Norwegian University
of Life Sciences

Master Thesis 2016 60 ECTS
Faculty of Veterinary Medicine and Biosciences
Department of Chemistry, Biotechnology and Food Science

The potential role of tap water bacteria in inflammatory bowel disease

Christine Thorsrud
MSc Microbiology

Acknowledgements

First, I would like to express my gratitude towards my primary supervisor Professor Knut Rudi of the MiDiv Lab at the Norwegian University of Life Sciences. I would like to thank Knut for the sincere devotion he has dedicated to his job as my supervisor and for always making me feel welcome if in need for advice. Whenever in times of feeling disoriented and overwhelmed with work, Knut has always found a way to guide me in the right direction and spark my motivation. His encouragement and profound passion for microbiology has inspired me to produce a Master`s thesis I am proud of.

Second, I would like to give my deepest appreciation to the work my co-supervisor Jane Ludvigsen, PhD Candidate of the MiDiv Lab has dedicated to my thesis. Her support in the laboratory, in addition to the help provided with computer analysis have been of immense value. Together, my supervisors have provided me with invaluable knowledge within the field of microbial research. Thanks to their help, the shape of my learning curve has been under constant acceleration during the past year.

I would also like to thank Post-doctor Ekaterina Avershina of the MiDiv Lab for her essential contributions to my thesis. I would further like to direct my courtesies to the members of the MiDiv team as a whole. They have not only inspired me to learn a lot, but most importantly, they have made me come to the acknowledgement that there is a lot left to be learned. Their genuine enthusiasm for the field has inspired me to continue striving for improvements as a microbial researcher.

Finally, I want to express my deepest gratitude to Morten H. Vatn, Professor of Clinical Epidemiology of Akershus University Hospital for his contributions of valuable material for analysis. I will forever be grateful and feel humbled by the responsibilities I have been trusted with respect to this material. This thesis could not have been accomplished without his generosity.

Ås, May 2016

Christine Thorsrud

Abstract

Microbial dysbiosis is implicated to play a substantial part in several pathophysiological processes, and one disease receiving great attention in recent years owing to its affiliation with an abnormal microbial state of the gut is inflammatory bowel disease. Being a multifactorial disease, other factors implicated to contribute to its commence include genetics, components of the immune system and environmental factors. Albeit suggestions of tap water serving as an environmental trigger in the aetiology of IBD has been made, its potential impact on the gastrointestinal microbiota remains an untouched area of investigation. Thus, in this study we sought to investigate associations of tap water on the microbiota of gastrointestinal mucosa that could substantiate research conducted to unveil environmental and microbial factors contributing to the onset and/or maintenance of this disease. A total of 426 biopsies and 227 water samples retrieved from 129 and 223 patients respectively, served as material for analysis, and included both adult and pediatric patients from Norwegian IBD and control cohorts. The V3-V4 region of the 16S ribosomal ribonucleic acid gene was amplified using a nested approached to polymerase chain reaction, and sequenced by use of the Illumina MiSeq sequencing platform. Our findings exposed significant associations between tap water and biopsies with respect to an operational taxonomic unit belonging to *Enterobacteriaceae* at a p-value of 0,016 using Fisher exact as statistical approach. We further disclosed highly significant increases of the same OTU in pediatric IBD sufferers, especially in the ulcerative colitis cohort compared to cohorts of both age groups. This gave a p-value <0,05 when pairwise comparisons with the Conover-Inman method was employed on the median percentagewise prevalence of this OTU. Further analysis by Conover-Inman test also revealed augmented levels of this OTU in biopsies of inflamed origin compared to biopsies of normal state at a p-value of 0,000. Thus, our results serve as important contributors to research on the environmental aspects of IBD, and also with respect to the role of *Enterobacteriaceae* as a potential microbial key player in the onset and/or maintenance of this disease.

Abstrakt

Mikrobiell dysbiose er implisert å spille en vesentlig rolle i flere patofysiologiske prosesser, og en sykdom som de senere år har mottatt stor oppmerksomhet på grunn av sin assosiasjon med anormale mikrobielle forhold i tarmen er inflammatorisk tarmsykdom. Som en multifaktoriell sykdom, antas det at også genetikk, immunologiske komponenter og miljøpåvirkninger medvirker til dens oppblomstring. På tross av forslag om drikkevann som en mulig miljøtrigger i etiologien av IBD, er dens påvirkning på gastrointestinal mikrobiota forblitt et relativt urørt forskningsområde. Derfor ønsket vi i denne studien å undersøke sammenhenger mellom drikkevann og den gastrointestinale mikrobiotaen i mukosa, og bidra med avdekkingen av miljømessige og mikrobielle faktorer som kan medvirke til oppblomstringen og/eller opprettholdelsen av denne sykdommen. Totalt 426 biopsier og 227 vannprøver fra 129 og 223 pasienter ble benyttet som analysemateriale og inkluderte prøver fra både voksne og barn fra en norsk IBD og kontroll kohort. V3-V4 regionen av 16S rRNA genet ble amplifisert ved å bruke en nestet tilnærming til polymerase kjedereaksjon, og sekvensert ved å bruke Illumina MiSeq som sekvensplattform. Vi avdekket signifikante sammenhenger mellom drikkevann og biopsier når det kom til en operasjonell taksonomisk enhet tilhørende *Enterobacteriaceae* med en p-verdi på 0,016 når Fisher exact ble benyttet som statistisk tilnærming. Vi fant og signifikante økninger av den samme OTUen i barn med IBD, da spesielt i ulcerøs kolitt kohorten sammenlignet med kohorter av begge aldersgrupper. Dette viste en p-verdi <0,05 når parvise sammenligninger med Conover-Inman av medianen av denne OTUens prosentvis prevalens ble benyttet. Videre analyse med Conover-Inman avdekket og økte mengder av denne OTUen i inflammert vev sammenlignet med normalt vev med en p-verdi på 0,000. Våre resultater utgjør dermed viktige bidrag i forskningen på det miljømessige aspektet av IBD, og også i forskning som omhandler *Enterobacteriaceae* som en potensiell nøkkelbakterie i oppblomstringen og/eller opprettholdelsen av denne sykdommen.

