / 30sec; 50°C / 30 sec; 72°C / 35 sec
6	1	72°C / 8 min
7	hold	4°C

Fungal amplification conditions:

step	number of cycles	conditions
1	1	95°C / 5 min
2	35	95°C / 45sec; 56°C / 45 sec; 72°C / 90 sec
3	1	72°C / 10 min
4	hold	4°C

AmoA PCR1 amplification conditions:

step	number of cycles	conditions
1	1	95°C / 5 min
2	10	94°C / 30sec; 55°C / 30sec; 72°C / 60sec
3	25	92°C / 30sec; 55°C / 30sec; 72°C / 40sec
4	1	72°C / 10 min
5	hold	4°C

AmoA PCR2 amplification conditions:

step	number of cycles	conditions
1	1	95°C / 5 min
2	10	94°C / 30sec; 65°C -> 55°C / 30sec; 72°C / 60sec
3	25	92°C / 30sec; 55°C / 30sec; 72°C / 40sec
4	1	72°C / 10 min
5	hold	4°C

3.2.3 Pyrosequencing

Amplified 16S rDNA and ITS1 fragments were purified using 1:1 volume of Agencourt AmPure XP beads (Beckman-Colter, Brea, CA). The purified PCR products from the stool samples were pooled in equal amounts prior to pyrosequencing using Roche/454 Genome Sequencer Junior. DNA pools were separated by amplicon type and all samples were submitted for sequencing, even if no visible amplicon was observed in agarose gels, and in this case, 40 out 50 μL of the bead-purified PCR product was pooled with the other samples for sequencing.

3.2.4 Sequence manipulation

Sequences obtained were decoded and quality controlled using the QIIME pipeline ((Caporaso, Kuczynski *et al.*, 2010)). OTU's were formed at 97% and 95% similarity for archaeal and fungal sequences respectively, and were considered for further analysis if they had a minimum of 5 sequences across all samples detected. Taxonomy was assigned to OTU representative sequences using the RDPclassifier (Cole, Wang *et al.*, 2009) for archaeal sequences and BROCC (Dollive, Peterfreund *et al.*, 2012) for fungal sequences. All taxonomy assignments were manually curated to check for accuracy and nomenclatural agreement. OTU sequence counts for each sample were aggregated at genus level. All downstream analysis was done at Genus level using the R statistical environment unless otherwise noted. Genera were considered in the analysis if present in at least 9 out of the 96 available samples, and its absolute sequence count was equals or greater than 10. All novel sequence data was deposited at NCBI's Sequence Read Archive with accession number SRP021021.

3.2.5 Beta diversity

Taxonomic information can be used as a proxy for the relationships between distinct species (Lozupone e Knight, 2008). The taxonomic information was used to obtain the Taxonomic Distance between each genus using the R package Ade4, version 1.5 (Dray e Dufour, 2007). The taxonomic distance matrix was used as input to calculate Unifrac distances using the R package GUniFrac, version 1.0 (Chen, Bittinger *et al.*, 2012).

3.2.6 Inter-generic relationships

Effects on the fungi by the bacteria present in the microbiome, effects of the fungi on the bacteria, and the effects of archaea on the bacterial and fungal fractions were investigated using a Permanova test. Within sample genus proportions were used to calculate Spearman correlations between bacterial genera and fungal genera. As only one or very few archaeal genera were detected in any sample, sequence proportions would be greatly skewed invalidating any potential correlation results. Instead, samples were classified according to the archaea genera detected and bacterial and fungal proportions were used on as input for Mann-Whitney tests. P-values were considered significant using a FDR of 25%.

3.2.7 Co-occurrence and Network reconstruction

The Dice index (Dice, 1945) was used to determine the co-occurrence of genera in the dataset. Genera were considered present in a sample if its sequence proportion was at least 0.01.

The Dice index matrix was used to construct a co-occurrence network. Genera were considered co-occurring for all Dice indexes of 0.3 or above. Modules were detected using

the Walktrap algorithm. All network calculations and visualizations were made using R package igraph, version 0.6-3.

3.3 Results

3.3.1 Sample dataset used for the analysis

A total of 98 samples were collected from healthy volunteers, and sequences from 96 of these samples were used in this analysis after quality filtering. The archaeal and fungal components of the microbiota were assessed using the 16S rRNA gene and the ITS1 rRNA gene tags, respectively (Delong, 1992; Heuer, Krsek *et al.*, 1997; Ghannoum, Jurevic *et al.*, 2010). The bacterial population of these samples was characterized previously by 454 pyrosequencing of V1V2 segments of the 16S rRNA gene (Wu, Chen *et al.*, 2011). Volunteers were screened to be free of chronic gastrointestinal disease, cardiac disease, diabetes mellitus or immunodeficiency diseases, to have a normal bowel frequency (minimum once every 2 days, maximum 3 times per day), and body mass index (BMI) between 18.5 and 35.

3.3.2 The archaea of the gut microbiome

A total of 99,131 archaeal sequence reads were obtained, resulting in the detection of 5 genera (Figure 1). A total of 44 of the 96 samples analyzed were positive for at least one species. *Methanobrevibacter sp* was detected in 30 samples, and *Nitrososphaera sp* was detected in 16 samples (Figure 1C). The two genera were usually mutually exclusive, coexisting in only 6 samples.

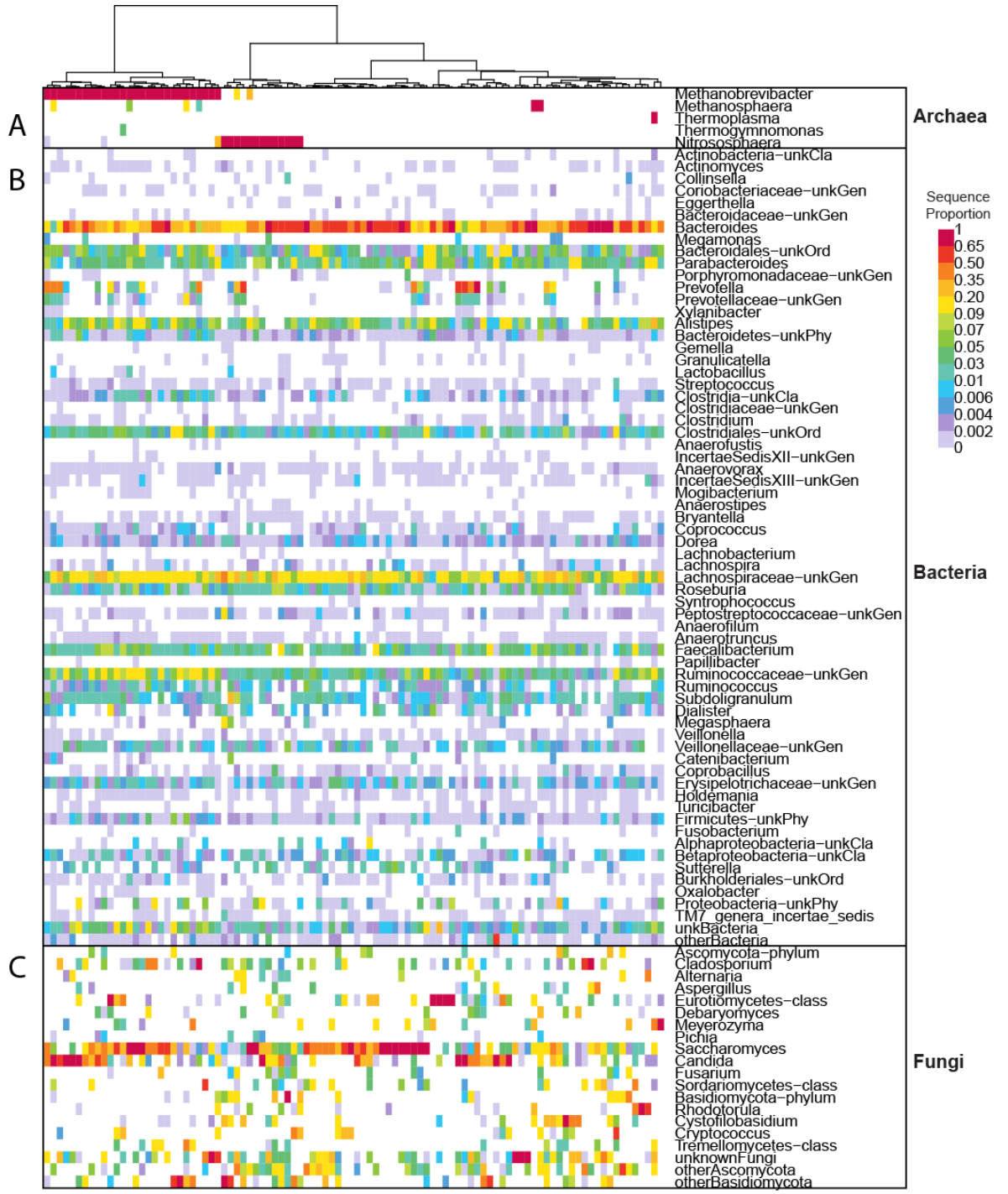


Figure 2-7: The fungal and archaeal members of the human gut microbiome. The heatmap show the relative proportions of archaeal genera (A), bacterial genera (B), and fungal genera (C). Proportions were calculated for each sequencing effort separately. Other Ascomycota and Basidiomycota are composed of genera which were detected in only one sample.

The detection of *Nitrososphaera* was surprising and so was investigated further. Comparison of amplification efficiency suggested that the *Methanobrevibacter* when present was relatively abundant, while the *Nitrososphaera* was less abundant.

Nitrosophaera was not detected after control amplifications with archaeal primers using products of blank DNA purifications as template followed by 454 pyrosequencing (no positives out of eight tested), indicating that *Nitrososphaera* DNA was not a contaminant introduced during purification.

To validate the detection of *Nitrososphaera*, we used a nested PCR assay to detect the AmoA gene, which encodes the ammonia mono-oxyenase enzyme, and is distinctive for *Nitrososphaera*. We found an association between an AmoA positive PCR and *Nitrososphaera* detection in the samples (p=0.014, Fisher's exact test). The association, while significant, was not invariant, probably because of difficulties in detection due to the low level of *Nitrosophaera* in the samples, and possible presence of AmoA in other microbes or food materials. Inspection of published work showed detection of *Nitrosophaera* in metagenomic sequences from one of two individuals studied by (Eckburg, Bik *et al.*, 2005), and in both 16S and metagenomic sequences in another cohort (Yatsunenko, Rey *et al.*, 2012). These data do not distinguish whether *Nitrosophaera* is replicating in the human gut or a transient present in food.

3.3.3 The fungi of the gut microbiome

The fungal sequencing effort yielded 332,659 sequence reads, resulting in detection of 66 genera and 13 other lineages that could not be classified to the genus level (Figure 1B). Fungal sequences were detected in every sample analyzed. Only 12 fungal genera were detected in 9 or more samples (Figure 2). The phyla Ascomycota and Basidiomycota were mostly inversely correlated (Figure 2). The most prevalent genus in this sample set was *Saccharomyces* (present in 89% of the samples), followed by *Candida* (57%) and *Cladosporium* (42%) (Figure 2).

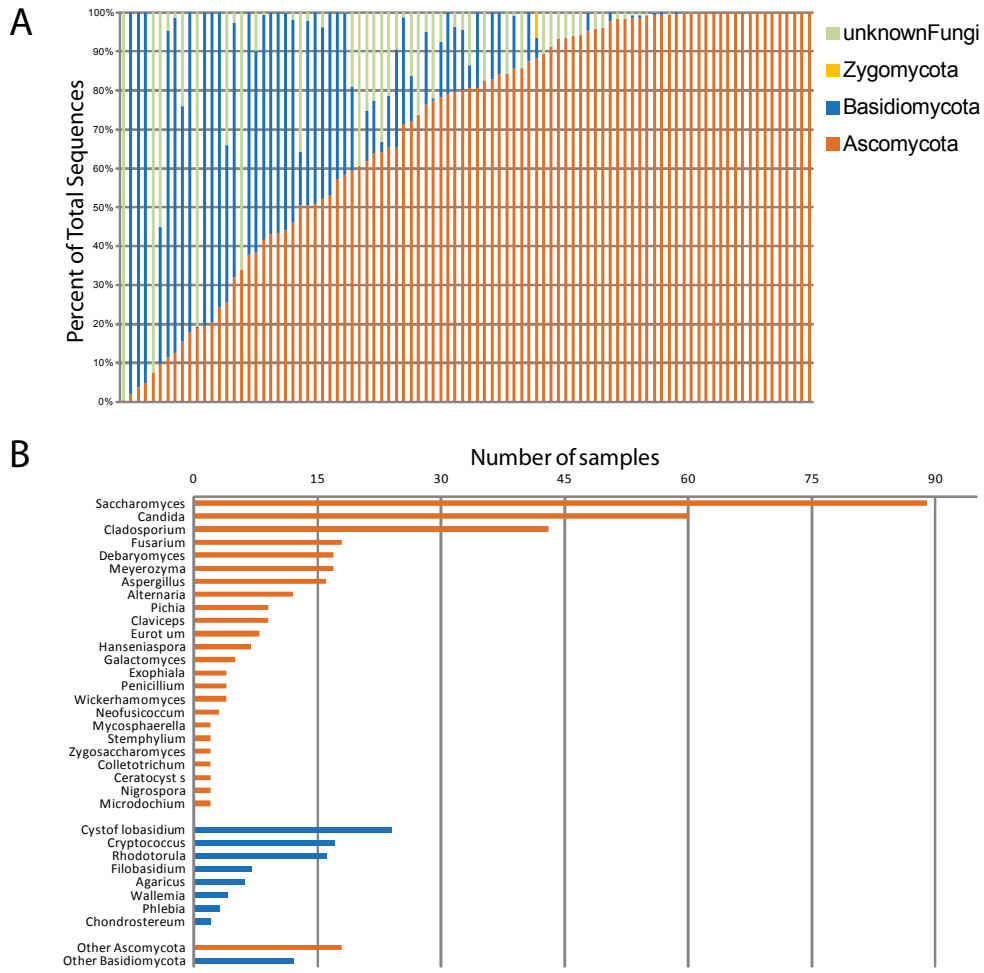


Figure 2-8. Fungi detected and their prevalence. The Fungal phyla detected are shown as sequence proportions within each sample (A). The prevalance of each fungal genera detected across all samples is depicted in (B). Genera are grouped by their phylum affiliation and only genera sequences that could be assigned to the genus level are shown.

3.3.4 Covariation among microbial lineages

We next examined covariation among the three domains, taking into account the relative abundance of each lineage in addition to presence-absence information. For the fungi and bacteria, multiple lineages were seen, and multiple different lineages were present in all samples, allowing use of correlation-based methods. However, for the archaea, only two lineages were detected with substantial frequency, *Methanobrevibacter* and *Nitrosophaera*, and these were mostly mutually exclusive. Thus the archaea were simply divided into three categories (containing either of the two archaea or no archaea), and co-occurring bacteria and fungi scored. For those few cases where both archaea were seen in a

single sample, there was a substantially greater abundance of one of the two, and the sample was assigned on that basis.

As a first global test of association among the three Domains, we conducted a Permanova test using the newly developed Generalized Unifrac distance (Chen, Bittinger *et al.*, 2012), which showed significance (Tables 3 and 4). To characterize the lineages involved, we used a non-parametric Kruskal-Wallis test to determine which bacterial and fungal genera covaried with the archaeal categories (Figure 3A).

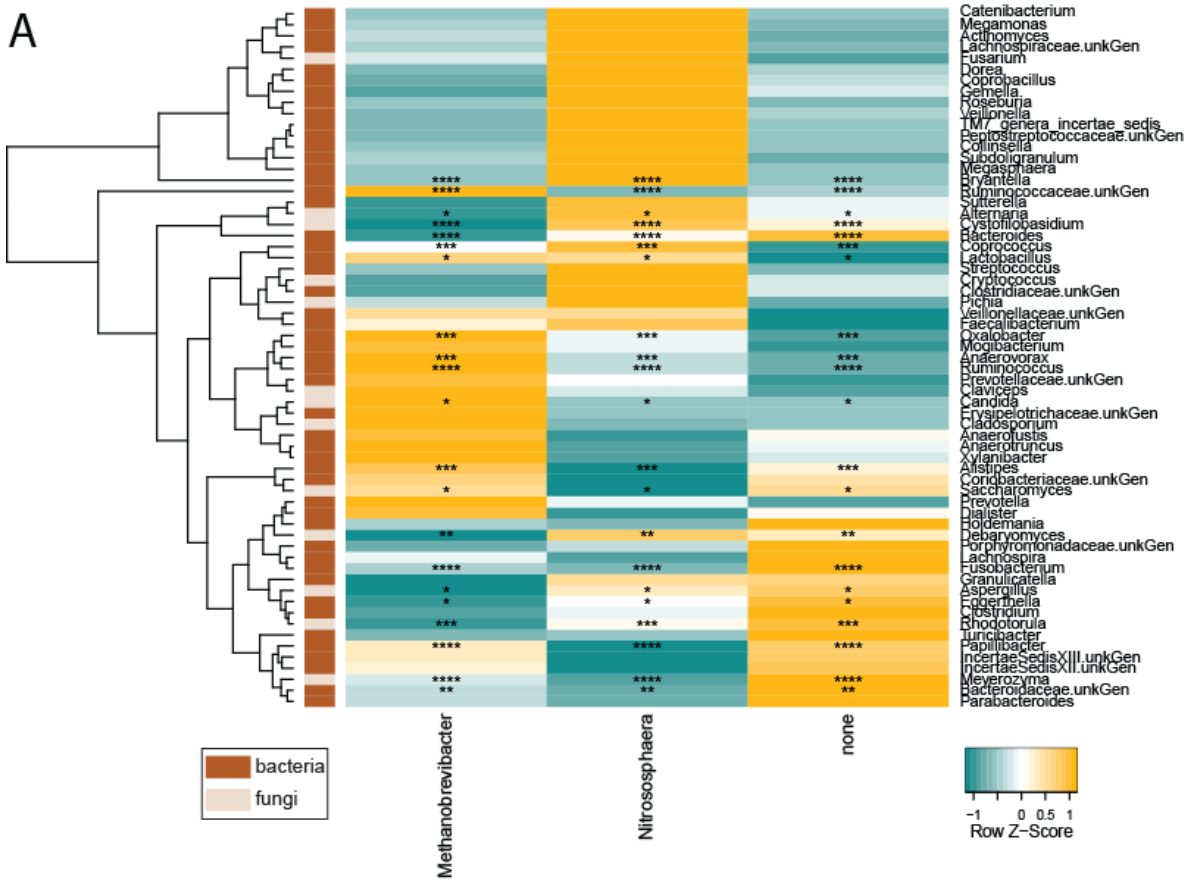
Table 2.6: Permanova of Archaea effects on the Bacterial and Fungal parts of the microbiome.

Permanova pvalues (10000 permutations)	Presence/Absence of any Archaea	Archaeal Groups
Fungal	**0.0517**	**0.001**
Bacterial	**0.0004**	**0.0015**

Table 2.7: Fungal Phyla permanova with Bacteria

	GUniFrac	unweighted	weighted
Ascomycota	0.0202	0.2948	0.0205
Basidiomycota	0.0037	0.5404	0.004
unknownFungi	0.6174	0.3934	0.9009

Several lineages co-varied with *Methanobrevibacter*, including the commonly encountered bacteria *Ruminococcus*, and rarer lineages such as *Oxalobacter* and *Papillibacter*. Of the fungi, *Candida* and *Saccharomyces* were both positively associated with *Methanobrevibacter*. Both fungal genera were negatively associated with *Nitrososphaera*.


A
Catenibacterium
Megamonas
Actinomyces
Lachnospiraceae.unkGen
Fusarium
Dorea
Coprobacillus
Gemella
Roseburia
Veillonella
TM7_genera_incertae_sedis
Peptostreptococcaceae.unkGen
Collinsella
Subdoligranulum
Megasphaera
Bryantella
Ruminococcaceae.unkGen
Sutterella
Alternaria
Cystofilobasidium
Bacteroides
Coprococcus
Lactobacillus
Streptococcus
Cryptococcus
Clostridiaceae.unkGen
Pichia
Veillonellaceae.unkGen
Faecalibacterium
Oxalobacter
Mogibacterium
Anaerovorax
Ruminococcus
Prevotellaceae.unkGen
Claviceps
Candida
Erysipelotrichaceae.unkGen
Cladosporium
Anaerofustis
Anaerotruncus
Xylanibacter
Alistipes
Coriobacteriaceae.unkGen
Saccharomyces
Prevotella
Dialister
Holdemania
Debaryomyces
Porphyromonadaceae.unkGen
Lachnospira
Fusobacterium
Granulicatella
Aspergillus
Eggerthella
Clostridium
Rhodotorula
Turicibacter
Papillibacter
IncertaeSedisXIII.unkGen
IncertaeSedisXII.unkGen
Meyerozyma
Bacteroidaceae.unkGen
Parabacteroides
bacteria
fungi
Methanobrevibacter
Nitrososphaera
none
−1 0 0.5 1
Row Z−Score

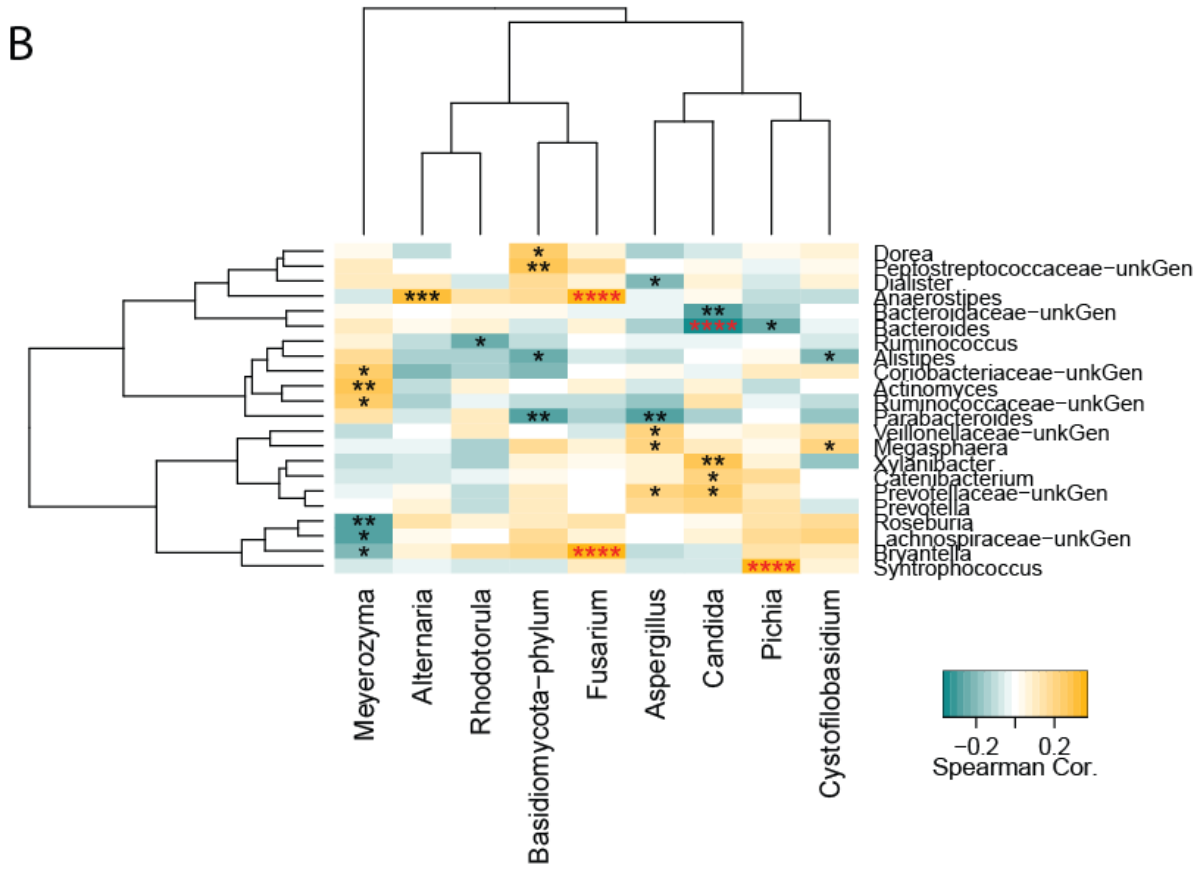

B
Dorea
Peptostreptococcaceae−unkGen
Dialister
Anaerostipes
Bacteroidaceae−unkGen
Bacteroides
Ruminococcus
Alistipes
Coriobacteriaceae−unkGen
Actinomyces
Ruminococcaceae−unkGen
Parabacteroides
Veillonellaceae−unkGen
Megasphaera
Xylanibacter
Catenibacterium
Prevotellaceae−unkGen
Prevotella
Roseburia
Lachnospiraceae−unkGen
Bryantella
Syntrophococcus
Meyerozyma
Alternaria
Rhodotorula
Basidiomycota−phylum
Fusarium
Aspergillus
Candida
Pichia
Cystofilobasidium
−0.2 0.2
Spearman Cor.

Figure 2-9. Inter-generic relationships: cross-domain correlations. The heatmaps quantify the intergeneric relationships. (A) Normalized z-score of the bacterial and fungal proportions for samples grouped according to their archaeal status (Methanobrevibacter positive, Nitrososphaera positive, or archaea negative). Asterisks indicate Kruskall-Wallis significant comparisons after FDR adjustment (FDR of 25, 20, 15, and 10% are marked with 1, 2, 3 or 4 asterisks, respectively). Domain membership is color-coded on the left. (B) Spearman correlations between Fungi and Bacteria. Asterisks in red indicate FDR adjusted significant correlations (FDR 20%) and the remaining raw p-values are shown to illustrate general patterns within the data.

The *Prevotella/Bacteriodes* ratio was implicated as important for the gut microbiome structure in previous studies (Yatsunenko, Rey *et al.*, 2012), so we assessed correlations with fungal and archaeal taxa. We performed a Permanova test using the *Prevotella*/*Bacteroides* ratio and the generalized Unifrac matrices obtained from the fungal composition data. A significant relationship was observed (p=0.0146) indicating a potential influence of the *Prevotella/Bacteriodes* ratio on fungi. A post-hoc test with the individual weighted and unweigheted unifrac matrices showed that there was a significantly correlation between the *Prevotella/Bacteriodes* ratio and the weighted Unifrac distances (p=0.0133), but not with the unweighted Unifrac matrix. Thus the *Prevotella*/*Bacteroides* ratio correlated with the amounts of fungi present, but not the types. For the archaea, *Bacteriodes* was significantly negatively correlated with *Methanobrevibacter*, but positively correlated with both *Nitrososphaera* and no archaea. *Prevotella* showed a reciprocal pattern, but it did not achieve significance, probably because of the lower numbers of *Prevotella*-postive samples.

We also used a PermanovaG test to determine whether the proportion of fungal phyla (Ascomycota or Basidiomycota) correlated with the gut bacterial lineages. Both Ascomycota and Basidiomycota were significantly correlated with the bacterial lineages (p=0.0202 and p=0.0037, respectively). A post-hoc Permanova test with the individual weighted and unweighted Unifrac matrices was significant only when using the weighted Unifrac matrix (p=0.0205 and p=0.004, for the Ascomycota and Basidiomycota, respectively). A targeted analysis of the Fungi (Figure 3B) showed a negative association of *Candida* with *Bacteriodes*. Together these results indicate that the types of fungal species in

the gut were not correlated with the bacterial taxa present, but rather their relative proportions.

3.3.5 Co-occurrence analysis using the Dice index

Microbial communities in diverse settings have been shown to form syntrophic communities, in which metabolic outputs from one microbe provide needed inputs to another. An initial analysis of community structure was done to determine which types of microbes co-occur, as scored by the Dice index. For this analysis, we only used data of relatively abundant genera (within Domain sample proportion of 0.01 or greater).

Numerous associations were detected. Figure 4 shows these interactions, incorporating data over all three Domains and indicating co-occurrence by the color code. Inspection of the figure shows several examples of co-occurrence involving a high proportion of samples (Figure 4, lower left corner, warmer colors). *Bacteroides* occurred commonly with *Parabacteroides*, *Lachnospiraceae*, and *Ruminococcaceae*. *Lachnospiraceae* occurred commonly with *Faecalibacterium* and *Ruminococcaceae* as well as *Bacteroides*. In contrast, co-occurance of *Methanobrevibacter* and *Nitrososphaera* was low, as mentioned above.

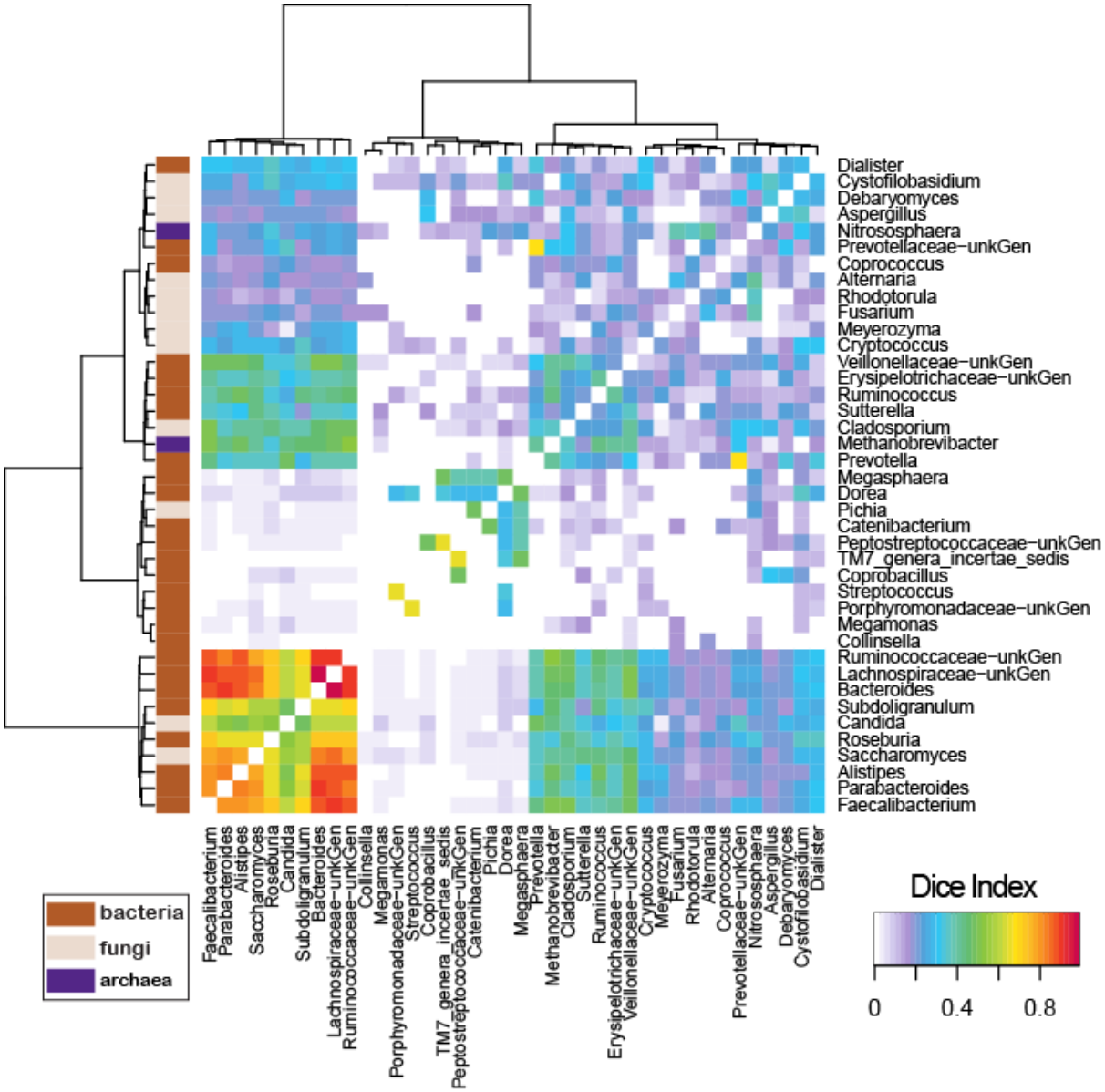


Figure 2-10. Analysis of co-occurance among microbial lineages scored using the Dice index. The calculated Dice indexes across all genera pairs which are present across the samples at a proportion >= 0.01 are shown as a heatmap. Clustering was carried out using Ward's criteria, based on the Euclidian distance between each genus pair using their Dice index across all other genera. Domain membership is color-coded on the left.

Candida and *Saccharomyces* were both associated with the group containing the above bacteria. Several other fungi achieved levels sufficient for inclusion in the analysis, but showed less clearcut co-occurrence with other community members. *Methobrevibacter* showed modest levels of association with the above group of bacteria, but was similarly also associated with *Prevotella* and several other bacterial groups. *Nitrososphaera* showed only

modest associations with other lineages. The frequent co-occurrence of some of these microorganisms suggests candidate interactions among gut microbes for further investigation.

3.3.6 The Gut Microbial Network

The Dice index matrix was used to construct a co-occurrence network (Figure 5). Genera were considered co-occurring for all Dice indexes of 0.3 or above. Modules were detected using the Walktrap algorithm and no similar structure was detected in randomly created networks based on the real data relationships. Figure 5A shows the graph representation of the co-occurrence matrix with the detected modules colored. A Core-Periphery structure is observed, where all members of the core area are highly interconnected. The structure fails to conform with Scale free or random networks with a bimodal degree distribution (Figure 5C). Four main modules were detected: the first included the *Nitrososphaera* archaeon and several fungi genera, the second and third modules included only bacterial species and could potentially be merged into a single module, and the fourth contained the network core, including genera *Bacteroides*, *Prevotella*, *Methanobrevibacter*, *Candida* and *Sacharomyces*. The same network structure is represented as a tree (Figure 5B), to highlight the high level of compartmentalization present in the network.

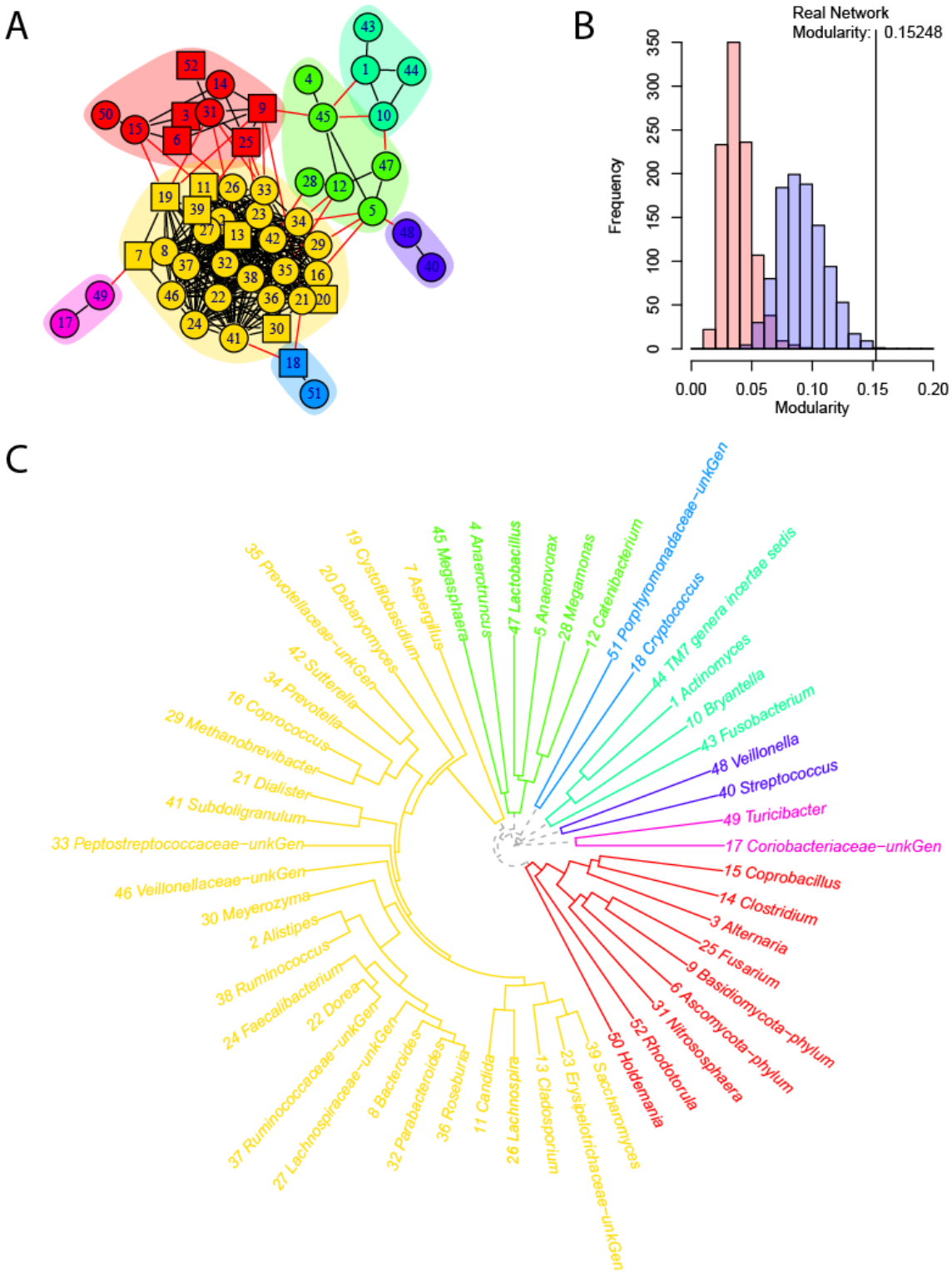


Figure 2-11. Microbial Network present in the gut microbiome. The Dice index matrix was used to create a co-occurrence matrix. For bacteria and fungi, genera were considered present in a sample if their relative proportion was equal or higher then 0.01% of the total sequences. Genera were deemed co-occurring if they had a Dice index of 0.3 or higher. A) Network Graph of the co-occurring genera in the gut microbial ecosystem. Modules detected using the Walktrap algorithm are represented with distinct colors. Red edges indicate between module connections. Genera nodes are labeled as panel C. B) Histogram distribution of the modularity detected in simulated networks. Histogram in red shows the modularity of random networks obtained using the same node degrees distribution as present in the real network. Histogram in blue distribution shows the modularity of random networks created following a power-law degree distribution, containing the same number of nodes and edges as the real network. C) The same network structure represented as a dendrogram.

3.4 Discussion

Here we investigated the fungal and archaeal composition of the gut microbiome in a cohort of subjects previously characterized for their gut bacteria composition, determining the relationships existing between the Bacteria, the Fungi and the Archaea across all individuals in this sample dataset.