Abbreviations

AIEC – Adherent invasive Escherichia coli

ARG16L1 – β2-adrenogenicreceptor 16L1

bp – base pairs

BLAST – Basic local alignment search tool

CD – Crohn`s disease

DC – Dendritic cell

DDH – DNA-DNA hybridization

ddNTP – Dideoxynucleotide triphosphate

DNA – Deoxyribonucleic acid

dsDNA – double stranded DNA

G-/G+ - Gram positive/ gram negative

GI – Gastro intestinal

GNP – Gross national product

GWAS – Genome wide association studies

IBD - Inflammatory bowel disease

MAP – *Mycobacterium avium* subspecies *paratuberculosis*

M-cells – Microfold-cells

MLST – Multilocus sequence typing

mRNA – messenger ribonucleic acid

Muc2 – Mucin2

NGS – Next generation sequencing

NKT – natural killer T-cell

NOD – Nucleotide-binding oligomerization domain

OTU – Operational taxonomic unit

PCoa – Principal coordinates analysis

PCR – Polymerase chain reaction

PRR – Pattern recognition receptor

QIIME – Quantitative insight into microbial ecology

QPCR – Quantitative polymerase chain reaction

ROS – Reactive oxygen species

rRNA – ribosomal ribonucleic acid

SBS – Sequencing by synthesis

SCFA – Short chain fatty acids

SRB – Sulphate Reducing Bacteria

T-cell – Thymus-cell

TLR – Toll-like receptors

UC – Ulcerative colitis

Contents

1.0 Introduction……1

1.1 Human gut microbiota……..1
1.1.1 Environmental significance in shaping the gut microbiota……1
1.1.2 Spatial composition and diversity of gut microbes……2
1.1.3 Is there a microbial congruity between individuals?.....3
1.1.4 Gut microbial influence on human health……..4
1.1.5 Gut homeostasis and immunologic tolerance……4

1.2 Inflammatory bowel disease……..…..5
1.2.1 Disturbation of gut homeostasis and immunologic tolerance in IBD……6
1.2.2 Susceptibility genes……7
1.2.3 Microbial diversity and composition in IBD……..8
The linkage between microbial dysbiosis and IBD……...8
Spatial arrangement of the gut bacteria in IBD-patients……8
Microbiome of the IBD-microbiota……...10
1.2.4 Environmental influence……11

1.3 Tap water and its significance on human health…….....11
1.3.1 Distribution systems as important microbial reservoirs……11
1.3.2 Is there a role for tap water in the aetiology of IBD?.....12

1.4 Obtaining DNA for metagenomics analysis……...13
1.4.1 Challenges when extracting prokaryotic DNA from gut biopsies……...14

1.5 Prokaryotic markers for taxonomic assignment……..……14
1.5.1 The 16S rRNA gene……...……....15
1.5.2 Prokaryotic species definition……15

1.6 Polymerase Chain Reaction…….……...16
1.6.1 Quantitative PCR……17
1.6.2 Quandaries associated with PCR of gut biopsies……17

1.7 DNA sequencing……...……18
1.7.1 First generation sequencing……18
1.7.2 Second generation sequencing……18
NGS by the Illumina platform……19
1.7.3 Third generation sequencing……21

1.8 Sequence analysis through QIIME…… 21
1.8.1 Upstream analysis……22
Preprocessing of input data……22
OTU designation……22
1.8.2 Downstream analysis……...23
Intragroup diversity analysis……23
Intergroup diversity analysis……23

1.9 Aim of project……24

2.0 Materials and methods……..25

2.1 Study material……25

2.2 Obtainment of DNA and quality assurance……28
2.2.1 Cell lysis and extraction of DNA……..28
2.2.2 Quantification of prokaryotic DNA……..28
2.2.3 Quality assurance……...29

2.3 Amplicon library preparation……...29
2.3.1 Nested PCR……...30
2.3.2 PCR product purification……...31
2.3.3 Sequencing preparations……31
2.3.4 Library denaturation and MiSeq sequencing……32

2.4 Analysis of sequencing data…….....33
2.4.1 Analyses in QIIME…….....33
2.4.2 Statistical analysis of datasets……33
2.4.3 Analysis of associations between OTUs in water and biopsies……..34
Identifying and selecting matches……...34
Statistical testing with Fisher exact……35

3.0 Results……36

3.1 Library preparation……...36
3.1.1 Illumina sequencing……..38

3.2 Sequence processing……...38
3.2.1 Intragroup diversity analysis……38

3.3 Statistical testing of biopsy and water sample data set……...……41
3.3.1 Statistical analysis of biopsy data set……41
Testing for interactions and group differences……..……41
OTU4 in inflammation and age……43
3.3.2 Statistical testing of water sample data set……44

3.4 Overlapping OTUs between water and biopsy data set……..44

3.5 Associations between water and biopsy OTUs……45
3.5.1 Matches determined by water and biopsy data set……46
3.5.2 Matches determined by Jukes-Cantor data set……..47
3.5.3 Match 4/2 with respect to diagnosis……..50

4.0 Discussion……...……52

4.1 Transmission of Enterobacteria from water to mucosa?.....52
4.1.1 Tap water as a potential causative agent for the precedence of OTU 4?.....52
Enterobacteria in drinking water……...52

4.1.2 A direct transmission of OTU 4 from water to mucosa?.....53
4.2 Enterobacteria in IBD.....54
4.2.1 Where do we stand so far?.....54
4.2.2 Potential mechanisms for enterobacterial thrift in IBD.....54
4.2.3 Enterobacteria and its preeminence in pediatric IBD.....54
4.2.4 Is there a connection to the extent of inflammation?.....55
4.2.5 Could the precedence of OTU 4 be explained by AIEC?.....56
4.2.6 Could a potential transmission of OTU 4 be attributed to IBD patients alone?.....57
4.3 Bacterial composition of OTU matches.....58
4.3.1 Proteobacteria and its contributions to the microbiota of drinking water.....58
4.3.2 Tap water in relation to previous research.....60
4.4 Possible transmission of other OTUs from water to mucosa?.....61
4.4.1 OTU matches from Jukes-Cantor dataset.....61
4.4.2 OTU matches from biopsy data set.....61
4.4.3 Could tap water introduce perturbations to the gut microbiota by indirect means?.....62
4.5 Analysis in QIIME.....62
4.5.1 Intragroup diversity analysis.....62
4.6 Library preparation and sequencing.....63
4.7 Critical appraisals and possible artifacts.....64
4.7.1 Technical issues.....64
4.7.2 Research design.....65
4.7.3 Mechanisms of contamination.....66
4.8 Concluding remarks.....67
4.9 Future research.....67

References.....69

Appendix.....79
Appendix A – Overview of samples and patients.....79
Appendix B – Results of alpha diversity estimates in QIIME.....88
Appendix C – Output of Kruskal-Wallis test on age-diagnosis.....89
Appendix D – Output of Conover-Inman test on age-diagnosis.....90
Appendix E – Output of Kruskal-Wallis test on age-inflammation.....90
Appendix F – Output of Conover-Inman test on age-inflammation.....91
Appendix G – Significant results Kruskal-Wallis test on OTUs from biopsies and water samples..92
Appendix H – Results of OTU matching using Jukes-Cantor.....100
Appendix I – Matches from Jukes-Cantor dataset used in Fisher exact testing.....113
Appendix J – Matches from biopsy dataset used in Fisher exact testing.....123
Appendix K – Top 50 water OTUs identified as match by Jukes-Cantor.....132
Appendix L – Results of ASCA-ANOVA analysis.....135

1.0 Introduction

1.1 Human gut microbiota

The gut microbiota comprise a vast and extremely complex community of bacteria, and our understanding of its influence on human health is steadily increasing along with the advancements in microbial technologies. The bacterial number in the human gastrointestinal tract equals a total of 10^{14}, with the colon being the most densely populated. (Biedermann & Rogler 2015) It has for long been recognized that this bacterial community constitutes 10 times as much cells as the number of cells in a human body, (Sekirov et al. 2010) although recent publications a somewhat lower ratio. (Sender 2016) Numerically speaking, this bacterial community collectively possess a number of genes that outcompete the human genome by a hundredfold. (Fava & Danese 2011). Although several studies published in high profile articles still report of the gut microbiota consisting of more than 1000 species, research based on novel methods presents estimates of 100-200 species. Based on this dissension, it has been proposed that a stronger consensus with respect to diversity estimates will be of great importance for further advances in studies concerning microbial composition and function of the human gut. (Avershina & Rudi 2015)

1.1.1 Environmental significance in shaping gut microbiota

Owing to findings of a bacterial community in meconium, it is assumed that colonization of the gut commence in utero before birth. (Jimenez et al. 2008) Several environmental factors such as mode of delivery (Dominguez-Bello et al. 2010), and mode of feeding (Koenig et al. 2011) will subsequently shape this process of colonization until a microbial profile with resemblance to an adult microbiota is reached after 3-5 years. (Rodriguez 2015) Although the adult microbiota is considered to be more resilient than the microbiota of infants due to higher diversity and stability, it is still prone to influences from several environmental factors. (Satokari 2015) This includes nutrition, (Wu et al. 2011) use of antibiotics, (Perez-Cobas et al. 2013) physical exercise, (Clarke et al. 2014) smoking (Biedermann et al. 2013) and aging. (Claesson et al. 2011) Although sparsely studied, our genome is also presumed to have an impact on the bacterial composition, (Satokari 2015) much because of interactions and cross-reactions between metabolites synthesised by bacteria and its host. (Biedermann & Rogler 2015) Whether environmental perturbations will disrupt the stable state depends on the resilience of the microbiota, that is the amount of stress or perturbations the microbiota can tolerate before a new equilibrium state is reached. This is thought to differ between individuals and exert an influence on how susceptible these individuals are to develop

diseases associated with a degraded microbiota, such as inflammatory bowel disease (IBD) (Lozupone et al. 2012)

1.1.2 Spatial composition and diversity of gut microbes.

The bacterial density steadily increases as one move down the lower GI-tract starting with 10^1 microbial cells per gram of content in the stomach, until 10^{12} cells per gram content is reached in the colon. Differences in density can be seen across the GI-tract as well, with the mucosa containing a significantly lower microbial density than the lumen. This spatial increase in microbial density seems to be accompanied by increased diversity. (Sommer & Backhed 2013) The microbial composition of the small intestine divaricates from the microbiota in the large intestine (Berry & Reinisch 2013) while the mucosa associated microbiota is thought to differ from the microbiota of the feces. (Zoetendal et al. 2002) Thus, microbial profiling should ideally include both mucosal and fecal samples. (Satokari 2015) However, most studies seeking to investigate the microbial taxonomy and diversity of the gut seem to employ the latter material for analysis.

Although the microbial composition to a large extent varies between individuals, some conjectures apply for most individuals. Firmicutes and Bacteroidetes seems to be the two most dominating phyla in the fecal gut microbiota, constituting 64% and 23% of the gut microbiota respectively. (Sartor et al. 2012) The microbiota also harbors Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia although these phyla are somewhat less prominent. The most prevalent genera however are *Bacteroides, Faecalibacterium* and *Bifidobacterium,* respectively belonging to the first three phyla. A clear consensus with respect to species composition however, seem to be absent. (Arumugam et al. 2011) Although the most exhaustive research has been done with respect to the bacterial component of the microbiota, it should be mentioned that the GI-tract harbours other microbial residents as well. With an estimated total of 10^{15}, viruses comprise the most immense population of the gut, numerically speaking. (Sheehan et al. 2015) Although greatly outnumbered by bacteria and viruses, several archaea and fungi present itself in the gut as well. In terms of prevalence, the archael genera *Methanobrevibacter* and *Nitrososphaera,* and the fungal genera *Saccharomyces, Candida* and *Cladosporium* constitutes important contributions to the archaeal and fungal load of the gut microbiota. (Hoffmann et al. 2013)

The difference in microbiota in fecal and mucosa becomes particularly evident in a study by Eckburg et al (2005) where they found relatively few sequences belonging to the phyla

Proteobacteria, Actinobacteria, Fusobacteria and Verrumicrobia in the latter material. However, most of the sequences in their study belonged to Firmicutes and Bacteroidetes as which is in concordance with findings from the fecal microbiota. 95% of the Firmicutes belonged to the class Clostridia, where a considerable amount of these bacteria were butyrate producers of the Clostridial clusters IV XIVa and XVI. It has been proposed that the mucosa associated bacteria in the colon is more or less uniform due to the close interaction between host and bacterium. (Zoetendal et al. 2002) However a study of the microbiota of mucosal samples by Frank et al (2007) found that the distribution of several bacterial groups might differ between the gastrointestinal compartments. Amongst their findings were increased abundance of the Actinobacterial phylum and the class Bacilli, and decreased levels of Bacteroidetes and *Lachnospiraceae* in the small intestine compared to colon. Still, most sequences were designated to Firmicutes and Bacteroidetes regardless of anatomical origin, although these bacterial phyla showed less overall diversity in the small intestine. (Frank et al. 2007) Other bacteria that has been proposed to be of increased prevalence in the mucosa includes *A. muciniphila* and several proteobacteria.