We detected no fewer than 62 fungal genera and 184 species level OTUs, paralleling and extending a study of one subject, which also yielded a high number of fungal lineages (Hamad, Sokhna *et al.*, 2012). Which of these lineages are true gut residents, and which are transients in food is unknown. Six individuals had sequences belonging to genus *Agaricus*, the white button mushroom, which is consumed as food. This genus was among those filtered out of the analysis due to low prevalence in the dataset, suggesting that relatively little fungal DNA in food may survive digestion. However, we cannot exclude the possibility that the high prevalence of *Sacharomyces* in the fungal data is due to the ingestion of yeast-containing foods such as bread and beer.

Nitrososphaera, which were encountered with unexpected frequency in our data, are different enough from other archaeal groups to be placed in their own phylum, the *Thaumarchaeota*. Members of the *Nitrososphaera* genus are able to oxidize ammonia and degrade urea, which presumably would also feed nitrogen into the gut microbial community. *Nitrososphaera* may have been previously underappreciated in microbiome studies due to its low abundance. Here, it was detected in 16% of the samples analyzed, though in low abundance, in a mutually exclusive pattern with *Methanobrevibacter*. The basis of possible antagonism between *Methanobrevibacter* and *Nitrososphaera* is unknown,

and we cannot exclude the possibility that *Nitrososphaera* was ingested with foods, being only a transient member of the gut microbiome.

Previously, interactions among microbial lineages of the gut were proposed to separate human populations into "enterotypes", leading to considerable controversy. Arumugam et al. (Arumugam, Raes *et al.*, 2011) described three interacting networks of microbial lineages, centered on the presence of *Prevotella*, *Bacteriodes*, and *Ruminococcus*, together with other interacting taxa. However, subsequent work indicated that the most prominent feature of the data was an inverse correlation between the *Prevotella* and *Bacteroides* genera (Wu, Chen *et al.*, 2011; Yatsunenko, Rey *et al.*, 2012). Here we show a positive association of *Methanobrevibacter* and *Candida* with the *Prevotella* group. A negative correlation of *Methanobevibacter* and *Bacteroides* was also observed, paralleling the original Arumugam et al. paper. However, we also observed a strong positive relationship between *Methanobrevibacter* and *Ruminococccus*, which were initially described to belong to distinct enterotypes. Thus our findings associate archaea and fungi with aspects of the enterotype concept.

The network analysis detected 4 main modules with core-perifery structure, with a high level of redundancy in the relationships present within the core area of the network. This highlights the high level of interpersonal variation observed across individuals. Moreover, the high number of connections within the core indicates a high level of resilience in the gut microbiome to perturbations in the system. Species may only be active in certain situations, for example: as diet shifts daily, distinct species would fulfill the same physiological roles, supplying the remaining members of the microbiome with the necessary nutrients to suport their growth. Periphery clusters may suggest the presence of specialized

metabolism which is only used when certain nutrients are present in the human diet. As mentioned above, *Nitrososphaera*, is reported to use urea and ammonia for its essential metabolism, and these two compounds are byproducts of protein metabolism. Thus, the cluster containing the *Nitrososphaera* OTU may be responsible for the degradation of proteins or the clearance of the metabolic byproducts produced by the host's protein metabolism. In Chapter 4 we explore the relationships of the microbiome and the human diet in greater depth.

3.5 Conclusion

The findings presented here provide a broader picture of the human gut microbiome, integrating the inter-domain relationships that are bound to exist in this complex ecosystem. The associations detected, together with other work (Ley, Hamady *et al.*, 2008; Koenig, Spor *et al.*, 2011; Yatsunenko, Rey *et al.*, 2012) and the work explored in chapter 5, also allow the proposal of specific relationships between nutrition and microbial consortia within the human gut.

3.6 References

ANDERSON, H. W. Yeast-Like Fungi of the Human Intestinal Tract. *The Journal of Infectious Diseases*, v. 21, n. 4, p. 13, 1917.

ARUMUGAM, M. *et al.* Enterotypes of the human gut microbiome. *Nature*, v. 473, n. 7346, p. 174-80, May 12 2011.

CAPORASO, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, v. 7, n. 5, p. 335-6, May 2010.

CHEN, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, v. 28, n. 16, p. 2106-13, Aug 15 2012.

COLE, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, v. 37, n. Database issue, p. D141-5, Jan 2009.

COSTELLO, E. K. *et al.* The application of ecological theory toward an understanding of the human microbiome. *Science*, v. 336, n. 6086, p. 1255-62, Jun 8 2012.

DELONG, E. F. Archaea in coastal marine environments. *Proc Natl Acad Sci U S A*, v. 89, n. 12, p. 5685-9, Jun 15 1992.

DETHLEFSEN, L. *et al.* The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS biology*, v. 6, n. 11, p. e280, Nov 18 2008.

DICE, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology*, v. 26, n. 3, p. 6, jul., 1945 1945.

DOLLIVE, S. *et al.* A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol*, v. 13, n. 7, p. R60, Jul 3 2012.

DRAY, S.; DUFOUR, A. B. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, v. 22, n. 4, p. 20, 2007.

DRIDI, B. *et al.* High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol. *PLoS ONE*, v. 4, n. 9, p. e7063, 2009.

DRIDI, B. *et al.* Archaea as emerging organisms in complex human microbiomes. *Anaerobe*, v. 17, n. 2, p. 56-63, Apr 2011.

ECKBURG, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science*, v. 308, n. 5728, p. 1635-8, Jun 10 2005.

FAUST, K. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol*, v. 8, n. 7, p. e1002606, 2012.

GEVERS, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biol*, v. 10, n. 8, p. e1001377, 2012.

GHANNOUM, M. A. *et al.* Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog*, v. 6, n. 1, p. e1000713, Jan 2010.

GILL, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science (New York, N.Y.)*, v. 312, n. 5778, p. 1355-1359, Jun 2 2006.

GRICE, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science*, v. 324, n. 5931, p. 1190-2, May 29 2009.

GUMBO, T. *et al.* Candida glabrata Fungemia. Clinical features of 139 patients. *Medicine (Baltimore)*, v. 78, n. 4, p. 220-7, Jul 1999.

HAMAD, I. *et al.* Molecular detection of eukaryotes in a single human stool sample from Senegal. *PLoS One*, v. 7, n. 7, p. e40888, 2012.

HEUER, H. *et al.* Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl Environ Microbiol*, v. 63, n. 8, p. 3233-41, Aug 1997.

ILIEV, I. D. *et al.* Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science*, v. 336, n. 6086, p. 1314-7, Jun 8 2012.

KOENIG, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*, v. 108 Suppl 1, p. 4578-85, Mar 15 2011.

KUCZYNSKI, J. *et al.* Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol*, v. 11, n. 5, p. 210, May 5 2010.

LEPP, P. W. *et al.* Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A*, v. 101, n. 16, p. 6176-81, Apr 20 2004.

LEY, R. E. *et al.* Evolution of mammals and their gut microbes. *Science (New York, N.Y.)*, v. 320, n. 5883, p. 1647-1651, Jun 20 2008.

LOZUPONE, C. A.; KNIGHT, R. Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev*, v. 32, n. 4, p. 557-78, Jul 2008.

MATARAZZO, F. *et al.* The domain Archaea in human mucosal surfaces. *Clin Microbiol Infect*, v. 18, n. 9, p. 834-40, Sep 2012.

MCKENNA, P. *et al.* The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS pathogens*, v. 4, n. 2, p. e20, Feb 8 2008.

MILLER, T. L. *et al.* Isolation of Methanobrevibacter smithii from human feces. *Appl Environ Microbiol*, v. 43, n. 1, p. 227-32, Jan 1982.

MINOT, S. *et al.* Hypervariable loci in the human gut virome. *Proc Natl Acad Sci U S A*, v. 109, n. 10, p. 3962-6, Mar 6 2012.

MINOT, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res*, Aug 31 2011.

MINOT, S. *et al.* Conservation of Gene Cassettes among Diverse Viruses of the Human Gut. *PLoS One*, v. 7, n. 8, p. e42342, 2012.

MURDOCH, T. B. *et al.* Pattern recognition receptor and autophagy gene variants are associated with development of antimicrobial antibodies in Crohn's disease. *Inflamm Bowel Dis*, v. 18, n. 9, p. 1743-8, Sep 2012.

OTT, S. J. *et al.* Fungi and inflammatory bowel diseases: Alterations of composition and diversity. *Scand J Gastroenterol*, v. 43, n. 7, p. 831-41, 2008.

REYES, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*, v. 466, n. 7304, p. 334-8, Jul 15 2010.

REYES, A. *et al.* Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol*, v. 10, n. 9, p. 607-17, Aug 6 2012.

SAMUEL, B. S. *et al.* Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc Natl Acad Sci U S A*, v. 104, n. 25, p. 10643-8, Jun 19 2007.

SCUPHAM, A. J. *et al.* Abundant and diverse fungal microbiota in the murine intestine. *Appl Environ Microbiol*, v. 72, n. 1, p. 793-801, Jan 2006.

SEGATA, N. *et al.* Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*, v. 13, n. 6, p. R42, 2012.

SEIBOLD, F. *et al.* Anti-Saccharomyces cerevisiae antibodies in inflammatory bowel disease: a family study. *Scand J Gastroenterol*, v. 36, n. 2, p. 196-201, Feb 2001.

STEWART, J. A. *et al.* Carriage, quantification, and predominance of methanogens and sulfate-reducing bacteria in faecal samples. *Lett Appl Microbiol*, v. 43, n. 1, p. 58-63, Jul 2006.

THE HMP CONSORTIUM. Structure, function and diversity of the healthy human microbiome. *Nature*, v. 486, n. 7402, p. 207-14, Jun 14 2012.

TOURNA, M. *et al.* Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci U S A*, v. 108, n. 20, p. 8420-5, May 17 2011.

URSELL, L. K. *et al.* The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol*, v. 129, n. 5, p. 1204-8, May 2012.

WU, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science*, v. 334, n. 6052, p. 105-8, Oct 7 2011.

YATSUNENKO, T. *et al.* Human gut microbiome viewed across age and geography. *Nature*, v. 486, n. 7402, p. 222-7, Jun 14 2012.

CHAPTER 4. DIET AND THE HUMAN GUT MICROBIOME

The contents of this chapter have been submitted for publication or published as:

> Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, and Bushman FD. (2013) Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. PLoS ONE. 8(6): e66019. doi:10.1371/journal.pone.0066019.
>
> Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, et al. (2011) Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. Science 334 (6052):105-108. DOI: 10.1126/science.1208344

4.1 Introduction

The intestinal microbiota, until recently considered a mere bystander, has been elevated almost to the status of an active organ, playing a role in a wide range of host physiological processes. Examples of these host-microorganism relationships are the fermentation of indigestible carbohydrates, the synthesis of certain essential vitamins, the degradation of dietary oxalates, the development of the immune system, the protection of the epithelial tissue, and stimulation of angiogenesis (Hooper e Gordon, 2001). It is composed of >100 trillion microorganisms such as bacteria, archaea, micro-eukaryotes and viruses, which are, for the most, not yet cultured (Ley, Peterson *et al.*, 2006). The combined genome of this microbial community is at least 10-fold greater than the human genome.

The microbiota composition differs according to body site, and these differences are present not only between distinct body sites (such as mouth versus skin) but also between regions that could be considered more similar to each other (such as regions of the gastrointestinal tract)(Costello, Lauber *et al.*, 2009). A distinct pattern can also be observed in the same body site according to health and disease state such as inflammatory bowel disease (Sartor, 2008), obesity (Ley, Backhed *et al.*, 2005), and colon cancer (Scanlan,

Shanahan *et al.*, 2006). Several factors affect the composition and maintenance of the gut microbiome, including genetics and the host phenotype, the immune system (Macpherson e Harris, 2004), the intake of antibiotics, the environment, and the diet (Faith, Mcnulty *et al.*, 2011). Of these factors, diet, the main source of energy for the microbiome, is the easiest to manipulate in a clinical setting, with far fewer potential side effects.

Although the microbiota is different from person to person, individuals within a human population can be grouped according to their gut microbiome (Arumugam, Raes *et al.*, 2011; Wu, Chen *et al.*, 2011). Indeed, we have recently showed that the human population could be distinguished in at least two groups, the so called enterotypes, and within each group the microbiota is dominated by one or a few species. The bacterial species signature present in each cluster has been strongly associated with long-term dietary patterns. Particularly, the consumption of protein and animal fat was specific to the *Bacteroides* cluster, whereas carbohydrates consumption was associated with the *Prevotella*-dominated cluster. Another study recently reported that the diversity of the gut microbiota was linked with long-term diet, where a more diverse diet was correlated with an increased gut bacterial diversity. The authors also reported that long term dietary interventions (as long as 1 year) were needed to completely switch the intestinal bacterial composition in the cohort studied (Claesson, Jeffery *et al.*, 2012).

These results demonstrate that diet can modulate the gut microbiota composition and therefore have an impact on human health; however, the mechanisms by which diet influences the gut microbiome remain to be fully understood. Here we access the fungal and archaeal components of the human microbiome and associate them with the human diet.

4.2 Methods

4.2.1 Samples:

The sample set described in (Wu, Chen *et al.*, 2011) was used in this study. The bacterial 16S sequences data published therein were used in this analysis. Briefly, healthy volunteers were screened to be free from any chronic gastrointestinal disease, cardiac disease, diabetes mellitus or immunodeficiency diseases, to have a normal bowel frequency (minimum once every 2 days, maximum 3 times per day), and body mass index (BMI) between 18.5 and 35. One stool sample was provided per subject and kept frozen at -80^{o}C until processed for DNA extraction (Wu, Chen *et al.*, 2011).

16S rDNA gene and ITS1 region PCR amplification and pyrosequencing, as well as sequence analysis were performed as described in Chapter 3.

4.2.2 Beta diversity:

Taxonomic information can be used as a proxy for the relationships between distinct species (Lozupone e Knight, 2008). The taxonomic information was used to obtain the Taxonomic Distance between each genus using the R package Ade4, version 1.5 (Dray e Dufour, 2007). The taxonomic distance matrix was used as input to calculate Unifrac distances using the R package GUniFrac, version 1.0 (Chen, Bittinger *et al.*, 2012).

4.2.3 Diet Analysis:

The same dietary information published previously (Wu, Chen *et al.*, 2011) was used in this analysis. Usual diet was obtained using the Willett food frequency questionnaire (Willett, Stampfer *et al.*, 1983). Recent diet was obtained from 3 interviews recalling all

consumed food on 3 days within the week preceding the sample acquisition (NHANES method, ((Nchs). 2011)). All interviews were carried out by trained nutritionists. Nutrient measurements across individuals obtained from the dietary questionnaires were used in a clustering procedure to reduce the number of comparisons to be made.

First, Spearman correlations were calculated pairwise for all nutrient variables available and this correlation matrix was used as input in a clustering analysis using Ward's criterion, to reduce within cluster variance. Clusters were selected to reduce the number of variables to approximately 10% of the total number of nutrients available. Nutrients within each cluster were submitted to a Principal Component Analysis (PCA) and the first principal component values were extracted to be used as a surrogate dietary measurement (Nutrient Cluster Measurement). The same procedure was carried out for the usual and recent derived nutrients, yielding 20 and 14 clusters respectively.

The nutrient cluster measurements obtained in the PCA were used on a Permanova analysis together with the taxonomy based, weighted and unweighted Unifrac distances calculated using the R packages Ade4 and GUniFrac. Clusters which were significant in the Permanova analysis were further used to calculate Spearman rank correlations using the proportion for each bacterial and fungal genus across all samples. The Nutrient Cluster Measurements for each sample were also classified according to their archaeal status (*Methanobrevibacter* positive, *Nitrososphaera* positive, or Archaea negative) and submitted to a Kruskall-Wallis test to determine archaea/diet relationships. False Discovery Rates were calculated to assess p-values significance and values of 25% or less were considered significant.

4.3 Results

4.3.1 Clustering of Dietary Nutrients

The dietary information produced by recall or food frequency questionnaires is composed of many co-varying variables, several of which are slightly different representations of the same nutrient. For example, all amino acids will be highly positively correlated with each other and total protein. A clustering procedure was implemented to reduce the number of comparisons being made while still being able to accurately represent the diet intake. Figure 1 shows heatmaps constructed with the whole correlation matrix, as well as selected clusters. Cross-cluster correlations are also shown to demonstrate the high level of within cluster correlations. The procedure is able to correctly discriminate, without supervision, the main nutrient classes: Protein, carbohydrate and fatty acids, as well as other major nutrient classes (Tables 1 through 4).

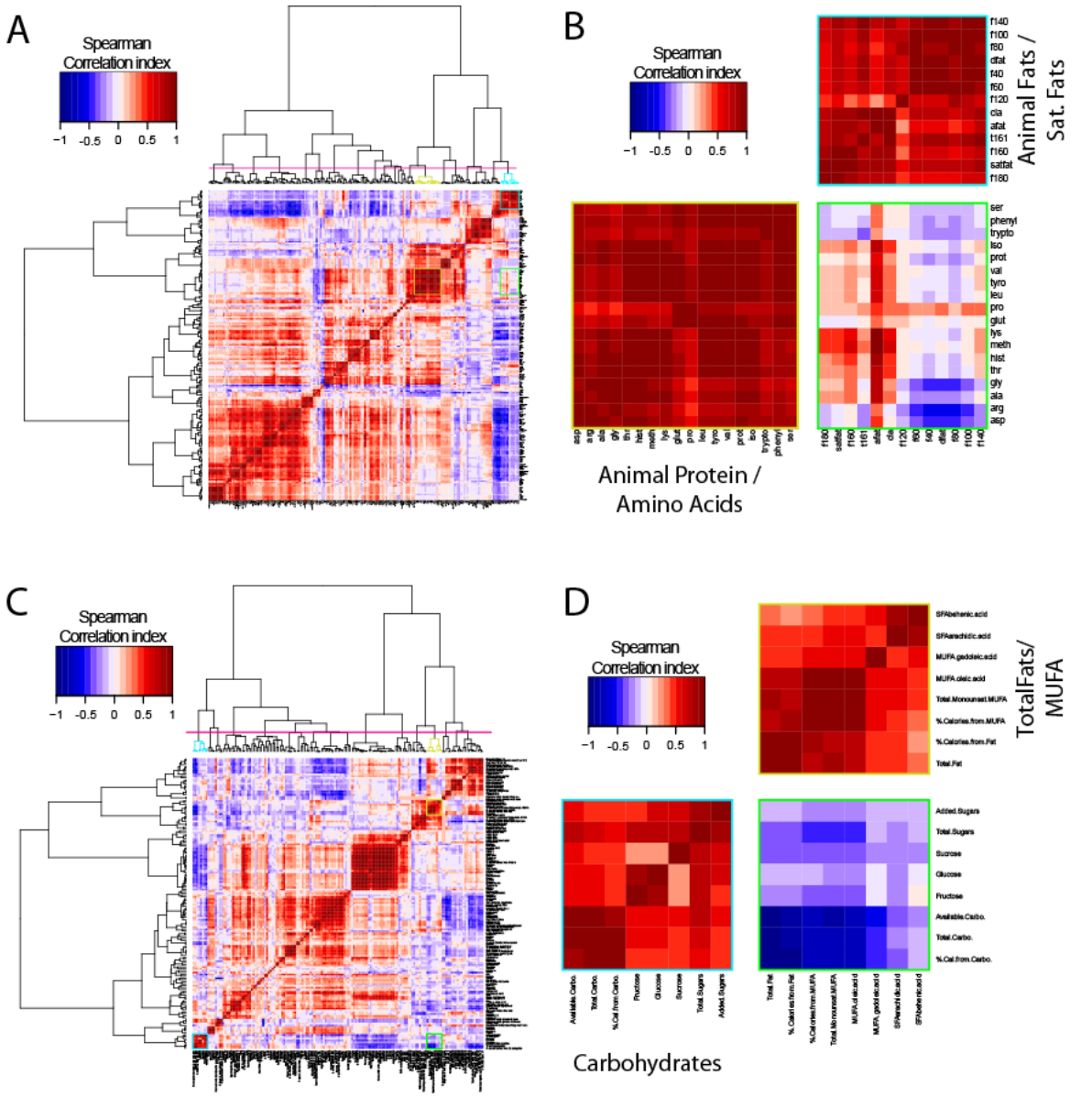


Figure 2-12. : Diet Nutrient Clustering. Dietary nutrients were measure using a food frequency questionnaire to access the subjects usual diet and using a recall interview to access the recently consumed diet. Nutrients were correlated using Spearman correlation to obtain representative nutrient clusters. A) Correlations between all nutrients from the usual diet are shown demonstrating strong positive correlations for groups of similar nutrients. The cut off used to create the nutrient clusters is indicated by the magenta line on the top dendrogram. B) Representative clusters from the usual diet marked with the yellow and blue boxes are highlighted as well as their cross-correlations (green box). C) and D) are the same representations but using nutrients accessed from the subjects' recent diet. For a complete list of nutrients and their cluster assignment, see Appendix II.

4.3.2 Archaeal associations with diet

Correlations between diet and bacteria have been reported previously for this data set (Wu, Chen *et al.*, 2011) and were recapitulated here after using the pre-clustered dietary data (below). As a first step, Permanova tests were carried out using the nutrient cluster

data and the weighted and unweighted Unifrac distance matrixes to determine if the Fungal and Archaeal fractions of the microbiome were also influenced by diet. Multiple significant associations were detected and are cataloged in (Tables 1 through 4).

Table 2.8. Usual diet nutrient clusters unweighted unifrac.

Cluster Number	cluster name	Bacteria	Fungi	Bacteria + Fungi	Archaea + Bacteria + Fungi
1	Calories/ GI / Betaine	0.97	0.5977	0.9114	0.9061
2	Amino Acids	0.1655	0.2825	0.0194	0.0181
3	animal fat / saturated fat / dairy fat	0.191	0.8183	0.3604	0.422
4	vegetable fat	0.9732	0.7841	0.9824	0.9054
5	Simple Sugars	0.0027	0.6927	0.002	0.0002
6	*Vitamin A / minerals*	0.9191	0.8993	0.9111	0.9245
7	calcium/lactose/dairy protein	0.4535	0.601	0.6198	0.6685
8	iron / zinc	0.9142	0.1565	0.7424	0.7494
9	vitamin C / Flavones / flavonones	0.5851	0.4466	0.5739	0.5109
10	B vitamins	0.5541	0.6177	0.807	0.8301
11	poly unsaturated fats	0.2634	0.7527	0.41	0.2758
12	choline	0.4226	0.5441	0.0653	0.0538
13	*alcohol / flavonols / antocyanidins*	0.1897	0.4156	0.2455	0.2582
14	animal protein / poly unsaturated fats	0.7377	0.367	0.2787	0.2232
15	omega and long chain fats	0.9238	0.4135	0.6611	0.599
16	trans fats	0.1228	0.2395	0.0964	0.1121
17	fiber (bran, wheat germ)	0.6723	0.7467	0.755	0.6908
18	aspartame	0.0068	0.9754	0.0518	0.0497
19	chatechins	0.8434	0.5505	0.8929	0.8954
20	*antocyanidins*	0.5426	0.7521	0.7326	0.7984

Table 2.9. Usual diet nutrient clusters and weighted unifrac.

Cluster	cluster name	Bacteria	Fungi	Bacteria +	*Archaea +*

Number				Fungi	*Bacteria + Fungi*
1	Calories/ GI / Betaine	0.4613	0.6318	0.5861	*0.6251*
2	Amino Acids	0.0015	0.7253	0.0157	*0.0173*
3	animal fat / saturated fat / dairy fat	0.1009	0.1401	0.0526	*0.2181*
4	vegetable fat	0.6136	0.1501	0.3001	*0.0396*
5	Simple Sugars	0.0053	0.0461	0.0028	*0.0001*
6	*Vitamin A / minerals*	0.8172	0.9664	0.9845	*0.9918*
7	calcium/lactose/dairy prot	0.1583	0.7369	0.5681	*0.774*
8	iron / zinc	0.9851	0.1204	0.3626	*0.4054*
9	vitamin C / Flavones / flavonones	0.506	0.8689	0.7858	*0.5164*
10	B vitamins	0.8657	0.5819	0.8417	*0.9467*
11	poly unsaturated fats	0.9624	0.587	0.8442	*0.0587*
12	choline	0.0062	0.8195	0.0349	*0.0384*
13	*alcohol / flavonols / antocyanidins*	0.4206	0.7883	0.6427	*0.7757*
14	animal protein / poly unsaturated fats	0.0069	0.2854	0.0077	*0.0131*
15	omega and long chain fats	0.3298	0.2102	0.1637	*0.1163*
16	trans fats	0.2128	0.228	0.1977	*0.6007*
17	fiber (bran, wheat germ)	0.7467	0.6048	0.7082	*0.4321*
18	aspartame	0.3106	0.9944	0.8965	*0.4343*
19	chatechins	0.7408	0.3535	0.4508	*0.6509*
20	*antocyanidins*	0.5948	0.4921	0.6509	*0.9543*

Table 2.10. Recent diet nutrient clusters unweighted unifrac.

Cluster Number	cluster name	Bacteria	Fungi	Bacteria + Fungi	Archaea + Bacteria + Fungi
1	Energy/Glycemic Load/ Cholest2SFA ratio	0.0079	0.3449	0.0225	0.0233
2	Total Fat/ %cal from fat	0.0728	0.3847	0.0714	*0.023*
3	Total Carb/ simple sugars/ Total Sugar/ Added Sugars	0.1148	0.3822	0.0846	0.0572

4	Total Prot/ AA/ %cal From Prot/ Nitrogen	0.7405	0.5593	0.6949	0.7413
5	Total, soluble, insoluble fiber/ Veg Prot	0.0341	0.9058	0.1798	0.2362
6	*Alcohol/ vit A, beta carotene/ betaine*	0.023	0.9968	0.1716	0.2126
7	PUFA (Arachdonic, EPA, DHA, DPA)/ choline/ cholesterol	0.776	0.8822	0.8451	0.8293
8	total SFA/ %cal from sfa/ CLA/ myristoleic and palmitoleic MUFA	0.2533	0.6779	0.4673	0.4139
9	total polyUnsaturated FA/ omega3	0.1529	0.3506	0.179	0.1995
10	maltose/ starch/ calcium/ galactose/ lycopene/ sweetners	0.3038	0.6802	0.4807	0.47
11	lactose/ retinol/vit D/ B complex/ Iron	0.0832	0.2216	0.0681	0.0762
12	vit E	0.0681	0.7118	0.1895	0.2147
13	*Sat. short chain FA*	0.4068	0.585	0.475	0.5179
14	**glycemic index/ trans fat**	0.0863	0.2675	0.0654	0.0632
15	Daidzein/ Genistein/ Glycitein/ Pinitol	0.1682	0.7607	0.3877	0.2994

Table 2.11. Recent diet nutrient clusters and weighted unifrac.

Cluster Number	cluster name	Bacteria	Fungi	Bacteria + Fungi	*Archaea + Bacteria + Fungi*
1	Energy/Glycemic Load/ Cholest2SFA ratio	0.2366	0.4101	0.4496	0.2547
2	Total Fat/ %cal from fat	0.677	**0.0952**	0.2178	0.0083
3	Total Carb/ simple sugars/ Total Sugar/ Added Sugars	0.3148	**0.008**	0.0128	0.0184
4	Total Prot/ AA/ %cal From Prot/ Nitrogen	0.6167	0.1384	0.2051	0.3144
5	Total, soluble, insoluble fiber/ Veg Prot	0.0675	0.9087	0.4663	0.8005
6	*Alcohol/ vit A, beta carotene/ betaine*	0.1254	0.9083	0.4644	0.7831
7	PUFA (Arachdonic, EPA, DHA, DPA)/ choline/ cholesterol	0.4506	0.6412	0.5281	0.5617
8	total SFA/ %cal from sfa/ CLA/ myristoleic and palmitoleic MUFA	0.0199	**0.0429**	0.0102	0.0297
9	total polyUnsaturated FA/ omega3	0.0739	0.751	0.4135	0.5166
10	maltose/ starch/ calcium/ galactose/ lycopene/ sweetners	0.1362	0.9875	0.6697	0.881
11	lactose/ retinol/vit D/ B complex/ Iron	0.2989	0.6899	0.5792	0.5361
12	vit E	0.6347	0.6584	0.7542	0.7051

13	*Sat. short chain FA*	0.0432	**0.0708**	0.0229	0.077
14	**glycemic index/ trans fat**	0.0064	0.9117	0.1252	0.1756
15	Daidzein/ Genistein/ Glycitein/ Pinitol	0.5967	0.6106	0.6851	0.561

Relationships between archaea and nutrients were explored using a Kruskall-Wallis test on the Permanova selected nutrient cluster measurements. As above, samples were separated into archaea-negative, *Methanobrevibacter*-positive or *Nitrososphaera*-positive groups. A higher intake of carbohydrates was correlated with *Methanobrevibacter*-positive samples. This trend was observed in comparisons to both the long-term and short-term dietary data. For the long-term diet, samples with *Nitrososphaera* or no archaea were enriched for clusters representing vegetable fat and poly-unsaturated fats. Samples with no archaea were enriched for a cluster representing total fat and total mono-unsaturated fats in the recent diet data (Figure 2).

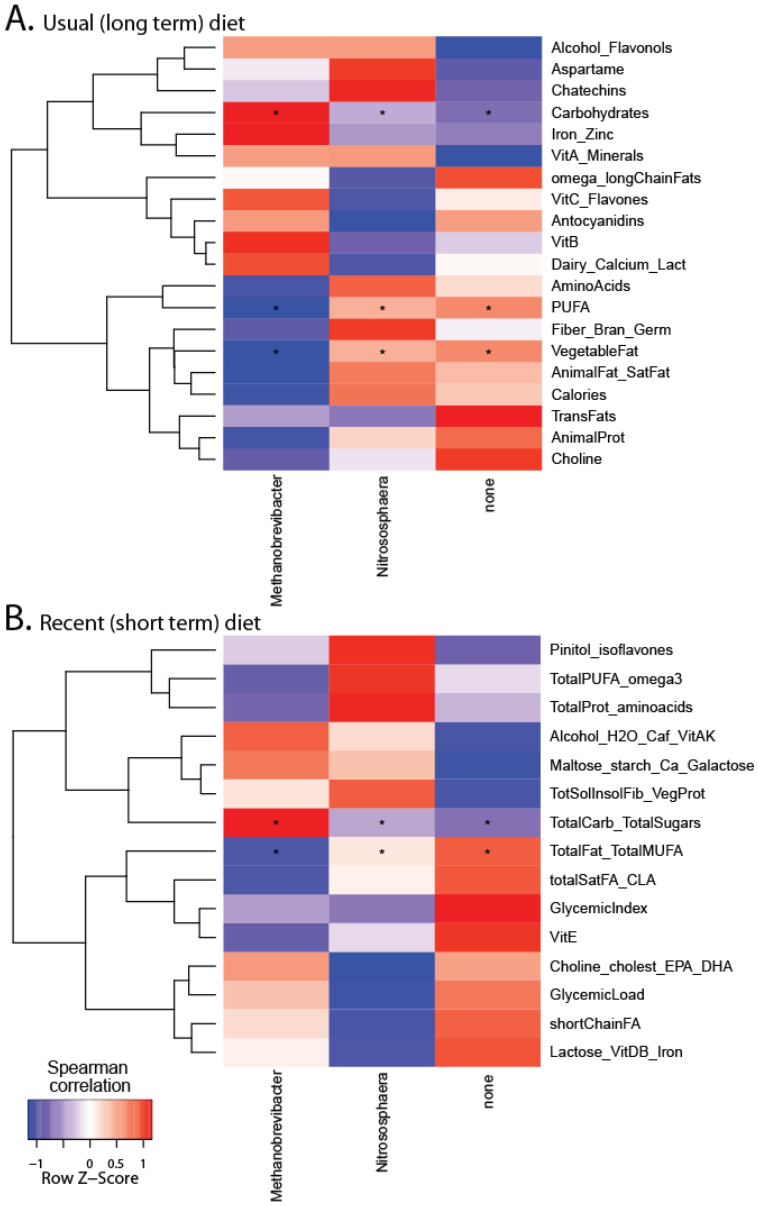


Figure 2-13. Archaea-Diet relationships. Heatmap of normalized average means for nutrient cluster measurements of the samples classified according to the dominant archaeal genus. Usual diet (A) and recent diet (B). Asterisks indicate Kruskall-Wallis significant comparisons after FDR adjustment.

4.3.3 Fungal associations with diet

Correlations between fungi and bacteria and the nutrient clusters are shown in Figure 3. Fungal and bacterial proportions were scored versus usual diet (Figure 3A) or versus recent diet (Figure 3B). Only diet categories with at least one significant association are shown. As reported previously, bacterial proportions were most strongly correlated with

components of the long term diet--*Bacteriodes* was more abundant in individuals eating high levels of animal protein, amino acids, and fats, while *Prevotella* was higher among those eating higher proportions of carbohydrates. For fungi significant correlations were observed only with the recent diet inventory, differing from the observations with bacteria. In the recent diet data, *Candida* was positively correlated with carbohydrates and negatively with total saturated fatty acids. A trend in the same direction was seen for *Candida* in the usual long-term diet, but it did not achieve significance. *Aspergillus* was negatively correlated with short chain fatty acids in the recent diet data. No trends were seen with *Saccharomyces*. Thus, these data indicate that fungal abundance is particularly strongly associated with the composition of recently consumed foods.

4.4 Discussion:

It was reported previously, for this same set of samples, that patterns in the bacterial part of the gut microbiome correlated with long-term diet. Here we recapitulate those finding using a novel dietary clustering method: *Prevotella* and *Ruminnococcus* were both positively correlated with the nutrient clusters representing Carbohydrate ingestion and negatively correlated with the ingestion of proteins, while *Bacteroides* presented the inverse pattern.

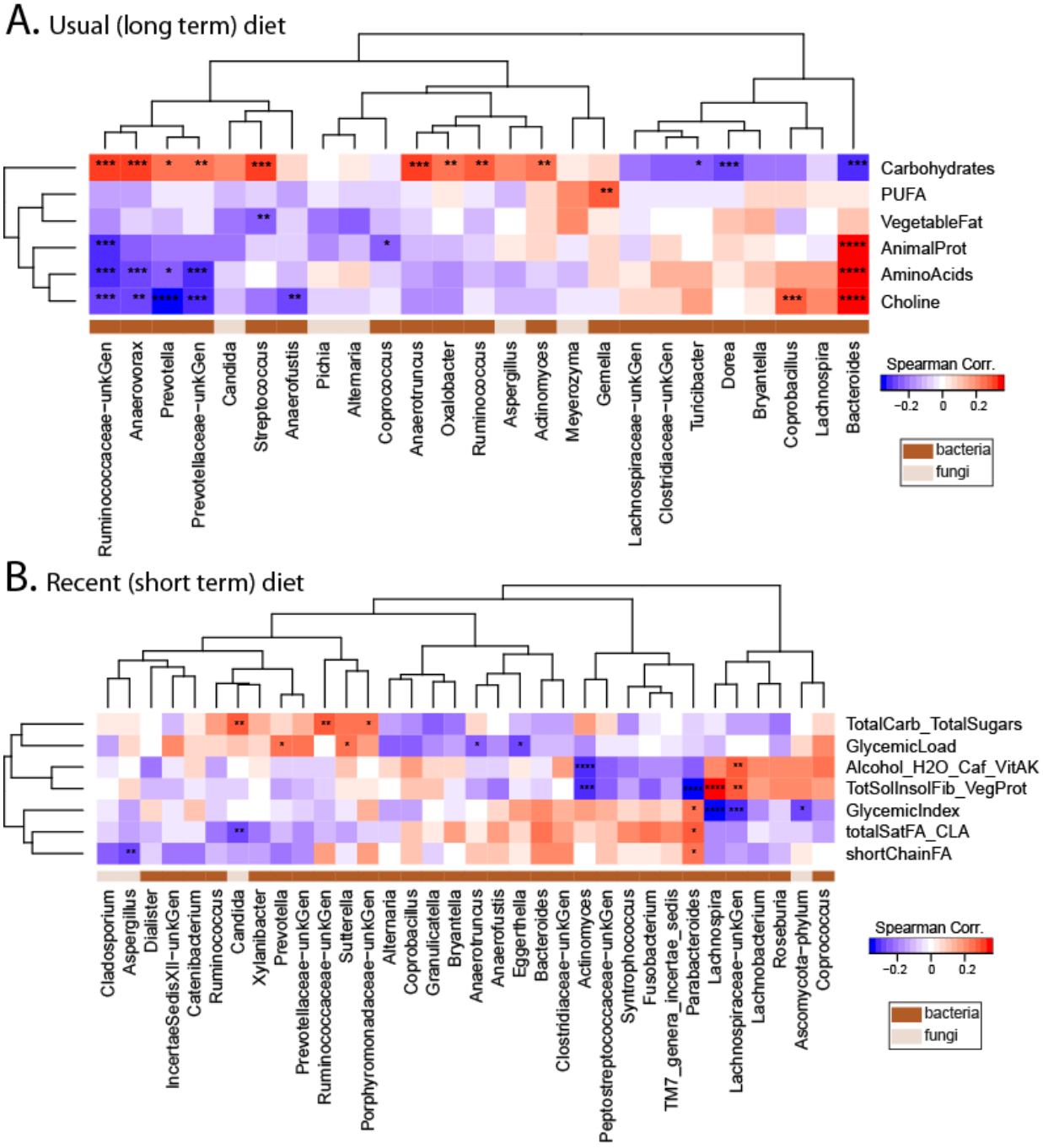


Figure 2-14. Fungi-Diet and Bacteria-Diet relationships. Heatmap of Spearman correlations between nutrient clusters and the bacterial and fungal genera detected in the dataset. Correlations which were considered significant after FDR correction using the Usual (A) and the Recent (B) diet data are marked with asterisks (FDR of 25, 20, 15, and 10% are marked with 1, 2, 3 or 4 asterisks, respectively). Domain membership is color-coded on the bottom.

We also characterized the dataset for fungal and archaeal relationships with diet intake. Among the many interactions between microorganisms and nutrients reported here, we found that *Methanobrevibacter* and *Candida* were positively correlated with the

ingestion of carbohydrates. This was most notable in short-term diet data for both groups, and in fact only achieved significance at all for *Candida* in the short-term diet data. These data support specific proposals for the interactions of members of the gut microbiome with dietary components and with each other.

Nitrososphaera also showed a positive association with the ingestion of proteins and amino acids, both for usual and recent diet. The correlation was not sufficient to survive correction for multiple comparisons, but may nevertheless indicate utilization of ammonia and or urea to meet their energy and carbon requirements. Alternatively, we cannot exclude the possibility that *Nitrososphaera* was ingested with foods, possibly associated with meats.