1.1.3 Is there a microbial congruity between individuals?

Due to large variations in the taxonomic profiles between individuals, it has been proposed that a functional core microbiome is being shared, rather than a core microbiota, with the latter being more variable. (Turnbaugh et al. 2009) (Lozupone et al. 2012) (Sartor et al. 2012) This is to some extent reflected in a study by Qin et al (2010) where deep metagenomic sequencing of fecal samples from 124 Europeans showed that almost 40% of the genes from each individual overlapped with at least half of the cohort. The idea of a functional stability across individuals has however encountered criticism for not sufficiently taking a possible interplay between phylotype and function into consideration. This is primarily because of the repercussion phylotypes exert on the functional characteristics in the gut, and its potential role as an interface for functionality. (Avershina & Rudi 2013) Furthermore, revelations of core phylogroups belonging to *Lachnospiraceae* by phylogroup-independent approaches provides reinforcement to theories embracing the existence of a core microbiota. (Sekelja et al. 2011) Suggestions of the human gut microbiota allegedly being divided into clusters of enterotypes, each with a characteristic microbial profile, has also emerged. The enterotype is determined by variations in the levels of *Bacteroides*, *Prevotella* and *Ruminococcus*, strengthening ideas of a limiting numbers of community compositions across individuals. (Arumugam et al. 2011)

1.1.4 Gut microbial influence on human health

The assembly of microorganisms is often referred to as its own organ which presents itself with a number of important functions impacting human health. First of all, the gut microbes have the ability to produce an array of important vitamins like vitamin K and several B-vitamins such as B12. (LeBlanc et al. 2013) We are also supplied with other substances of significance, most notably short chain fatty acids (SCFA) such as butyrate resulting from digestion of dietary fiber from certain bacteria. In addition to having anti-inflammatory properties (Tedelind et al. 2007) these fatty acids are the primary energy source for colonocytes, (Thibault et al. 2010) and of importance to the expression of tight-junctions and hence the integrity of the epithelial barrier. (Bordin et al. 2004) The importance of the microbiota on human health becomes particularly evident because of its ability to outcompete potential pathogens for nutrients and attachment sites whilst simultaneously stimulating and developing the gut associated immune system. (Sommer & Backhed 2013) The latter observation is being reflected by gnotobiotic animals having a lesser developed immune system in comparison with non-gnotobiotic counterparts. (Bouskra et al. 2008) A better exploitation of ingested nutrients are also being provided by the microbiota, mainly due to their ability to induce genes in epithelial cells important for digestive processes (Hooper et al. 2001) and by their ability to break down several indigestible sugars. Gnotobiotic animals being dependent on a higher caloric intake than non-gnotobiotic animals in order to retain the same body mass illustrates these observations. (Coates 1973)

There seem to be an increasing acceptance that alterations in the gut microbiota has the potential to exert an influence on several pathophysiological processes. This includes diseases such as inflammatory bowel diseases (IBD) like Crohn`s disease and ulcerative colitis, (Frank et al. 2007) obesity, (Ley et al. 2006) colon cancer (Lupton 2004) and several metabolic diseases such as diabetes. (Alkanani et al. 2014) However, the idea that several psychopathological pathways are affected by an aberrant gut microbiota seems to be accentuating as well, and a possible connection to mental disorders such as anxiety and chronic stress has been presented. (Dinan & Cryan 2013)

1.1.5 Gut homeostasis and immunologic tolerance

A thin layer of several types of epithelial cells is separating the lamina propria with its associated adaptive and innate immune cells, from the myriad of antigens in the intestinal lumen. These epithelial cells include i.a goblet cells, paneth cells, M-cells, enteroendocrine cells and absorptive enterocytes (Maynard et al. 2012) which are being replenished every 2-3 days. (Satokari 2015) In

the colon of healthy individuals, this epithelial cell-lining is fortified by two layers of mucin, produced by goblet cells. The inmost layer formed by Muc2, is the most dense and is virtually sterile due to its immense occurrence of antimicrobial peptides. The outer layer is less dense and serve as an important habitat for many commensals. (Maynard et al. 2012) However, the composition and thickness of this layer is to a large extent dependent on the microorganisms residing inside the GI tract (Sommer & Backhed 2013) and certain pathogens of Fusobacteria and Enterobacteria are able to imperil this protective layer. (Swidsinski et al. 2009)

The baseline for communication between the luminal microbes and epthelial cells and innate immune cells is the pattern recognition receptors (PRR) TLR and NOD, recognizing conserved structures in the microbiota. (Satokari 2015) Epithelial cells in the distant ileum and the colon normally express low amounts of TLR because of their close proximity to luminal microorganisms. (Sartor 2006) Signal mediation through PRR are thought to have an impact on the tolerogenic training of innate immune cells, and is therefore of importance for homeostasis. (Elson & Cong 2012) Dendritic cells (DC) possess the ability to express all the TLR and NODs, permitting them to distinguish between pathogens and commensals (Baumgart & Carding 2007) and under homeostatic conditions, their antigen presentation will promote immunologic tolerance against commensals. (Davies & Abreu 2015) In order for an intestinal homeostasis to be achieved, an intricate and delicate communication between the epithelium and its cellular components on each side must be obtained. (Goll & Granlund 2015) However, if proper controlling of this communication is not established, either as a result from defects in the host or an aberrant microbiota, decreased tolerogenic responses towards commensal bacteria with subsequent inflammations might arise, as hypothesized in IBD patients (Satokari 2015)

1.2 Inflammatory bowel disease

Inflammatory bowel diseases encompass the chronic relapsing disorders Crohn`s disease (CD) and ulcerative colitis (UC) and is characterized by intestinal inflammation, where the severity and localization along the intestine depends on diagnosis. UC is in general confined to the colon with ulcers and inflammation of the mucosal layer being characteristic symptoms. Goblet cells are often depleted, while micro abscess forming neutrophils often present themselves in large numbers in lamina propria and crypts. CD on the other hand can emerge along the entire GI tract, but is generally restricted to the ileum, where it presents itself as a deep and transmural inflammation,

often in segments. Aggregates of macrophages forming non-caseating granulomas are common histopathological feature of the latter disease. (Davies & Abreu 2015) (Xavier & Podolsky 2007) Despite the fact that the disease course of both CD and UC often alternates between relapse and remission, the anatomical location of inflammation show little signs of variation, although some extensions of inflammation have been observed in the latter diagnosis. (Burisch & Munkholm 2015) Both diseases are to date incurable, although treatment with probiotics, antibiotics, anti-inflammatory and/or immunosuppressive drugs can prove to be supportive. (Sartor & Mazmanian 2012) (Frank et al. 2007)

Albeit extensive researching efforts in order to unveil the causation of IBD has been done, the precise aetiology remains unknown. However, there seem to be an increasing evidence and general acquiescence as regards to IBD being a multifactorial disease, where several factors contribute to its commence. At its core is a deviant interaction between the gut microbiota and the immune system in genetically susceptible hosts, with environmental factors being of importance to the onset and maintenance of disease. (Berry & Reinisch 2013) (Sartor 2006)