These interactions with nutrients presented here, taken together with the inter-generic relationships presented in chapter 3, further support models for specific interactions among microbes in the human gut. An example of syntrophism has been previously described for *Rumminococcus* and methanogens, where the methanogens' consumption of H_2 allowed *Rumminococcus* to produce twice as many ATP molecules from the same amount of substrate (Stams e Plugge, 2009). One possible syntrophic guild specified in our data includes *Candida*, *Prevotella*, *Rumminococcus* and *Methanobrevibacter* (Figure 4). *Candida* is able to degrade starches, especially when those are pre-treated with amylases (Iannotti, Kafkewitz *et al.*, 1973)—amylase proteins are encoded by the human host and present in the mouth and small intestine. Thus, in one model, *Candida* may assist in breaking down starch in carbohydrate rich foods, which in turn liberates simpler sugars to be fermented by bacteria such as *Prevotella* and *Rumminococcus.* Fermentation byproducts produced would then be consumed by *Methanobrevibacter* with the subsequent production

of CO_2 and/or CH_4 (Stams e Plugge, 2009). Alternatively, *Prevotella* might degrade starch (pre-treated or not by human alpha amylases) and mannan containing polysaccharides from food to smaller poly- and monosaccharides (Downes, Tanner *et al.*, 2010). *Prevotella* takes up the smaller mono and polysaccharides, catabolizing them to produce succinate and other byproducts (Van Gylswyk, 1995; Kovatcheva-Datchary, Egert *et al.*, 2009). All such hydrolysis is extra cellular, which would provide *Candida* with simpler sugars for fermentation (potentially down to acetate). *Ruminococcus* might then consume the succinate produced by *Prevotella* and produce H_2 or $acetate^-/H^+$ for consumption by *Methanobrevibacter* (Iannotti, Kafkewitz *et al.*, 1973; Kovatcheva-Datchary, Egert *et al.*, 2009; Purushe, Fouts *et al.*, 2010).

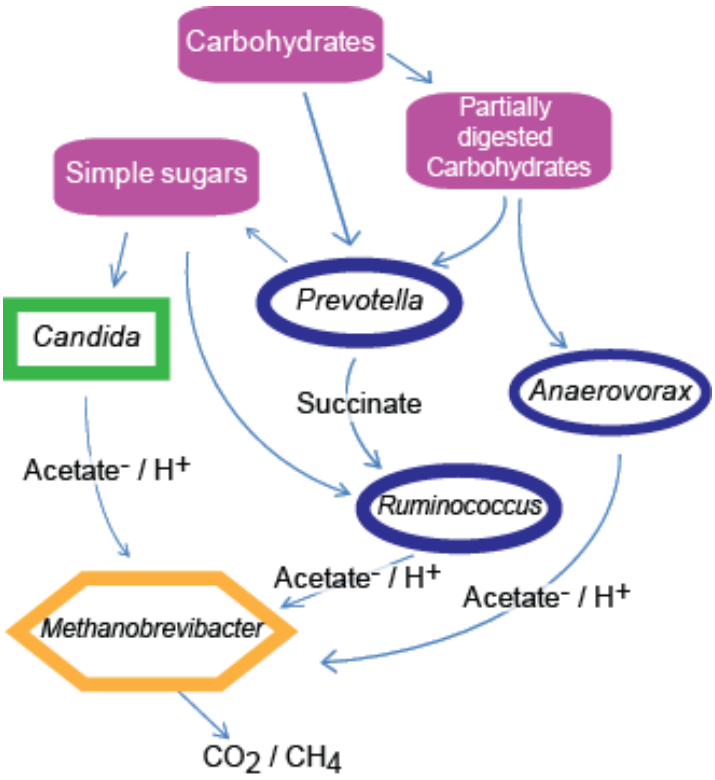


Figure 2-15: Possible syntrophic relationships in the human gut. This is only one possible set of syntrophic relationships present in the human gut, consistent with the data presented.

4.5 Conclusion

The findings presented here provide a broader picture of the human gut microbiome, integrating its full diversity with human diet. The associations presented here, together with other work (Ley, Hamady *et al.*, 2008; Koenig, Spor *et al.*, 2011; Yatsunenko, Rey *et al.*, 2012), allow the proposal of specific relationships between nutrition and microbial consortia within the human gut.

4.6 References

(NCHS)., C. F. D. C. A. P. C. N. C. F. H. S. National Health and Nutrition Examination Survey Questionnaire. . Hyattsville, MD.: Centers for Disease Control and Prevention (CDC). 2011.

ARUMUGAM, M. *et al.* Enterotypes of the human gut microbiome. *Nature*, v. 473, n. 7346, p. 174-80, May 12 2011.

CHEN, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, v. 28, n. 16, p. 2106-13, Aug 15 2012.

CLAESSON, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature*, v. 488, n. 7410, p. 178-84, Aug 9 2012.

COSTELLO, E. K. *et al.* Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science (New York, N.Y.)*, Nov 5 2009.

DOWNES, J. *et al.* Prevotella saccharolytica sp. nov., isolated from the human oral cavity. *Int J Syst Evol Microbiol*, v. 60, n. Pt 10, p. 2458-61, Oct 2010.

DRAY, S.; DUFOUR, A. B. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, v. 22, n. 4, p. 20, 2007.

FAITH, J. J. *et al.* Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science*, v. 333, n. 6038, p. 101-4, Jul 1 2011.

HOOPER, L. V.; GORDON, J. I. Commensal host-bacterial relationships in the gut. *Science (New York, N.Y.)*, v. 292, n. 5519, p. 1115-1118, May 11 2001.

IANNOTTI, E. L. *et al.* Glucose fermentation products in Ruminococcus albus grown in continuous culture with Vibrio succinogenes: changes caused by interspecies transfer of H 2. *J Bacteriol*, v. 114, n. 3, p. 1231-40, Jun 1973.

KOENIG, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*, v. 108 Suppl 1, p. 4578-85, Mar 15 2011.

KOVATCHEVA-DATCHARY, P. *et al.* Linking phylogenetic identities of bacteria to starch fermentation in an in vitro model of the large intestine by RNA-based stable isotope probing. *Environ Microbiol*, v. 11, n. 4, p. 914-26, Apr 2009.

LEY, R. E. *et al.* Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, v. 102, n. 31, p. 11070-5, Aug 2 2005.

LEY, R. E. *et al.* Evolution of mammals and their gut microbes. *Science (New York, N.Y.)*, v. 320, n. 5883, p. 1647-1651, Jun 20 2008.

LEY, R. E. *et al.* Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, v. 124, p. 837-484, 2006.

LOZUPONE, C. A.; KNIGHT, R. Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev*, v. 32, n. 4, p. 557-78, Jul 2008.

MACPHERSON, A. J.; HARRIS, N. L. Interactions between commensal intestinal bacteria and the immune system. *Nature reviews.Immunology*, v. 4, n. 6, p. 478-485, Jun 2004.

PURUSHE, J. *et al.* Comparative genome analysis of Prevotella ruminicola and Prevotella bryantii: insights into their environmental niche. *Microb Ecol*, v. 60, n. 4, p. 721-9, Nov 2010.

SARTOR, R. B. Microbial influences in inflammatory bowel diseases. *Gastroenterology*, v. 134, n. 2, p. 577-94, Feb 2008.

SCANLAN, P. D. *et al.* Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol*, v. 44, n. 11, p. 3980-8, Nov 2006.

STAMS, A. J.; PLUGGE, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol*, v. 7, n. 8, p. 568-77, Aug 2009.

VAN GYLSWYK, N. O. Succiniclasticum ruminis gen. nov., sp. nov., a ruminal bacterium converting succinate to propionate as the sole energy-yielding mechanism. *Int J Syst Bacteriol*, v. 45, n. 2, p. 297-300, Apr 1995.

WILLETT, W. C. *et al.* Validation of a dietary questionnaire with plasma carotenoid and alpha-tocopherol levels. *Am J Clin Nutr*, v. 38, n. 4, p. 631-9, Oct 1983.

WU, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science*, v. 334, n. 6052, p. 105-8, Oct 7 2011.

YATSUNENKO, T. *et al.* Human gut microbiome viewed across age and geography. *Nature*, v. 486, n. 7402, p. 222-7, Jun 14 2012.

CHAPTER 5. DIETARY INTERVENTION EFFECTS ON THE HUMAN GUT MICROBIOME

5.1 Introduction

Several studies have linked the gut microbiome and diet related health issues, such as obesity (Turnbaugh, Ley *et al.*, 2006) and coronary disease(Tang, Wang *et al.*, 2013). Obesity has been linked to a distinct bacterial community composition in a cohort of identical, but discordant twins(one obese, one lean)(Turnbaugh, Hamady *et al.*, 2009). Studies in mouse models have also shown that germ-free mice have a significantly higher weight gain when transplanted with the gut microbiome from obese individuals than from lean individuals (Turnbaugh, Ley *et al.*, 2006). This raises several important questions when considering that humans are germ free when born, and are colonized throughout the first year of life (Gonzalez, Clemente *et al.*, 2011). If the microbiome is at least partially responsible for an obese state, can obesity be treated by manipulating the microbiome, either via diet or via direct interventions?

A recent report indicates that diet has quick impact on the gut microbiome (Wu, Chen *et al.*, 2011). Subjects were placed in a high or low fat diet for 10 days and the intestinal bacterial community was sequenced. Effects were observed both on community composition as well species proportions in as little as 24 hours after initiation of the intervention. Nevertheless , these effects were modest and intersubject variability still predominated in the bacterial composition of the gut microbiome (Figure 1).

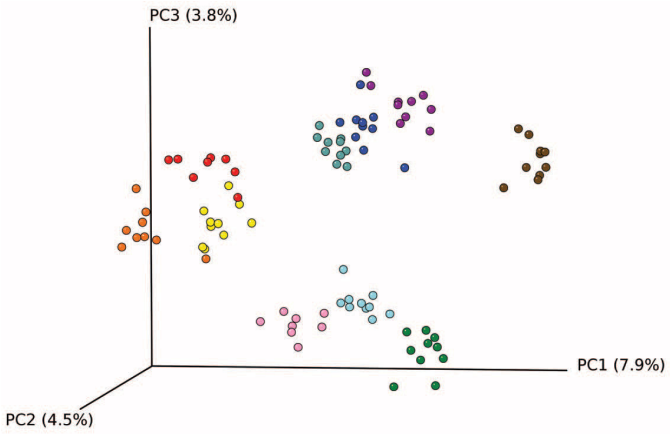


Figure 2-16. Changes in bacterial communities during controlled feeding experiment. Ten subjects were randomized to high-fat/low-fiber or low-fat/high-fiber diets, and microbiome composition was monitored longitudinally for 10 days by sequencing 16S rDNA gene tags.The figure shows a cluster diagram based on a principal coordinates analysis using unweighted UniFrac. Colors indicate samples from each individual. Statistical analysis demonstrated the day 1 samples are outliers compared to all other days, indicating change in the gut microbiome within 24 hours of initiating controlled feeding (Wu, Chen *et al.*, 2011).

The linkage between diet and the microbiome is not exclusive to human. The relationship between mammals and their intestinal microbiome, for example, was recently shown to reflect each group's dietary needs: mammal groups can be separated using only the bacterial data into carnivores, omnivores and herbivores (Ley, Hamady *et al.*, 2008; Ley, Lozupone *et al.*, 2008). Another study in Burmese pythons (Costello, Gordon *et al.*, 2010) has also demonstrated that drastic changes in the intestinal microbiome of these snakes happen after a meal, and such changes start at 1 day post feeding. The state induced post-feeding is dramatically distinct from that of fasting, and the source of such shift is unlikely to be transient bacteria introduced with the food intake. Such data leads inevitably to the question: is it possible to remodel the human gut microbiome from one state to another only via a change in diet, and how long does that take?

These questions are addressed here with two dietary intervention studies. One places individuals on a diet containing higher or lower calorie levels then their usual intake. The second intervention studies the effect of a simpler intervention, where diet is

maintained 'ad lib" and a high dosage omega-3 fatty acid is introduced. Pyrosequencing of a bacterial 16S rDNA amplicon is used to access the changes in the gut microbiome in these interventions.

5.2 Methods

5.2.1 Samples and dietary interventions

Healthy volunteers were recruited and subjected to 2 distinct interventions: the omega-3 intervention and the calorie intake intervention.

12 patients were placed in the omega intervention. Each subject was placed on a high dose of omega-3 fish oil (17.6g per day) for 4 weeks. Samples were collected weekly, with 2 samples prior to treatment initiation, 4 during the treatment, and 2 after treatment cessation.

6 patients were placed in the calorie intake intervention. A cross over design was used, where subjects were randomly assigned to the receive two dietary courses, one with a high calorie intake, and one with low calorie intake, with relation to each subject's normal caloric intake. Half of the subjects were placed first on a high calorie diet, followed by a 1 month wash out period, and then placed in a low calorie diet. The other half of the subjects was placed in the inverse order. Each dietary change was preceded by a period of isocaloric diet to normalize the most recent diet intake for all patient. The period of intervention for each stay was 10 days. Samples were collected daily during the dietary interventions.

5.2.2 DNA extraction, PCR and sequencing

DNA was extracted from stool samples using the PSP stool plus DNA extraction kit (Invitek, Berlin, Germany). The V1V2 region of the 16S rDNA gene was amplified and used to

survey the bacterial fraction of the gut microbiome as described in chapter 2. PCR was set up in quadruplicate using the Accuprime polymerase (Invitrogen, Carlsbad, CA) with conditions as described in Chapter 3. PCR was performed on an ABI 2720 Thermocycler using the following conditions: Initial denaturing at 95˚C for 5 minutes followed by 20 cycles of 95˚C X 30 seconds, 56˚C X 30 seconds, and 72˚C X 1 minute 30 seconds. The reaction was terminated after an 8 minute extension at 72˚C. PCR products were bead purified using the AmPure beads (Beckman-Coulter, Brea, CA).

5.2.3 454/Roche pysequencing methods

Purified amplicon DNAs were quantified using Quant-iT PicoGreen kit (Invitrogen, Carlsbad, CA) and pooled in equal amounts for pyrosequencing. Pyrosequencing using the 454/Roche GS FLX chemistry was carried out according to the manufacturer's instructions. Pyrosequencing using the Titanium method was carried out using the Titanium genomic kit. Primers for use with each method are in Table 1.

Pyrosequence reads were uploaded into QIIME and processed as described (Caporaso, Kuczynski *et al.*, 2010). Briefly, QIIME accepts as input bar coded 16S rRNA gene sequences, classifies them using the RDPclassifier (Wang, Garrity *et al.*, 2007), aligns them using Pynast (Caporaso, Bittinger *et al.*, 2010), constructs phylogenetic trees using FastTree2 (Price, Dehal *et al.*, 2009), calculates Unifrac distances, and generates data summaries that proportions of taxa present and PCA plots of Unifrac distances. We used 97% OTUs in the analysis. For the RDP classifier, we required >50% confidence for all calls. OTU's were only considered in further analysis if they had at least 5 sequences detected across all samples sequenced.

5.2.4 Statistical methods

Samples characteristics were used to perform a Permanova analysis with the weighted and unweighted Unifrac distance matrices. For the omega-3 intervention saples were classified as pre-treatment, during treatment, and post treatment. For the calorie intake intervention samples were classified as iso-caloric diet, high calorie diet, low calorie diet. Permutations were constrained to each subject.

For the Calorie intake intervention, the proportion of each genus was used in a Kruskall-Wallis test to access if any genera was enriched for a given treatment. To be considered in this analysis, a genus had to be present in at least 50% of the all samples sequenced.

5.3 Results

5.3.1 Calorie intake intervention

Six healthy subjects were subjected to a high calorie and low calorie diets, using a cross over design, with a period of 1 month between treatments. The unweighted Unifrac PCoA results (Figure 3)indicated that intersubject variability was still the main cause of sample clustering. However, the diet had an effect on the weighted Unifrac analysis (Permanova, p=0.238). As an overall effect was observed on the weighted analysis (Figure 4), samples from each subject were subjected to a PCoA (Figure 4D). Samples obtained while under a high calorie diet tended to form clusters separated from either isocaloric or low calorie diet samples, but not enough data was available to achieve statistical significance. As an effect was observed for the type of diet, samples were partitioned according to diet and the genera proportion within each samples used to determine diet intake/genus relationships (Figure 5). *Dorea* and *Coprobacillus* were enriched in samples

obtained while on a low calorie diet, while unknown Lachnospiraceae were enriched in samples from a high calorie diet.

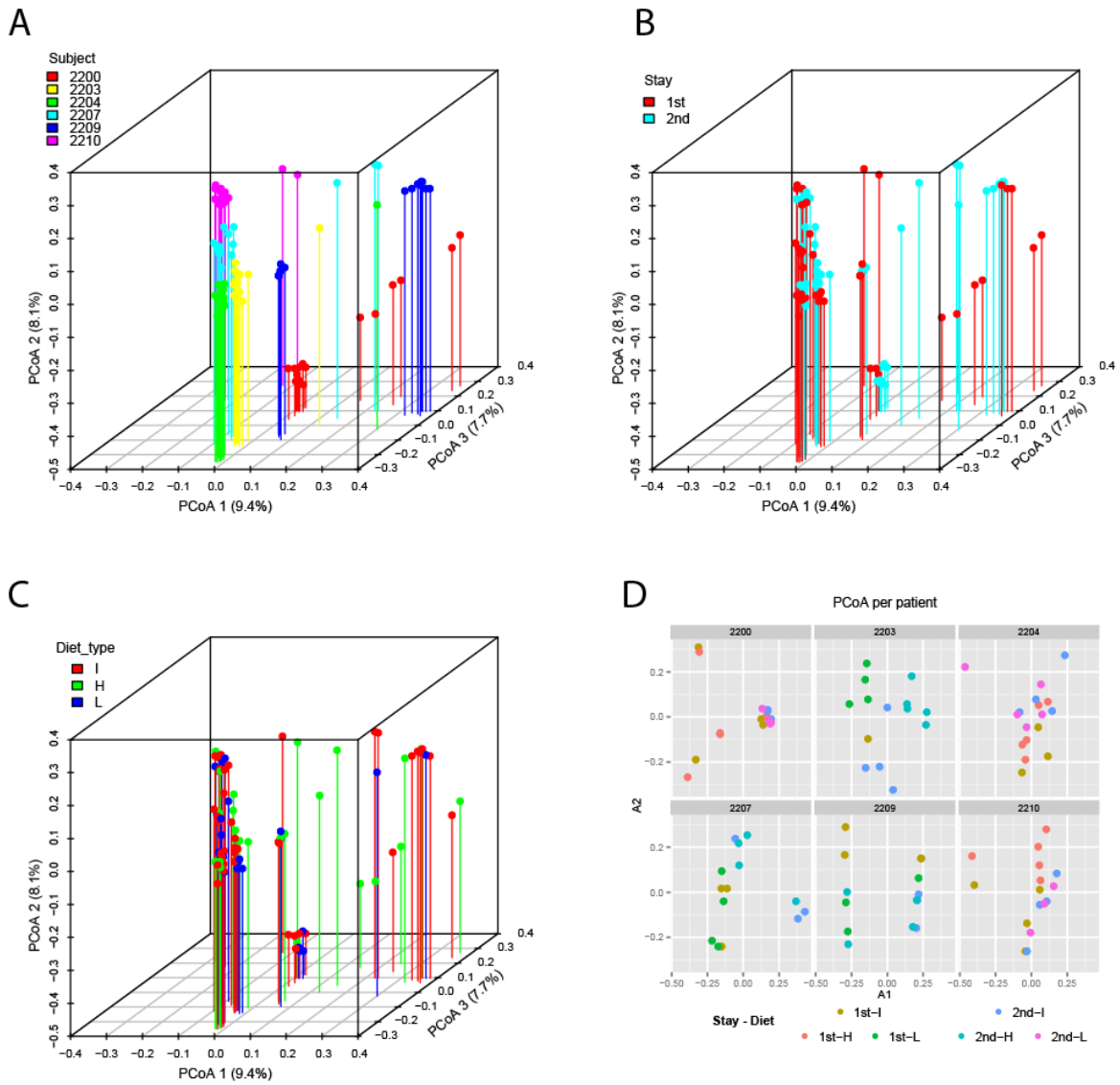


Figure 2-17. Dietary intervention changing calorie intake represented with unweighted Unifrac PCoA. Panels A, B and C shown the first 3 axis from a principal coordinate analysis (PCoA) using the unweighted Unifrac distances between samples. Samples are color coded by Subject (A), Stay during from the treatment course (B) and the Diet type the subjects were on (C). Percent variation explained by each axis is shown in parentheses. D shows the results for PCoA done using only within subject samples. Color code indicated the stay and diet assigned to subject for each sample. Inter-subject variation still predominates the sample clustering patterns observed.