1.2.1 Disturbation of gut homeostasis and immunological tolerance in IBD

It has been proposed that a rupture or leakage of the epithelial barrier might serve as the initiation factor of the inappropriate immune response observed in IBD. This might be a result of dissatisfactory replenishment of epithelial cells, ineffective tight junctions (Goll & Granlund 2015) defective mucus barrier (Swidsinski et al. 2009) or an infection of the epithelial barrier, which eventually might expose the immune system to the luminal antigens. (Sheehan et al. 2015) Although the innate immune system is considered to be of great importance to the maintenance of homeostasis, model systems have shown that defects in this part of the immune system alone is not sufficient for developing inflammations. It is thought to be dependent on an adaptive immune response to the microbiota. (Elson & Cong 2012) If commensal bacteria gain access to the underlying mucosal tissue, the DC which under homeostatic conditions would promote immunologic tolerance, might regard these cells as pathogens. Consequently these cells would initiate the differentiation of naive T-cells to effector cells such as Th1, Th2 and Th17, and natural killer T-cells (NKT). (Baumgart & Carding 2007) Activated DC and Macrophages have shown to be increased in IBD patients, as well as the amount of pro-inflammatory cytokines and chemokines. While the activation profiles of innate immune cells is thought to be the same in both UC and DC (Sartor 2006) it is presumed that there is a considerable variation with respect to the T-helper

response. While the immune response in CD is dominated by Th1, the most prevalent effector cell in UC seem to be Th2. Regardless of disease, the T-regs seem to be subordinate to the effector T-cells. (Baumgart & Carding 2007) If the exposure to luminal antigens are of repetitive nature, a loss of tolerance to the gut microbiota and an accumulation of memory T-cells against commensals might arise. (Cammarota et al. 2015)

1.2.2 Susceptibility genes

Genome-wide association studies (GWAS) have revealed 163 loci associated with IBD where 110 are shared between CD and UC. The remaining 30 and 23 loci are distinct for the two diseases respectively. There seem to be a considerable overlap between IBD susceptibility loci and loci associated with several other immune-mediated diseases. (Jostins et al. 2012) Many of these genes are associated with functions of the epithelial barrier, immunoregulation, components of the innate immune system (Sartor & Mazmanian 2012) and dendritic cells (DC). (Davies & Abreu 2015) One of the most eminent susceptibility genes in CD is NOD2, a PRR which initiates the secretion of alpha-defensins in Paneth cells. Impaired NOD2 might lead to the mucosa being more easily invaded. (Cammarota et al. 2015) ARG16L1 and Muc2 on the other hand, are susceptibility genes in both UC and CD where variants of the latter gene might allow for a weakened inner mucus layer and reduced homeostasis (Elson & Cong 2012) Variants of the ARG16L1 on the other hand have shown to give impaired autophagy and exocytosis in Paneth cells. (Goll & Granlund 2015) It is believed that defects in several of the susceptibility genes will have to be present in order to develop IBD (Elson & Cong 2012) The assumption that there is an interaction between several of the genes i.a NOD2 and ARG16L1 might further complicate our understanding of the genetic influence. Interactions might also affect the severity of the diseases. (Sheehan et al. 2015) A family history with IBD is considered to be the primary risk factor for disease development (Baumgart & Carding 2007) and seem to be somewhat stronger for the development of CD than UC. (Xavier & Podolsky 2007) However, research with "induced mutant" mice who developed IBD as a result of either knockout or overexpression of certain genes, has shown a ceasing of the disease as soon as the mutants were made germ free, demonstrating the importance of microbiota in in disease development. (Elson & Cong 2012)

1.2.3 Microbial diversity and composition in IBD

The linkage between microbial dysbiosis and IBD

There seem to be a general acquiescence as regards to IBD patients having an altered microbial composition and a reduced diversity compared to healthy controls, both in fecal and mucosal samples. (Berry & Reinisch 2013) (Sheehan et al. 2015) This is referred to as a dysbiotic microbiota which also tend to exhibit a lower stability than the microbiota of a healthy adult. (Satokari 2015) The microbial dysbiosis is most noticeable when the inflammation is active. (Biedermann & Rogler 2015) Research have shown that diversity can vary between non-inflamed and inflamed areas of the intestines of the same individual, with the latter displaying less alpha-diversity. (Sepehri et al. 2007) Interestingly, it has also been shown that inflammations of the colon might lead to depletion of bacteria in the feces, whilst simultaneously giving increased bacterial concentrations in the crypts. (Swidsinski et al. 2005) Based on findings of the microbiota of UC patients presenting itself with a lower diversity than that of CD patients, as well as different prevalence of certain bacteria, it has been suggested that the bacterial diversity of IBD is disease specific. (Ott et al. 2004) (Swidsinski et al. 2009) The possibility that disease phenotype might exert an influence on the microbial composition and diversity in IBD patients has also been proposed based on findings in a study by Willing et al (2010), showing that the microbial profile of patients with ileal CD differs from patients with colonic CD. Regarding the other microbial residents of the gut, it has been independently shown that CD patients carry an increased fungal diversity (Ott et al. 2008) and higher phage numbers compared to healthy counterparts. (Lepage et al. 2008)

Spatial arrangement of gut bacteria in IBD patients

The microbiota of the mucosa and lumen might be expected to differ. (Frank et al. 2007) According to a study by Gevers et al (2014), some microbial differences between CD patients and healthy controls only became evident when mucosal samples were analyzed as compared to fecal samples. This included a reduction in *Bifidobacteriaceae*, and an increase in *Fusobacteriaceae* and *Enterobacteriaceae*. These observations led to proposals of mucosal bacterias being of greater significance for the aetiology of the disease (Baumgart & Carding 2007) and that IBD to a smaller extent affect the luminal microbiota. (Sheehan et al. 2015)

On a phylum level, the mucosal microbiota of IBD patients in general present itself with a decreased abundance of Firmicutes and Bacteroidetes an increased abundance of Actinobacteria and Proteobacteria. (Frank et al. 2007) Increased levels of the latter phyla includes *Desulfovibrio* in mucosa of UC patients (Rowan et al. 2010) and mucosa associated *Escherichia coli*. Increased

abundance of AIEC are particularly evident in CD patients where it has the potential to invade epithelial cells and replicate intracellularly. (Rolhion & Darfeuille-Michaud 2007) AIEC has further been suggested to be enriched in inflamed tissue in ileal CD as opposed to in normal tissue. (Baumgart et al. 2007) In addition, *Clostridium* (cluster XIV, XVIII, IV) which in a cooperative manner are able to stimulate T-reg cells, (Atarashi et al. 2013) are found to be depleted in IBD patients. (Kabeerdoss et al. 2015) These clusters include several important producers of SCFA such as *C. leptum* (cluster XIVa), *C. coccoides, Roseburia hominis* and *Faecalibacterium prausnitzii* (cluster IV) which are considered to be of great importance to the preservation of immunological balance and gut homeostasis. (Lopetuso et al. 2013) (Satokari 2015)