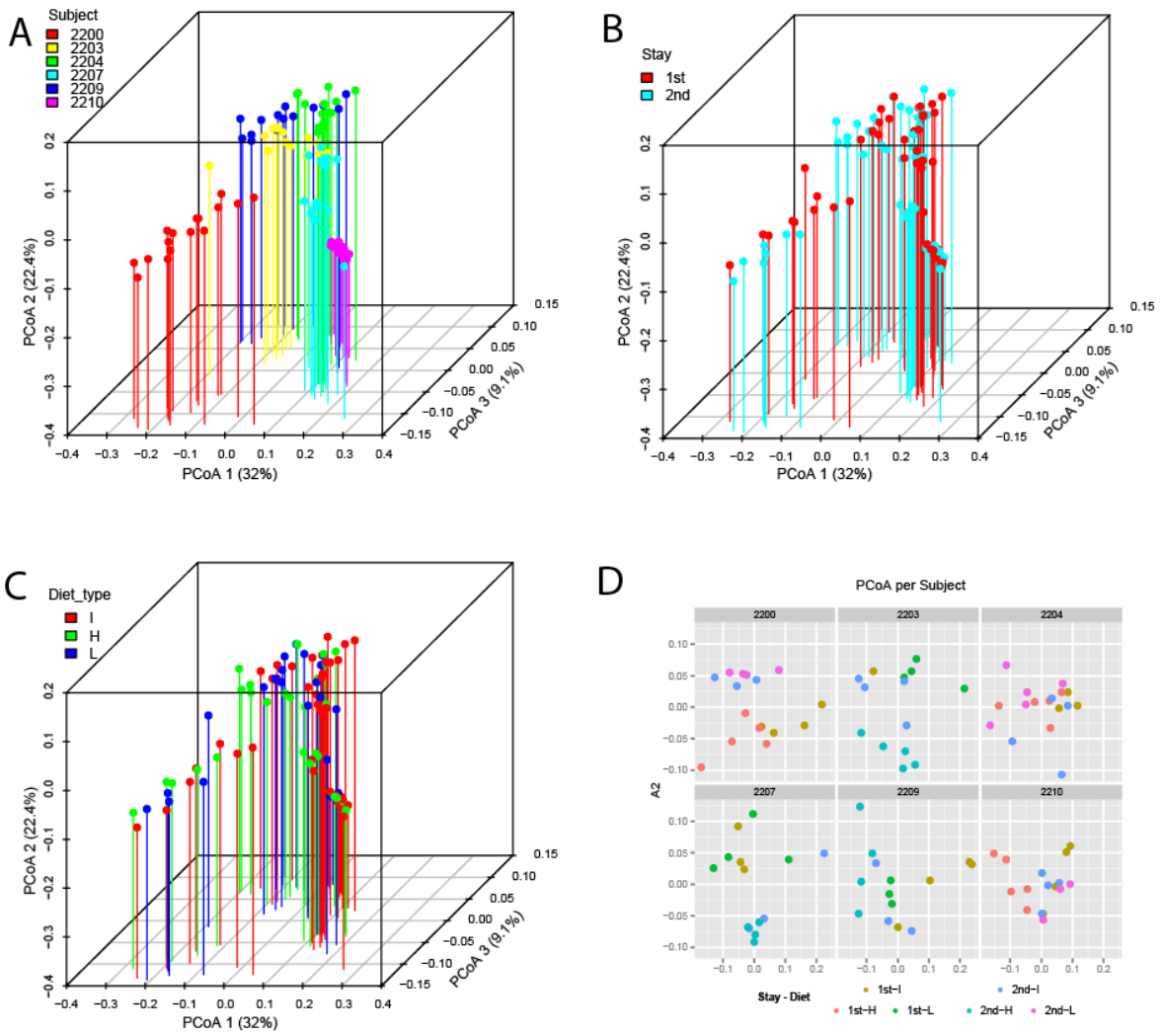


Figure 2-18. Dietary intervention changing calorie intake represented with weighted Unifrac PCoA. Panels A, B and C shown the first 3 axis from a principal coordinate analysis (PCoA) using the weighted Unifrac distances between samples. Samples are color coded by Subject (A), Stay during from the treatment course (B) and the Diet type the subjects were on (C). Percent variation explained by each axis is shown in parentheses. D shows the results for PCoA done using only within subject samples. Color code indicated the stay and diet assigned to subject for each sample.

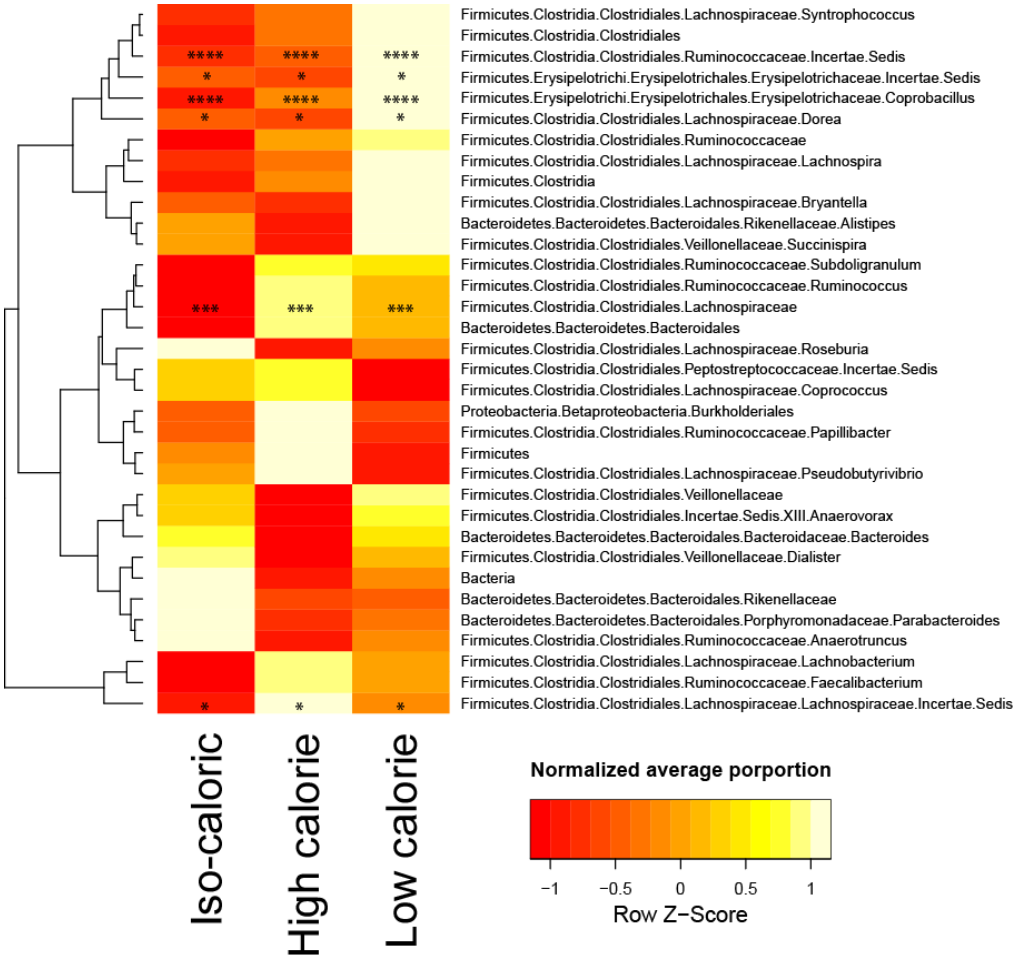


Figure 2-19. Heatmap of taxa detected in the Change in Calorie intervention. Operational taxonomic units were created and classified down to the genus or the lowest level possible. Only taxa present in at least 50 percent of the samples are included in the heatmap. asterisks indicate significant groups using a Kruskall-Wallis test after FDR correction (*<=25%, **<=20%, ***<=15% and ****<=10%).

5.3.2 Omega-3 intervention

Healthy subjects were recruited and assigned to a 4 week treatment with purified Omega-3 fatty acids and stool samples were collected weekly. Two samples were collected prior to treatment initiation, 4 samples were collected while on omega-3 treatment and 2 samples were collected after treatment was discontinued. Figure 2 shows the results of a PCoA using Unifrac distances calculated pairwise across all samples. Both unweighted (Figure 2A) and weighted unifrac (Figure 2B) colored by patient. No effects from the treatment were seen on the microbiome and the inter-subject variation predominated.

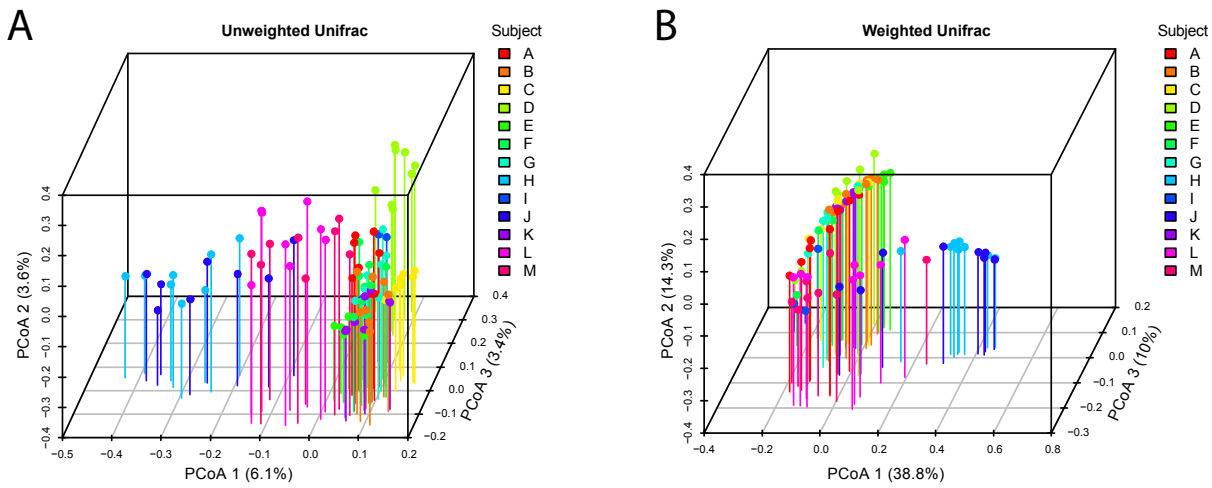


Figure 2-20: Community distances between samples for Omega-3 fatty acid intervention. Samples are color coder by subject and plots are the first 3 coordinates from a principal coordinate analysis of (A) unweighted and (B) weighted Unifrac distances. Percent variation explained by each coordinate is shown in parentheses. The inter-subject variation is still observable after the treatment course and no discernible changes are seen between samples taken pre-treatment, during treatment or after the treatment was stopped.

5.4 Discussion

The gut microbiome of healthy volunteers was analyzed for their bacterial composition following two dietary interventions. One intervention was broad, changing the calorie intake of the individuals, while the second intervention was specific, maintaining the diet of the subjects the same, but introducing a high level dose of omega-3 fatty acids.

The effects of the interventions on the human microbiome, when present, were modest, and the inter-subject variability still predominated in the across individuals. No discernible effects of the omega-3 fatty acid introduction were observed in the bacterial community of the individuals tested, even though the highest dose considered safe was used in this intervention. This may reflect the fact that a longer than 4 weeks intervention is needed to successfully shift the bacterial populations within the human intestine. However, we cannot exclude the possibility that changes in gene expression were significant, but bacterial populations remained stable, as this type of functional shift has been observed in other systems. For instance, a study measuring the effect of pro-biotic intake in a

fermented milk product has shown no significant effects on the intestinal bacterial community composition, of both humans or using a mouse model with a simplified humanized gut microbiome (Mcnulty, Yatsunenko *et al.*, 2011). However, the study was able to show that the microbiome's gene expression was specifically changed, indicating that functional changes may not be reflected in the bacterial species diversity and community evenness.

The caloric intake intervention showed an effect detectable only using the weighted Unifrac distance, indicating changes in the underlining relative proportions of bacterial groups present in the subject's gut. This finding is in contrast with the previously published using a change in the intake of fat (Wu, Chen *et al.*, 2011). Further analysis on the proportions of genera detected across the different types of diet revealed modest, but discernible changes. Genera were detected which were enriched in all 3 distinct dietary intakes, and all belonged to the Fimicutes phylum. Interestingly, two genera enriched in the low calorie diet, *Dorea* and *Coprobacillus*, were also reported to have higher proportions on elderly people with poorer, less diverse diets (Claesson, Jeffery *et al.*, 2012).

Both dietary interventions tested herein had mild effects in the gut microbiome. In fact, Claesson et al (Claesson, Jeffery *et al.*, 2012) were able to show that, in an elderly cohort of subjects, diet is completely changed within one month, but the microbiome takes between 6 months to 1 year to completely shift from one state to another. This emphasizes that dietary interventions aimed at modulating the gut microbiome may need to be carried out for extended periods of time to help ameliorate, if not completely alter, disease states.

5.5 Conclusions

The results presented here demonstrate that targeted dietary interventions, such as the use of an specific fatty acid, are not enough to change the bacterial species composition of the human gut microbiome, at least within a period of 4 weeks. The question of whether such target intervention is able to change the functional profile of the gut microbiome still remains open. Drastic shifts in diet can change the human intestinal microbiome, even within a period of ten days.

These and other results highlight the need for functional studies, and a broader understanding of the mechanisms that control microbial populations within the human gut.

5.6 References

CAPORASO, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, v. 26, n. 2, p. 266-7, Jan 15 2010.

CAPORASO, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, v. 7, n. 5, p. 335-6, May 2010.

CLAESSON, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature*, v. 488, n. 7410, p. 178-84, Aug 9 2012.

COSTELLO, E. K. *et al.* Postprandial remodeling of the gut microbiota in Burmese pythons. *ISME J*, v. 4, n. 11, p. 1375-85, Nov 2010.

GONZALEZ, A. *et al.* Our microbial selves: what ecology can teach us. *EMBO Rep*, v. 12, n. 8, p. 775-84, Aug 2011.

LEY, R. E. *et al.* Evolution of mammals and their gut microbes. *Science*, v. 320, n. 5883, p. 1647-51, Jun 20 2008.

LEY, R. E. *et al.* Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature reviews.Microbiology*, v. 6, n. 10, p. 776-788, Oct 2008.

MCNULTY, N. P. *et al.* The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med*, v. 3, n. 106, p. 106ra106, Oct 26 2011.

PRICE, M. N. *et al.* FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*, v. 26, n. 7, p. 1641-1650, Jul 2009.

TANG, W. H. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med*, v. 368, n. 17, p. 1575-84, Apr 25 2013.

TURNBAUGH, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature*, v. 457, n. 7228, p. 480-484, Jan 22 2009.

TURNBAUGH, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, v. 444, n. 7122, p. 1027-31, Dec 21 2006.

WANG, Q. *et al.* Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, v. 73, n. 16, p. 5261-5267, Aug 2007.

WU, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science*, v. 334, n. 6052, p. 105-8, Oct 7 2011.

CHAPTER 6. CONCLUSIONS AND PERSPECTIVES

The work presented here aims to create a more complete view of the human gut microbiome. It integrates species from all 3 domains of life, their human host, taking under consideration the human diet, the main nutrient source for the gut microbial community.

Chapter 2 dealt with the effects current methods have in the data acquisition employed for microbiome studies. Human interpersonal variability, caused by genetic and/or environmental factors, must be overcome to reveal general patterns in the species composition and community dynamics across human microbiome samples(Ursell, Clemente *et al.*, 2012). Therefore, large scale surveys need to be employed , highlighting the need for methods that are not only effective, but also convenient for the researcher to carry out. We have shown that biases exist regarding the species recovery using 16S rDNA based surveys, but that those are relatively small when compared to the human interpersonal variability.

Methods that favor a hasher cell lysis should be used when there is interest in harder to lyse cells, such as some Gram positive bacteria. We and others (Liu, Lozupone *et al.*, 2007) have shown that the choice of 16S rDNA amplicon used must also be taken under consideration in such microbial surveys, especially if data is to be directly compared with previously published data. Nevertheless, such sequence specific biases could be mitigated with an broader analysis, analyzing together the novel data and the previously published data, aided by the curated reference databases and sequence collections, instead of merely comparing the results.

Most studies published to date have looked at the bacterial components of the human gut microbiome (Thehmpconsortium, 2012). Such studies have estimated that approximately 2000 bacterial species are present therein(Frank e Pace, 2008), and although we started to elucidate the general patterns defining bacteria-human relationships, other microorganisms have been largely ignored. In Chapter 3 we explored the presence of over 60 Fungal genera detected in the human gut, as well as the detection of two highly prevalent Archaeal genera, the *Methanobrevibacter* and *Nitrososphaera*, present in 31% and 16% of samples analyzed, respectively.

Several cross domain relationships were detected between Bacteria, Archaea and Fungi. Most notably were the association of *Prevotella*, *Methanobrevibacter* and *Candida*. These associations indicate that several physiological pathways could be completed by the presence of very distinct microorganisms. When the human nutrient diet was taken under consideration, as present in Chapter 4, associations were detected that indicate the presence of microbial syntrophy. Syntrophic relationships are known to occur in the gut microbiome of ruminants for example(Stams e Plugge, 2009; Wrede, Dreier *et al.*, 2012). One such potential syntrophic relationship detected in the human gut includes *Methanobrevibacter*, *Candida*, *Ruminococcus* and *Prevotella*, all working in concert to utilize nutrients arising from a carbohydrate rich diet.

Although several relationships between human dietary intake are now present in the literature, including the one presented in Chapter 4, we have only started to understand how those relationships can be modified. Dietary interventions present an attractive pathway to threat several human diseases, such as inflammatory bowel disease and Crohn's disease, due to their mild side effects when compared to traditional therapies.

Nevertheless, short term dietary interventions may not be enough to substantially change the gut microbiome composition. Recent reports have demonstrated in a cohort of healthy elderly subjects that even though diet can be completely changed in as little as one month, the gut microbiome can take as long as one year or more to achieve a new community state. This does not preclude the possibility that still unknown specific microbial-human interactions could not be changed to the benefit of the human host(Claesson, Jeffery *et al.*, 2012).

The type of studies presented in this Thesis, where large surveys are performed, can aid in the inference of potential capabilities and the correlation of certain microbial species with host traits. However, the large inter-subject variability present in the human intestinal microbiome can mask the relationship patterns between host traits and the microbiome. Gene function, on the other hand, needs to be preserved if ecological roles are to be maintained across each person within the gut ecosystem, and several studies have shown thus far that the gene categories present in the gut of several people are indeed more similar to each other than their species composition (Lozupone, Stombaugh *et al.*, 2012).

Metagenomic studies of the human intestinal microbiota have been conducted trying to link gene content to host traits, but insofar as diet is concerned, they have dealt with specific questions, such as obese versus non-obese individuals(Turnbaugh, Ley *et al.*, 2006; Turnbaugh, Hamady *et al.*, 2009). investigations of the general patterns of dietary intake and their relationships with the gene content in microbial community present in the human gut would shed light in the genetic potential available to use the nutrients at hand. Furthermore, meta-transcriptomic studies would indeed link the gene function with nutrient metabolism. The integration of several research techniques, in a multi-omics approach, will

invariably complete our understanding of the this dynamic system and create a more complete view of the human meta-organism, both in health and disease.

6.1 References

CLAESSON, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature*, v. 488, n. 7410, p. 178-84, Aug 9 2012.

FRANK, D. N.; PACE, N. R. Gastrointestinal microbiology enters the metagenomics era. *Current opinion in gastroenterology*, v. 24, n. 1, p. 4-10, Jan 2008.

LIU, Z. *et al.* Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic acids research*, Sep 18 2007.

LOZUPONE, C. A. *et al.* Diversity, stability and resilience of the human gut microbiota. *Nature*, v. 489, n. 7415, p. 220-30, Sep 13 2012.

STAMS, A. J.; PLUGGE, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol*, v. 7, n. 8, p. 568-77, Aug 2009.

THE HMP CONSORTIUM. Structure, function and diversity of the healthy human microbiome. *Nature*, v. 486, n. 7402, p. 207-14, Jun 14 2012.

TURNBAUGH, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature*, v. 457, n. 7228, p. 480-484, Jan 22 2009.

TURNBAUGH, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, v. 444, n. 7122, p. 1027-31, Dec 21 2006.

URSELL, L. K. *et al.* The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol*, v. 129, n. 5, p. 1204-8, May 2012.

WREDE, C. *et al.* Archaea in symbioses. *Archaea*, v. 2012, p. 596846, 2012.

BIBLIOGRAPHY

ANDERSON, H. W. Yeast-Like Fungi of the Human Intestinal Tract. *The Journal of Infectious Diseases*, v. 21, n. 4, p. 13, 1917.

ARUMUGAM, M. *et al.* Enterotypes of the human gut microbiome. *Nature*, v. 473, n. 7346, p. 174-80, May 12 2011.

BERNSTEIN, C. N.; BLANCHARD, J. F. The epidemiology of Crohn's disease. *Gastroenterology*, v. 116, n. 6, p. 1503-1504, Jun 1999.

BINLADEN, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, v. 2, p. e197, Feb 14 2007.

BISWAS, B. *et al.* Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant Enterococcus faecium. *Infect Immun*, v. 70, n. 1, p. 204-10, Jan 2002.

BLANCHARD, J. F. *et al.* Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis. *American Journal of Epidemiology*, v. 154, n. 4, p. 328-335, Aug 15 2001.

CANTAREL, B. L. *et al.* Complex carbohydrate utilization by the healthy human microbiome. *PLoS One*, v. 7, n. 6, p. e28742, 2012.

CAPORASO, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, v. 26, n. 2, p. 266-7, Jan 15 2010.

CAPORASO, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, v. 7, n. 5, p. 335-6, May 2010.

CHEN, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, v. 28, n. 16, p. 2106-13, Aug 15 2012.

CHHIBBER, S. *et al.* Co-Therapy Using Lytic Bacteriophage and Linezolid: Effective Treatment in Eliminating Methicillin Resistant Staphylococcus aureus (MRSA) from Diabetic Foot Infections. *PLoS One*, v. 8, n. 2, p. e56022, 2013.

CHO, I.; BLASER, M. J. The human microbiome: at the interface of health and disease. *Nat Rev Genet*, v. 13, n. 4, p. 260-70, Apr 2012.

CLAESSON, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature*, v. 488, n. 7410, p. 178-84, Aug 9 2012.

COLE, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, v. 37, n. Database issue, p. D141-5, Jan 2009.

COSTELLO, E. K. *et al.* Postprandial remodeling of the gut microbiota in Burmese pythons. *ISME J*, v. 4, n. 11, p. 1375-85, Nov 2010.

COSTELLO, E. K. *et al.* Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science (New York, N.Y.)*, Nov 5 2009.

COSTELLO, E. K. *et al.* The application of ecological theory toward an understanding of the human microbiome. *Science*, v. 336, n. 6086, p. 1255-62, Jun 8 2012.

DELONG, E. F. Archaea in coastal marine environments. *Proc Natl Acad Sci U S A*, v. 89, n. 12, p. 5685-9, Jun 15 1992.

DETHLEFSEN, L. *et al.* The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS biology*, v. 6, n. 11, p. e280, Nov 18 2008.

DICE, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology*, v. 26, n. 3, p. 6, jul., 1945 1945.

DOLLIVE, S. *et al.* A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol*, v. 13, n. 7, p. R60, Jul 3 2012.

DOWNES, J. *et al.* Prevotella saccharolytica sp. nov., isolated from the human oral cavity. *Int J Syst Evol Microbiol*, v. 60, n. Pt 10, p. 2458-61, Oct 2010.

DRAY, S.; DUFOUR, A. B. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, v. 22, n. 4, p. 20, 2007.

DRIDI, B. Laboratory tools for detection of archaea in humans. *Clin Microbiol Infect*, v. 18, n. 9, p. 825-33, Sep 2012.

DRIDI, B. *et al.* High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol. *PLoS ONE*, v. 4, n. 9, p. e7063, 2009.

DRIDI, B. *et al.* Archaea as emerging organisms in complex human microbiomes. *Anaerobe*, v. 17, n. 2, p. 56-63, Apr 2011.

ECKBURG, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science*, v. 308, n. 5728, p. 1635-8, Jun 10 2005.

FAITH, J. J. *et al.* Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science*, v. 333, n. 6038, p. 101-4, Jul 1 2011.

FAUST, K. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol*, v. 8, n. 7, p. e1002606, 2012.

FRANK, D. N.; PACE, N. R. Gastrointestinal microbiology enters the metagenomics era. *Current opinion in gastroenterology*, v. 24, n. 1, p. 4-10, Jan 2008.

GEVERS, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biol*, v. 10, n. 8, p. e1001377, 2012.

GHANNOUM, M. A. *et al.* Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog*, v. 6, n. 1, p. e1000713, Jan 2010.

GILL, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science (New York, N.Y.)*, v. 312, n. 5778, p. 1355-1359, Jun 2 2006.

GONZALEZ, A. *et al.* Our microbial selves: what ecology can teach us. *EMBO Rep*, v. 12, n. 8, p. 775-84, Aug 2011.

GRICE, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science*, v. 324, n. 5931, p. 1190-2, May 29 2009.

GUMBO, T. *et al.* Candida glabrata Fungemia. Clinical features of 139 patients. *Medicine (Baltimore)*, v. 78, n. 4, p. 220-7, Jul 1999.

HAMAD, I. *et al.* Molecular detection of eukaryotes in a single human stool sample from Senegal. *PLoS One*, v. 7, n. 7, p. e40888, 2012.

HAMADY, M. *et al.* Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J*, v. 4, n. 1, p. 17-27, Jan 2010.

HAMADY, M. *et al.* Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature methods*, v. 5, n. 3, p. 235-237, Mar 2008.

HE, T. *et al.* The role of colonic metabolism in lactose intolerance. *European journal of clinical investigation*, v. 38, n. 8, p. 541-547, Aug 2008.

HEUER, H. *et al.* Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl Environ Microbiol*, v. 63, n. 8, p. 3233-41, Aug 1997.

HILDEBRANDT, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*, v. 137, n. 5, p. 1716-24 e1-2, Nov 2009.

HOFFMANN, C. *et al.* Community-wide response of gut microbiota to enteropathogenic Citrobacter infection revealed by deep sequencing. *Infection and immunity*, Jul 27 2009.

HOFFMANN, C. *et al.* DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic acids research*, v. 35, n. 13, p. e91, 2007.

HOOPER, L. V.; GORDON, J. I. Commensal host-bacterial relationships in the gut. *Science (New York, N.Y.)*, v. 292, n. 5519, p. 1115-1118, May 11 2001.

IANNOTTI, E. L. *et al.* Glucose fermentation products in Ruminococcus albus grown in continuous culture with Vibrio succinogenes: changes caused by interspecies transfer of H 2. *J Bacteriol*, v. 114, n. 3, p. 1231-40, Jun 1973.

ILIEV, I. D. *et al.* Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science*, v. 336, n. 6086, p. 1314-7, Jun 8 2012.

KOENIG, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*, v. 108 Suppl 1, p. 4578-85, Mar 15 2011.

KOVATCHEVA-DATCHARY, P. *et al.* Linking phylogenetic identities of bacteria to starch fermentation in an in vitro model of the large intestine by RNA-based stable isotope probing. *Environ Microbiol*, v. 11, n. 4, p. 914-26, Apr 2009.

KUCZYNSKI, J. *et al.* Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol*, v. 11, n. 5, p. 210, May 5 2010.

LEMON, K. P. *et al.* Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med*, v. 4, n. 137, p. 137rv5, Jun 6 2012.

LEPP, P. W. *et al.* Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A*, v. 101, n. 16, p. 6176-81, Apr 20 2004.

LEWIS, S. J.; HEATON, K. W. Stool form scale as a useful guide to intestinal transit time. *Scandinavian journal of gastroenterology*, v. 32, n. 9, p. 920-924, Sep 1997.

LEY, R. E. *et al.* Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, v. 102, n. 31, p. 11070-5, Aug 2 2005.

LEY, R. E. *et al.* Evolution of mammals and their gut microbes. *Science*, v. 320, n. 5883, p. 1647-51, Jun 20 2008.

LEY, R. E. *et al.* Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature reviews.Microbiology*, v. 6, n. 10, p. 776-788, Oct 2008.

LEY, R. E. *et al.* Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, v. 124, p. 837-484, 2006.

LIU, Z. *et al.* Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic acids research*, Sep 18 2007.

LOZUPONE, C. *et al.* UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC bioinformatics*, v. 7, p. 371, Aug 7 2006.

LOZUPONE, C.; KNIGHT, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*, v. 71, n. 12, p. 8228-35, Dec 2005.

LOZUPONE, C. A. *et al.* The convergence of carbohydrate active gene repertoires in human gut microbes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 39, p. 15076-15081, Sep 30 2008.

LOZUPONE, C. A. *et al.* Quantitative and Qualitative {beta} Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, v. 73, n. 5, p. 1576-1585, Mar 2007.

LOZUPONE, C. A.; KNIGHT, R. Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev*, v. 32, n. 4, p. 557-78, Jul 2008.

LOZUPONE, C. A. *et al.* Diversity, stability and resilience of the human gut microbiota. *Nature*, v. 489, n. 7415, p. 220-30, Sep 13 2012.

LUDWIG, W. *et al.* ARB: a software environment for sequence data. *Nucleic acids research*, v. 32, n. 4, p. 1363-1371, Feb 25 2004.

MACPHERSON, A. J.; HARRIS, N. L. Interactions between commensal intestinal bacteria and the immune system. *Nature reviews.Immunology*, v. 4, n. 6, p. 478-485, Jun 2004.

MARDIS, E. R. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, v. 9, p. 387-402, 2008.

MATARAZZO, F. *et al.* The domain Archaea in human mucosal surfaces. *Clin Microbiol Infect*, v. 18, n. 9, p. 834-40, Sep 2012.

MCKENNA, P. *et al.* The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS pathogens*, v. 4, n. 2, p. e20, Feb 8 2008.

MCNULTY, N. P. *et al.* The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med*, v. 3, n. 106, p. 106ra106, Oct 26 2011.

MILLER, T. L. *et al.* Isolation of Methanobrevibacter smithii from human feces. *Appl Environ Microbiol*, v. 43, n. 1, p. 227-32, Jan 1982.

MINOT, S. *et al.* Hypervariable loci in the human gut virome. *Proc Natl Acad Sci U S A*, v. 109, n. 10, p. 3962-6, Mar 6 2012.

MINOT, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res*, Aug 31 2011.

MINOT, S. *et al.* Conservation of Gene Cassettes among Diverse Viruses of the Human Gut. *PLoS One*, v. 7, n. 8, p. e42342, 2012.

MOROWITZ, M. J. *et al.* Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc Natl Acad Sci U S A*, v. 108, n. 3, p. 1128-33, Jan 18 2011.

MURDOCH, T. B. *et al.* Pattern recognition receptor and autophagy gene variants are associated with development of antimicrobial antibodies in Crohn's disease. *Inflamm Bowel Dis*, v. 18, n. 9, p. 1743-8, Sep 2012.(NCHS)., C. F. D. C. A. P. C. N. C. F. H. S. National Health and Nutrition Examination Survey Questionnaire. . Hyattsville, MD.: Centers for Disease Control and Prevention (CDC). 2011.

(NCHS)., C. F. D. C. A. P. C. N. C. F. H. S. National Health and Nutrition Examination Survey Questionnaire. . Hyattsville, MD.: Centers for Disease Control and Prevention (CDC). 2011.

NEEFS, J. M. *et al.* Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Res*, v. 21, n. 13, p. 3025-49, Jul 1 1993.

OTT, S. J. *et al.* Fungi and inflammatory bowel diseases: Alterations of composition and diversity. *Scand J Gastroenterol*, v. 43, n. 7, p. 831-41, 2008.

PALMER, C. *et al.* Rapid quantitative profiling of complex microbial populations. *Nuc. Acids Res.*, v. 10, p. e5, 2006.

POLZ, M. F.; CAVANAUGH, C. M. Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, v. 64, n. 10, p. 3724-3730, Oct 1998.

PRICE, M. N. *et al.* FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*, v. 26, n. 7, p. 1641-1650, Jul 2009.

PURUSHE, J. *et al.* Comparative genome analysis of Prevotella ruminicola and Prevotella bryantii: insights into their environmental niche. *Microb Ecol*, v. 60, n. 4, p. 721-9, Nov 2010.

REYES, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*, v. 466, n. 7304, p. 334-8, Jul 15 2010.

REYES, A. *et al.* Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol*, v. 10, n. 9, p. 607-17, Aug 6 2012.

SAMUEL, B. S. *et al.* Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc Natl Acad Sci U S A*, v. 104, n. 25, p. 10643-8, Jun 19 2007.

SARTOR, R. B. Microbial influences in inflammatory bowel diseases. *Gastroenterology*, v. 134, n. 2, p. 577-94, Feb 2008.

SCANLAN, P. D. *et al.* Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol*, v. 44, n. 11, p. 3980-8, Nov 2006.

SCUPHAM, A. J. *et al.* Abundant and diverse fungal microbiota in the murine intestine. *Appl Environ Microbiol*, v. 72, n. 1, p. 793-801, Jan 2006.

SEGATA, N. *et al.* Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*, v. 13, n. 6, p. R42, 2012.

SEIBOLD, F. *et al.* Anti-Saccharomyces cerevisiae antibodies in inflammatory bowel disease: a family study. *Scand J Gastroenterol*, v. 36, n. 2, p. 196-201, Feb 2001.

STAMS, A. J.; PLUGGE, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol*, v. 7, n. 8, p. 568-77, Aug 2009.

STEWART, J. A. *et al.* Carriage, quantification, and predominance of methanogens and sulfate-reducing bacteria in faecal samples. *Lett Appl Microbiol*, v. 43, n. 1, p. 58-63, Jul 2006.

TANG, W. H. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med*, v. 368, n. 17, p. 1575-84, Apr 25 2013.

THE HMP CONSORTIUM. Structure, function and diversity of the healthy human microbiome. *Nature*, v. 486, n. 7402, p. 207-14, Jun 14 2012.

TOURNA, M. *et al.* Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci U S A*, v. 108, n. 20, p. 8420-5, May 17 2011.

TURNBAUGH, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature*, v. 457, n. 7228, p. 480-4, Jan 22 2009a.

TURNBAUGH, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, v. 444, n. 7122, p. 1027-31, Dec 21 2006.

URSELL, L. K. *et al.* The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol*, v. 129, n. 5, p. 1204-8, May 2012.

VAN GYLSWYK, N. O. Succiniclasticum ruminis gen. nov., sp. nov., a ruminal bacterium converting succinate to propionate as the sole energy-yielding mechanism. *Int J Syst Bacteriol*, v. 45, n. 2, p. 297-300, Apr 1995.

VAN NOOD, E. *et al.* Duodenal infusion of donor feces for recurrent Clostridium difficile. *N Engl J Med*, v. 368, n. 5, p. 407-15, Jan 31 2013.

WANG, Q. *et al.* Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, v. 73, n. 16, p. 5261-5267, Aug 2007.

WHITMAN, W. B. *et al.* Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A*, v. 95, p. 6578-6583, 1998.

WHITTAKER, R. H. Evolution and Measurement of Species Diversity. *Taxon*, v. 21, n. 2/3, p. 213-251, may 1972 1972.

WILLETT, W. C. *et al.* Validation of a dietary questionnaire with plasma carotenoid and alpha-tocopherol levels. *Am J Clin Nutr*, v. 38, n. 4, p. 631-9, Oct 1983.

WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*, v. 74, n. 11, p. 5088-90, Nov 1977.

WREDE, C. *et al.* Archaea in symbioses. *Archaea*, v. 2012, p. 596846, 2012.

WU, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science*, v. 334, n. 6052, p. 105-8, Oct 7 2011.

WUTHRICH, M. *et al.* Adaptive immunity to fungi. *Annu Rev Immunol*, v. 30, p. 115-48, 2012.

XU, J. *et al.* Evolution of Symbiotic Bacteria in the Distal Human Intestine. *PLoS Biol.*, v. 5, n. 7, p. e156, Jun 19 2007.

YANG, S. K. *et al.* Epidemiology of inflammatory bowel disease in Asia. *Inflammatory bowel diseases*, v. 7, n. 3, p. 260-270, Aug 2001.

YATSUNENKO, T. *et al.* Human gut microbiome viewed across age and geography. *Nature*, v. 486, n. 7402, p. 222-7, Jun 14 2012.

APPENDIX 1 : SUPPLEMENTAL DATA FOR CHAPTER 2

Table 12: Summary of samples used in the storage and purification methods study. DNA was amplified using the BSF8 / BSR357 16S primer pair, purified in a gel, and sequenced in the reverse direction using the FLX platform.