An impoverished detection of mucosal SCFA-producing bacteria in IBD patients was also revealed in a study by Frank et al (2007) and Willings et al (2010), with the latter study presenting decreased levels of *Faecallibacterium* and *Roseburia* and increased levels of *E.coli* and *R.gnavus* from the *Enterobacteriace* in patients with ileal CD. CD but not UC patients have further been proposed to have increased amounts of *Mycobacterium avium* subspecies *paratuberculosis* (MAP), although these findings seem to vary between projects. (Feller et al. 2007) Depletion of lactic acid bacteria within *Lactobacillus* (phylum Firmicutes) has also been detected in IBD patients. (Ott et al. 2004) *Bacteroides* should normally be found mainly in feces, but adhesive and infiltrating bacteria of this genus has been found in inflamed mucosal tissues of the colon of IBD patients. (Swidsinski et al. 2005) Samples of both colon and small intestine of IBD patients have also proven to be deficient of the *Lachnospiraceae* family compared to healthy subjects. (Frank et al. 2007) When comparing biofilm-formation and bacterial density of the IBD mucosa to healthy counterparts, this is found to be significantly increased, with *B. fragilis* being responsible for the majority of the biofilm. (Swidsinski et al. 2005) Concentration of mucosal bacteria also seem to be positively correlated with disease severity, in both inflamed and non-inflamed colonic tissue. (Swidsinski et al. 2002)

Analysis of fecal microbiota in UC patients has also unveiled a reduced abundance of bacteria involved in SCFA-metabolism such as *R. bromii, Roseburia sp,* and *A. municiphila.* Bacteria of increased prevalence in UC patients included *Fusobacterium* sp. (Rajilic-Stojanovic et al. 2013) where certain strains of this genus possess invasive and proinflammatory properties. (Strauss et al. 2011) Increased numbers of *Helicobacter sp*. and *Campylobacter sp*. has also been found (Rajilic-Stojanovic et al. 2013) Other proteobacteria of exaggerated numbers in feces of IBD patients include the genera *Desulfovibrio* (Loubinoux et al. 2002) which possess toxigenic properties due to its ability to produce pro-inflammatory hydrogen sulphide. (Cammarota et al. 2015)

Microbiome of the IBD microbiota

As opposed to the extensive research that has been conducted on the taxonomic characteristics of the gut microbiota in IBD, research performed with respect to the microbiome are still scarce. However, a study by Morgan et al (2012) seeking to unveil functional perturbations of the IBD microbiome, found shifts in oxidative stress pathways, and a decreased expression of genes related to synthesis of SCFA and amino acids. Several genes involved in pathological processes, most notably adherence invasion and type 2 secretion systems were also found to be increased in patients with ileal CD. They also found an increase in cysteine metabolism along with increased N-acetylgalactosamine transporters, which potentially could indicate an abundance of bacteria metabolizing mucin. (Morgan et al. 2012)

It has been proposed that microbial anomalies observed in IBD could serve as useful biological markers for inflammation activity (Berry & Reinisch 2013) and diagnostic tests for microbial dysbiosis based on deviations from a healthy gut microbiota have already been developed. (Casen et al. 2015) Albeit the linkage between IBD and microbial dysbiosis has been known for long, the question of whether the aberrant microbiota is a cause or consequence of IBD remains unknown. (Baumgart & Carding 2007) (Maynard et al. 2012) (Mukhopadhya et al. 2012) (Sartor et al, 2015)

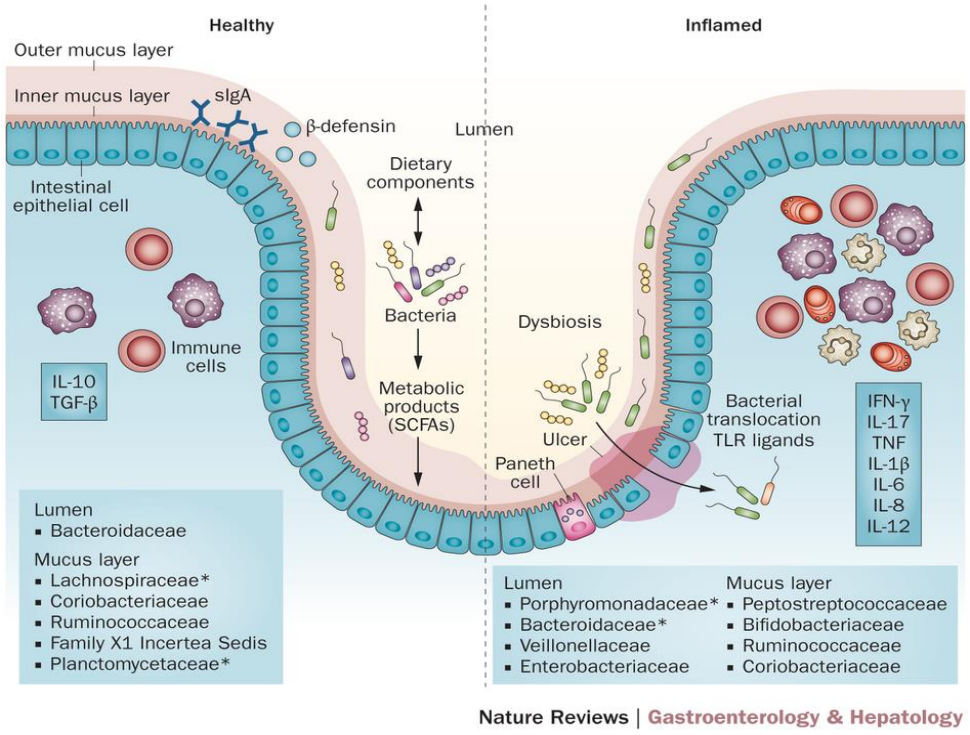


Figure 1.1: The figure gives a simplified illustration of layers of the gut mucosa and the interplay between a subset of the immunological, and microbial factors implicated to contribute in the development of IBD. Picture from (Sartor 2015)

1.2.4 Environmental influence

It has been independently shown that the prevalence of IBD is positively correlated with a nations GNP, (Burisch & Munkholm 2015) and that the number of incidents increases as a nation progresses from having a non-industrialized to an industrialized status. (Sheehan et al. 2015) This is somewhat reflected by the increased numbers of IBD incidents in emigrants from non-industrialized areas when exposed to a modern western lifestyle. (Barreiro-de Acosta et al. 2011) In already industrialized countries however, the prevalence of disease has stabilized. (Ng et al. 2013) It has been suggested that the reduced exposure to microbial antigens in areas with exaggerated hygienic conditions might debilitate the proper maturation of the immune system, and thereby increasing the risk of inappropriate immune responses (Baumgart & Carding 2007) Research has also shown that environmental factors possibly may exert a greater influence in the aetiology of IBD than genetic factors. (Sheehan et al. 2015) This is reflected in a study of monozygotic twins by Halfvarson et al (2003) presenting a concordance rate of <20% and 50% between twins with UC and CD respectively. Environmental and lifestyle factors thought to exert an influence on the development of IBD includes hygiene, microbial exposure, diet, use of antibiotics, pollution, smoking, (Ng et al. 2013) consumption of detergents and emulsifiers (Swidsinski et al. 2009) and water supply. (Frank et al. 2007) (Aamodt et al. 2008)