Sample	Subject	Storage method	DNA extraction method	Reads passing QC
A1	1001	frozen	qiagen	8992
A2	1001	frozen (secondary site)	qiagen	7584
A3	1001	frozen	mobio	9501
A4	1001	4C24h	qiagen	7006
A5	1001	4C48h	qiagen	8632
A6	1001	PSP48h	PSP	12804
A7	1002	frozen	qiagen	8454
A8	1002	frozen (secondary site)	qiagen	9241
A9	1002	frozen	mobio	8262
A10	1002	4C24h	qiagen	8558
A11	1002	4C48h	qiagen	10375
A12	1002	PSP48h	PSP	18647
A13	1002	frozen	qiagen (70C)	10075
A14	1003	frozen	qiagen	8806
A15	1003	frozen (secondary site)	qiagen	7324
A16	1003	frozen	mobio	6699
A17	1003	4C24h	qiagen	5130
A18	1003	4C48h	qiagen	733
A19	1003	PSP48h	PSP	8620
A20	1004	frozen	qiagen	10613
A21	1004	frozen (secondary site)	qiagen	8474
A22	1004	4C24h	qiagen	7804

A23	1004	4C48h	qiagen	7751
A24	1004	PSP48h	PSP	8875
A25	1004	fresh	qiagen	7869
A26	1006	frozen	qiagen	7989
A27	1006	frozen (secondary site)	qiagen	7641
A28	1006	4C24h	qiagen	8115
A29	1006	4C48h	qiagen	6715
A30	1006	PSP48h	PSP	8861
A31	1007	frozen	qiagen	7577
A32	1007	frozen	qiagen	8264
A33	1007	4C24h	qiagen	6221
A34	1007	4C48h	qiagen	6058
A35	1007	PSP48h	PSP	9953
A36	1008	frozen	qiagen	8431
A37	1008	frozen (secondary site)	qiagen	8079
A38	1008	4C24h	qiagen	9203
A39	1008	4C48h	qiagen	8033
A40	1008	PSP48h	PSP	9656
A41	1008	fresh	qiagen	7944
A42	1009	frozen	qiagen	7536
A43	1009	frozen (secondary site)	qiagen	8434
A44	1009	4C24h	qiagen	8233
A45	1009	4C48h	qiagen	7989
A46	1009	PSP48h	PSP	8347
A47	1010	frozen	qiagen	5438
A48	1010	frozen (secondary site)	qiagen	8266

A49	1010	4C24h	qiagen	8717
A50	1010	4C48h	qiagen	8634
A51	1010	PSP48h	PSP	9046
A52	1010	fresh	qiagen	7149
A53	1011	frozen	qiagen	8367
A54	1011	frozen (secondary site)	qiagen	7346
A55	1011	4C24h	qiagen	8860
A56	1011	4C48h	qiagen	7781
A57	1011	PSP48h	PSP	7457
			Total	473169

APPENDIX 2 : NUTRIENT CLUSTER COMPOSITION AS USED ON CHAPTER 4

Table 13: Nutrients detected in the recent diet and their cluster membership.

Nutrient	cluster
Cholesterol to Saturated Fatty Acid Index	1
Energy (kcal)	1
Glycemic Load - bread reference	1
Glycemic Load - glucose reference	1
MUFA gadoleic acid	2
MUFA oleic acid	2
Percent Calories from Fat	2
Percent Calories from MUFA	2
SFAarachidic acid	2
SFAbehenic acid	2
Total Fat	2
Total Monounsaturated Fatty Acids MUFA	2
Added Sugars	3
Available Carbohydrate	3
Fructose	3
Glucose	3
Percent Calories from Carbohydrate	3
Sucrose	3
Total Carbohydrate	3
Total Sugars	3
Alanine	4

Animal Protein	4
Arginine	4
Aspartic Acid	4
Cystine	4
Glutamic Acid	4
Glycine	4
Histidine	4
Isoleucine	4
Leucine	4
Lysine	4
Methionine	4
Nitrogen	4
Percent Calories from Protein	4
Phenylalanine	4
Proline	4
Serine	4
Threonine	4
Total Protein	4
Tryptophan	4
Tyrosine	4
Valine	4
X3 Methylhistidine	4
Beta Cryptoxanthin	5
Copper	5
Coumestrol	5

Inositol	5
Insoluble Dietary Fiber	5
Magnesium	5
Manganese	5
Natural Folate food folate	5
Oxalic Acid	5
Pectins	5
Phytic Acid	5
Potassium	5
Soluble Dietary Fiber	5
Total Dietary Fiber	5
Vegetable Protein	5
Vitamin C	5
Alcohol	6
Alpha Carotene	6
Beta Carotene	6
Beta Carotene Equivalents	6
Betaine	6
Caffeine	6
Lutein Zeaxanthin	6
Percent Calories from Alcohol	6
Total Vitamin A Activity I	6
Total Vitamin A Activity II	6
Total Vitamin A Activity IU	6
Vitamin K	6

Water	6
Cholesterol	7
Choline	7
MUFA erucic acid	7
PUFA arachidonic acid	7
PUFA DHA	7
PUFA DPA	7
PUFA EPA	7
PUFA parinaric acid	7
Selenium	7
CLA (cis 9,trans 11)	8
CLA (trans 10,cis 12)	8
MUFA myristoleic acid	8
MUFA palmitoleic acid	8
Percent Calories from SFA	8
SFA margaric acid	8
SFA palmitic acid	8
SFA stearic acid	8
Total Conjugated Linoleic Acid (CLA 18:2)	8
Total Saturated Fatty Acids SFA	8
TRANS fats (16:1)	8
Delta Tocopherol	9
Gamma Tocopherol	9
Omega 3 Fatty Acids	9
Percent Calories from PUFA	9

Polyunsaturated to Saturated Fat Ratio	9
PUFA linoleic acid	9
PUFA linolenic acid	9
Total Polyunsaturated Fatty Acids PUFA	9
Acesulfame Potassium	10
Ash	10
Aspartame	10
Biochanin A	10
Calcium	10
Erythritol	10
Formononetin	10
Galactose	10
Lycopene	10
Maltose	10
Mannitol	10
Phosphorus	10
Sodium	10
Sorbitol	10
Starch	10
Sucralose	10
Tagatose	10
Xylitol	10
Dietary Folate Equivalents	11
Iron	11
Lactose	11

Niacin vitamin B3	11
Niacin Equivalents	11
Pantothenic Acid	11
Retinol	11
Riboflavin vitamin B2	11
Synthetic Folate folic acid	11
Thiamin vitamin B1	11
Total Folate	11
Vitamin B12	11
Vitamin B6	11
Vitamin D	11
Zinc	11
Beta Tocopherol	12
Natural Alpha Tocopherol	12
Synthetic Alpha Tocopherol	12
Total Alpha Tocopherol Equivalents	12
Vitamin E IU	12
Vitamin E mg	12
SFA butyric acid	13
SFA capric acid	13
SFA caproic acid	13
SFA caprylic acid	13
SFA lauric acid	13
SFA myristic acid	13
Glycemic Index - bread reference	14

Glycemic Index - glucose reference	14
Total Trans Fatty Acids	14
TRANS fats (18:1)	14
TRANS fats (18:2)	14
Daidzein	15
Genistein	15
Glycitein	15
Pinitol	15

Table 14: Nutrients detected in the usual diet and their cluster membership.

Nutrient	Abreviation	Cluster
Betaine w/o suppl., choline derivative	betaine_wo	1
Betaine, choline derivative	betaine	1
Calories	calor	1
Glycemic Load	gl	1
Lycopene	lyco	1
Sodium	sodium	1
Alanine	ala	2
Arginine	arg	2
Asparate	asp	2
Glutamate	glut	2
Glycine	gly	2
Histidine	hist	2
Isoleucine	iso	2
Leucine	leu	2

Lysine	lys	2
Methionine	meth	2
Phenylalanine	phenyl	2
Proline	pro	2
Protein	prot	2
Serine	ser	2
Threonine	thr	2
Tryptophan	trypto	2
Tyrosine	tyro	2
Valine	val	2
Animal fat	afat	3
Butyric fatty acid	f40	3
c9,t11 conjug diene isomer 18:2 Linoleic	cla	3
Capric fatty acid	f100	3
Caproic fatty acid	f60	3
Caprylic fatty acid	f80	3
Dairy Fat	dfat	3
Lauric fatty acid	f120	3
Myristic fatty acid	f140	3
Palmitelaidic trans fatty acid	t161	3
Palmitic fatty acid	f160	3
Saturated fat	satfat	3
Stearic fatty acid	f180	3
Monunsaturated fat	monfat	4
Oleic	f181	4

Total Fat	tfat	4
Vegetable fat	vfat	4
Carbohydrates	carbo	5
Eriodictyol, flavonone	uerid	5
Fructose	fruct	5
Glucose	glu	5
Glycemic Index	gid	5
Maltose	malt	5
Sucrose	sucr	5
Total Sugars	sugtot	5
Alpha Carotene	acar	6
AOAC fiber	aofib	6
Beta Carotene	bcar	6
Beta Carotene w/o suppl.	bcar_wo	6
Carotene	carot	6
Copper	cu	6
Copper w/o suppl.	cu_wo	6
Free Choline w/o suppl.	frcho_wo	6
Free Choline, choline-contrib. metabolite	frcho	6
Lutein & Zeaxanthin	lut	6
Magnesium	magn	6
Magnesium w/o suppl.	magn_wo	6
Manganese	mn	6
Manganese w/o suppl.	mn_wo	6
Natural Food Folate	fdfol	6

Oxalatefort. foods	oxal06	6
Phylloquinone Vitamin K1	vitk	6
Phylloquinone Vitamin K1 w/o suppl.	vitk_wo	6
Potassium	k	6
Potassium w/o suppl.	k_wo	6
Retinol Activity Equivalents w/o suppl.	rae_wo	6
Retinol Equiv. of Vit. A w/o suppl.	reteq_wo	6
Total Carotene w/o suppl.	car_wo	6
Total Folate w/o vit. pills	fol98_wo	6
Vitamin A	vita	6
Vitamin A w/o vit. pills	vita_wo	6
Vitamin E w/o vit. suppl.	e02mg_wo	6
Calcium	calc	7
Calcium w/o vit. pills	calc_wo	7
Choline, Glycerophosphocholine	gpcho	7
Choline, Phosphocholine	pcho	7
Cystine	cys	7
Dairy Calcium	dcalc	7
Dairy Protein	dprot	7
Dairy Vitamin D	dvitd	7
Lactose	lact	7
Pantothenic Acid w/o suppl.	pant_wo	7
Phosphorous	ph	7
Phosphorous w/o suppl.	ph_wo	7
Retinol w/o vit. pills	ret_wo	7

Riboflavin B2 w/o vit. pills	b2_wo	7
Vitamin B12 w/o vit. pills	b12_wo	7
Vitamin D w/o vit. pills	vitd_wo	7
Added Germ from wheats	germa	8
Iron	iron	8
Iron w/o vit. pills	iron_wo	8
Niacin w/o suppl.	nia_wo	8
Pyridoxine B6 w/o vit. pills	b6_wo	8
Thiamine B1 w/o vit. pills	b1_wo	8
Vitamin E, Food Fortification	es2mg	8
Zinc	zn	8
Zinc w/o vit. pills	zn_wo	8
Beta Cryptoxanthin	bcryp	9
Hesperetin, flavanone	uhesp	9
Luteolin, flavone	ulutn	9
Naringenin, flavanone	unarg	9
Total flavanones	totusflavanone	9
Total flavones	totusflavone	9
Vitamin C	vitc	9
Vitamin C w/o vit. pills	vitc_wo	9
Folate Equivalents, suppl. & fort. foods	foleq	10
Folic Acid from suppl. & fort. foods	folic	10
Iodine	iodine	10
Niacin	niacin	10
Pantothenic acid	panto	10

Retinol	retinol	10
Retinol Activity Equivalents	rae	10
Retinol Equivalents of Vit A	reteq	10
Total Folate post 1998	fol98	10
Total Vitamin E, suppl. & fort. foods	e02mg	10
Vitamin B1	b1	10
Vitamin B12	b12	10
Vitamin B2	b2	10
Vitamin B6	b6	10
Vitamin D	vitd	10
Alpha + Gamma Linolenic	ag18302	11
Alpha linolenic fatty acid	a1830	11
Alpha Linolenic fatty acid	pfa183n3c02	11
Linoleic	f182	11
Linolenic fatty acid	f183	11
Omega 3	omg02	11
Omega 3 no vit. pill w/ alpha183	omg02_wo	11
Omega 6	pfn602	11
Omega 6 w/o suppl. 2002	pfn602_wo	11
Omega 6, no gamma	n602	11
Omega 6, no gamma 18:3 nosuppl.	n602_wo	11
Polyunsaturated fat	poly	11
Tot Omega 3 w/o suppl.	pfn302_wo	11
Total Omega 3	pfn302	11
Cholesterol	chol	12

Choline w/o suppl.	choline_wo	12
Choline, Phosphatidylcholine	ptdcho	12
Choline, Phosphatidylcholine w/o suppl.	ptdcho_wo	12
Sum of Betaine & Choline	betchol	12
Total Choline, no betaine	choline	12
Alcohol	alco	13
Apigenin, flavone	uapig	13
Caffeine	caff	13
Catechin, flavan-3-ol	ucat	13
Epicatechin, flavan-3-ol	uec	13
Isorhamnetin, flavonol	uisor	13
Kaempferol, flavonol	ukaem	13
Myricetin, flavonol	umyri	13
Proanthocyanidin, dimers	prodim	13
Proanthocyanidin, monomers	promon	13
Quercetin w/o suppl., flavonol	uquer_wo	13
Quercetin, flavonol	uquer	13
Total flavan-3ols	totusflavan3ol	13
Total flavonoids	totusflav	13
Total flavonols	totusflavonol	13
Total theaflavin & polymers proanthocyanidins	totuspolyflav	13
Animal Protein	aprot	14
Arachadonic fatty acid	f204	14
Choline, Sphingomyelin	sphingo	14
Docosapentaenoic fatty acid (DPA)	f225	14

Eicosenoic fatty acid	f201	14
Gamma linolenic fatty acid	g1830	14
Gamma Linolenic fatty acid	pfg183n6c02	14
Hydroxyproline	hydpro	14
Palmitoleic fatty acid	f161	14
Taurine	tau	14
Docosahexaenoic fatty acid (DHA)	f226	15
Eicosapentaenoic fatty acid (EPA)	f205	15
Long Chain Fatty Acid	lcn302	15
Long Chain Fatty Acid nosuppl.	lcn302_wo	15
Long Chain N3 Fatty Acid 205+225+226	lcn3	15
Long Chain N3 Fatty Acid no vit. pills	lcn3_wo	15
Omega 3,no alpha-linolenic acid	omega	15
11t-Eicosenoic fatty acid	t201	16
Acrylamide	acryl	16
Dihydrophylloquinone Vitamin K1	dvitk	16
Total Trans	trn02	16
Total Trans/Cis Trans Linoleic	t182	16
Trans Oleic fatty acid	t181	16
Added Bran from wheat,rice,oat,corns	brana	17
Bran Scores	bran	17
Germ Scores	germ	17
Natural Brans	brann	17
Natural Germs	germn	17
Whole Grain Amounts	whgrn	17

Whole Grain Score w/o Added Bran & Germs	whsub	17
Aspartamefort. foods	aspart	18
Aspartic Acid, Aspartame	aspa	18
Phenylalanine, Aspartame	phenyla	18
Epicatechin 3-gallate, flavan-3-ol	uecg	19
Epigallocatechin 3-gallate flavan-3-ol	uegcg	19
Epigallocatechin, flavan-3-ol	uegc	19
Flavonoids no prothocyanidins	usflavnopro	19
Gallocatechin, flavan-3-ol	ugcat	19
Theaflavin 3,3 digallate, flavan-3-ol	utfdg	19
Theaflavin 3-gallate, flavan-3-ol	utfg	19
Theaflavin 3-gallate, flavan-3-ol(2)	utf3g	19
Theaflavin, flavan-3-ol	utf	19
Thearubigins, flavan-3-ol	utrg	19
Cyanidin, anthocyanidin	ucyn	20
Delphinidin, anthocyanidin	udlp	20
Malvidin, anthocyanidin	umlv	20
Pelargonidin, anthocyanidin	upel	20
Peonidin, anthocyanidin	upeo	20
Petunidin, anthocyanidin	upet	20
Proanthocyanidin, 4-6mers	pro46	20
Proanthocyanidin, 7-10mers	pro710	20
Proanthocyanidin, polymers	propoly	20
Proanthocyanidin, trimers	protrim	20
Total anthocyanidins	totusanth	20

APPENDIX 3 : PEER-REVIEWED PAPERS PUBLISHED DURING DEGREE COMPLETION

All published papers are freely available online. The corresponding hyperlink follows each citation.

Hoffmann, C; Dollive, S; Grunberg, S; Chen, J; Li, H; Wu, GD; Lewis, JD; Bushman, FD. Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *Plos One*. 8:e66019. 2013.
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0066019

Dollive, S; Chen, Y; Grunberg, S; Bittinger, K; **Hoffmann, C**; Vandivier, L; Cuff, C; Lewis, JD; Wu, GD; Bushman, FD. Fungi of the murine gut: episodic variation and proliferation during antibiotic treatment. *PLoS ONE.* 8:e71806. 2013.
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0071806

Latulippe, ME; Meheust, A; Augustin, L; Benton, D; Berč; Birkett, A; Eldridge, AL; Faintuch, J; **Hoffmann, C**; Miller Jones, J; Kendall, C; Lajolo, F; Perdigon, G; Antonio Prieto, P; Rastall, RA; Sievenpiper, JL; Slavin, J; Wenzel De Menezes, E. ILSI Brazil International Workshop on Functional Foods: a narrative review of the scientific evidence in the area of carbohydrates, microbiome, and health. *Food & Nutrition Research*, 57:19214, 2013.
http://www.foodandnutritionresearch.net/index.php/fnr/article/view/19214/25929

Mei, J; Liu, Y; Dai, N; **Hoffmann, C**; Hudock, KM; Zhang, P; Guttentag, SH; Kolls, JK; Oliver, PM; Bushman, FD; Worthen, GS. Cxcr2 and Cxcl5 regulate the IL-17/G-CSF axis and neutrophil homeostasis in mice. *The Journal of Clinical Investigation*. 122:JCI6058. 2012.
http://www.jci.org/articles/view/60588

Chen, J; Bittinger, K; Charlson, E. S; **Hoffmann, C**; Lewis, J; Wu, GD; Collman, RG; Bushman, FD; Li, H. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*. 28:2106-2113, 2012.
http://bioinformatics.oxfordjournals.org/content/28/16/2106.full

Dollive, S; Peterfreund, GL; Sherrill-Mix, S; Bittinger, K; Sinha, R; **Hoffmann, C**; Nabel, C; Hill, DA; Artis, D; Bachman, MA; Custers-Allen, R; Grunberg, S; Wu, GD; Lewis, JD; Bushman, FD. A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biology*. 13:R60, 2012.
http://genomebiology.com/content/13/7/R60

Wu, GD; Chen, J; **Hoffmann, C**; Bittinger, K; Chen, Y; Keilbaugh, SA; Bewtra, M; Knights, D; Walters, WA; Knight, R; Sinha, R; Gilroy, E; Gupta, K; Baldassano, R; Nessel, L; Li, H; Bushman, FD; Lewis, JD. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*. 334:105-108, 2011.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3368382/

Wu, GD; Lewis, JD; **Hoffmann**, C; Chen, Y; Knight, R; Bittinger, K; Hwang, J; Chen, J; Berkowsky, R; Nessel, L; Li, H; Bushman, FD . Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiology*. 10:206, 2010.
http://www.biomedcentral.com/1471-2180/10/206

EPILOGUE

Quamvis procrastinatio ampla fuerit, sicut Iulius dixit:

Veni, Vidi et Vici.