1.3 Tap water and its significance on human health

1.3.1 Distribution systems as important microbial reservoirs

The drinking water in a country is normally treated in concordance with guidelines established by the respective countries official national guidelines. Norwegian drinking water is treated according to the Drinking Water Act (Drikkevannforskriften, www.lovdata.no) in order to remove contamination of any kind that could pose a threat on consumers health. However, in order to reach the consumer, the water must move through distribution systems where different influential factors might support bacterial growth. This includes parameters such as distribution time, arrangement of the pipes, temperature of the water to be distributed, and the concentration of disinfectant residuals and biodegradable organic matter (Pepper et al. 2015) The presence of bacteria in drinking water is also influenced by the frequency of usage of the tap, and can if not frequently used, give rise to potential human pathogens. (Rudi et al. 2009) The creation of bacterial biofilms on pipe surfaces and bacterial aggregates in the distribution water is of particular concern due to increased resistance to disinfectants (Williams et al. 2004) and better exploitation of available nutrients, thereby

reinforcing bacterial growth. (Pepper et al. 2015) Many of the bacteria in the distribution water are also able to grow with limited availability of nutrients, and thereby posing another major problem in water distribution systems.(Payment et al. 1991) Pathogens that are able to grow in distribution systems include *Legionella* spp, *Aeromonas* spp, *Mycobacterium* spp, and *Pseudomonas aeruginosa*. (Szewzyk et al. 2000) Furthermore, it has been demonstrated that the material of the distribution systems might exert significant impacts on the growth of bacteria such as atypical Mycobacterium (Schwartz et al. 1998) and different strains of Betaproteobacteria. (Kalmbach et al. 2000)

The microbiota of tap water seem to be dominated by Proteobacteria, although what constitutes the most prevalent proteobacterial class seem to vary between research projects. Both Alphaproteopacteria (Williams et al. 2004) and Betaproteobacteria have been proposed to be the most dominating class, and a possible interaction between the two has also been suggested. (Rudi et al. 2010) Within the latter proteobacterial class, several strains from the *Aquabacterium* genus seem to predominate the drinking water in distribution systems, including *A. parvum, A. commune* and *A.citratiphilum.* (Kalmbach et al. 2000) The proteobacterial phyla comprise several heterotrophic pathogens (pathogens using organic nutrients) that can be found in drinking water, such as *Desulfovibrio, Pseudomonas, E.coli, K. pneumoniae, Y. enterocolitica, E. cloacae* and *C. freundi.* (Allen et al. 2004) The presence of potential pathogens in tap water has i.a been shown in a study by Payment et al (1994) where the virulence of heterotrophic bacteria in tap water was investigated. This study found that 57% of the tap water samples contained cultivable cytolytic bacteria, and that 17% of the samples contained cytolytic bacteria possessing both adherent and hemolytic properties, which could give rise to diseases if present in adequate numbers. (Payment et al. 1994)

1.3.2 Is there a role for tap water in the aetiology of IBD?

Few studies have to date investigated the possible association between drinking water and gastrointestinal diseases. One popular theory regarding tap water as an environmental trigger behind IBD, encompassed the plausible association between *Mycobacterium avium* subsp. *paratuberculosis* (MAP) and CD. MAP has earlier been identified as the causative agent of Johne`s disease, a disease similar to CD in cattle. It is regarded as a bacterium that potentially could be transmitted to humans through water, owing to its high persistence in harsh environments and resistance against common chlorine disinfection concentrations used in distribution systems. (Naser

et al. 2014) As of today however, this hypothesis is to some extent regarded as controversial, owing to the lack of consistency between research projects. (Liverani et al. 2014) A study by Payment et al (1991) found a potential link between different gastrointestinal symptoms in Montreal and water supply in terms of the presence of heterotrophic pathogens. A resembling study by Aamodt et al (2008) found an association between water supply in terms of iron content and the prevalence of IBD in Norway. They suggested the potential pathogenicity of iron in the development of IBD in part could be explained by this chemical elements ability to increase oxidative stress and produce reactive oxygen species (ROS). ROS has been proposed to be of significance in the aetiology of IBD due to its ability to depolymerize mucine (Goll & Granlund 2015) Another plausible explanation for the observed association was that iron somehow affected the growth of the gut bacteria by changing the balance of the bacteria present, or increasing their virulence. (Aamodt et al. 2008) It has earlier been recognized that trace elements like iron and aluminium might have the potential to alter bacterial pathogenicity and thereby exacerbating the immune response towards these bacteria. (Perl et al. 2004) Furthermore, an unveilance of highly increased levels of a common drinking water bacterium of the Alphaproteobacteria in mucosal samples of IBD patients, has strengthened ideas regarding the microbiota of drinking water possibly exerting an influence on the development of IBD as well. (Frank et al. 2007)

1.4 Obtaining DNA for metagenomics analyses

Acquisition of bacterial DNA for metagenomics analyses often requires a lysis of bacterial cells. This is commonly achieved by mechanical, enzymatic or chemical means, sometimes applied in a combinatorial fashion. Mechanical lysis frequently involves the use of bead beating and represent to some extent a more rough method of treatment than the two latter options. (Salonen et al. 2010) Given the fact that rough treatment of cells might give more fragmented DNA, it has been proposed that the intensity of lysis should be put into context with the desired purpose of analysis. Shotgun metagenomic analyses will for instance demand longer fragments than metagenomics analyses based on sequencing of the16S rRNA gene. (Nannipieri & Smalla 2006) Ideally, lysis of cells should not be subject to constraints from the morphology of the cells, their growth phase, concentrations or method of preparations. (Zoetendal et al. 2001) Still, enzymatic and chemical methods have encountered criticism for its lack of ubiquity in bacterial targets. (Salonen et al. 2010) and not providing sufficient lysis of G+ cells. (Carbonero et al. 2011) Mechanical disruption also tend to favour lysis of G- over G+ cells, due to the rigidness of the peptidoglycan layer of the latter. (Tortora et al. 2010) However, the degree of cross-binding between peptides in this layer and hence

its strength, will to some extent vary between species and is influenced by growth phase, with cells in growth possessing a weakened layer. Also, size and shape of the cells might exert constraints on the efficiency of lysis by mechanical means, with large and/or rod-shaped cells being more easily ruptured than small and/or cocci-shaped cells. This could subsequently propagate an overrepresentation of easily lysed cells in downstream analyses. (Nannipieri & Smalla 2006)

1.4.1 Challenges when extracting prokaryotic DNA from gut biopsies

The nature of the material to be analysed, provide further implications with respect to what serve as the best method of lysis. Thus, contradistinctions exist as regards to what constitutes the best method of lysis for gut biopsies. It has been argued that chemical and enzymatic lysis should be favoured due to the vast amount of eukaryotic DNA a mechanical disruption will yield. Also, underrepresentation of certain microbial groups such as Sulphate Reducing Bacteria (SRB) and Methanogenic archaea when mechanical lysis of fecal samples was employed, makes it possible to believe that similar taxonomic biases might occur during analysis of gut biopsies. (Carbonero et al. 2011) There are however several studies commending the use of mechanical lysis when analysing gut biopsies, presenting results of smaller biases (Zoetendal et al. 2001) and better lysis of G+ cells such as those belonging to Firmicutes. (Cuiv et al. 2011) In addition, mechanical lysis has been proposed to be the best method of choice, owing to our current lack of understanding of the cell wall composition of bacteria in the gut. (Avershina et al. 2014)

1.5 Prokaryotic markers for taxonomic assignment

Prior to the 1970s, microbial classification was performed with respect to differences in physiological properties, thus giving scarce amounts of groups for microbial annotation. (Pepper et al. 2015) Along with advances in techniques for analysing differences in biological markers between microorganisms, new prokaryotic groups emerged. As of today, several taxonomic markers for phylogenetic classification of prokaryotes exist. Some includes chemotaxonomic markers such as teichoic acids (Fiedler & Schaffler 1987), flavonoids (Emerenciano et al. 2001), phospholipids and fatty acids. (Romano et al. 2000) Other taxonomic markers are based on sequence differences in housekeeping genes. This is a collective term embracing universal genes of vital proteins, such as *rpoB* and *gyrB*, the genes behind a RNA polymerase subunit and DNA gyrase respectively. (Pepper et al. 2015) The most recognized taxonomic marker to date however, is probably the 16S rRNA gene.

1.5.1 The 16S rRNA gene

The16S rRNA gene of the ribosomal operon in prokaryotes encodes for a part of the small ribosomal subunit in prokaryotes. Due to its importance in binding the Shine Dalgarno sequence in mRNA to be translated, the gene sequence is ubiquitous amongst prokaryotes, possess highly conserved properties and is seldom encountered to mutations. (Rajendhran & Gunasekaran 2011) (Willey et al. 2009) This gene contains nine variable regions (V1-V9) interspersed by highly conserved regions. (Mizrahi-Man et al. 2013) allowing for taxonomic assignments in both higher and lower hierarchic levels, (Willey et al. 2009) The conservative regions also allow for design of primers which normally are modified with degenerate positions in order to increase their coverage. Although the 16S rRNA gene extend over approximately 1500 bp, (Rajendhran & Gunasekaran 2011) massive high throughput sequencing technologies is usually limited to sequencing sections of this gene. Apparently, there is little consensus as regards to which of the hypervariable regions that should serve as target, although most studies seem to include V3, V4 or V6. (Mizrahi-Man et al. 2013) However, it has been proposed that the sequencing platform might exert an influence on what serve as the most optimal hypervariable region of choice. (Claesson et al. 2010) Using several different types of primers has also been suggested in order to avoid a possible primer bias and consequently an over or underrepresentation of specific taxa. (Hamady & Knight 2009)

1.5.2 Prokaryotic species definition

The definition of what constitutes a bacterial species has for long been a subject of debate, much due to the genetic elasticity of these organisms. (Pepper et al. 2015) Several approaches aiming at presenting a definition of bacterial species have been proposed, with DNA-DNA hybridization (DDH) being the most acknowledged method prior to the era of sequencing. Species definition by means of DDH involves the designation of two bacteria to the same species if their DNA molecules present a hybridization rate of >70%. (Konstantinidis et al. 2006) However, along with advances in sequencing technologies, sequencing of universal genes, most notably the 16S rRNA gene has become the method of choice for species definition. Here, the taxonomic designation on species level occurs for sequences with $\geq$97% identity, which are clustered into an operational taxonomic unit (OTU). (Pepper et al. 2015) The remaining 3% represent ~45 nucleotides located in so called hypervariable regions of the gene. (Stackebrandt & Goebel 1994) Species definition by means of OTUs has encountered criticism for being too categorical, (Avershina & Rudi 2013) and giving a pre-definition of bacterial species (Sekelja et al. 2011) and not being sufficiently discriminatory. A potential consequence of the latter disadvantage is that bacteria having $\geq$97% sequence similarity in

the gene encoding 16S rRNA, still might be below the threshold of 70% sequence homology if the traditional DNA-DNA hybridization was being used for species definition, and vice versa. (Stackebrandt & Goebel 1994) For this reason, multilocus sequence typing (MLST) has been proposed as an alternative method for taxonomic assignment on a lower hierarchical level. This method includes sequencing of several housekeeping genes with subsequent comparison of the resulting profile to sequence databases. (Pepper et al. 2015)

With that being said, analysis of16S rRNA sequences has not only made it possible to analyse several organisms simultaneously, (Pepper et al. 2015) but it has also circumvented the need for culturing and enabled the study of entire microbial communities in their natural environment. (Rajendhran & Gunasekaran 2011) This includes habitats such as soil and the human gut where it has been estimated that 99% and 60-80% of bacteria from the respective habitats cannot readily be cultivated. (Hirsch et al. 2010) (Suau et al. 1999) Furthermore, with the advent of quantitative PCR, employment of primers specific for the 16S rRNA gene allows for estimates of the total bacterial load in samples, which priory had proven to be difficult. (Pepper et al. 2015)

1.6 Polymerase Chain Reaction

In the mid 1980`s, the traditional Polymerase Chain reaction (PCR) was invented by Kerry Mullis & coworkers and enabled an amplification of DNA by performing three relatively simple steps in a repetitive manner. The first step, melting of dsDNA involves denaturation by an increase of temperature to 94-95^{o}C. The second step, primer annealing, allows for primers to bind to the 3`end of each strand at a temperature that ideally is 2-4^{o}C below the melting temperature of the primers. The final step involves the elongation of DNA at approximately 72^{o}C by a heat-stable polymerase isolated from the thermophilic bacterium *Thermus aquaticus*. The repetitive number of these steps, hereby referred to as cycles, normally differs between 25-40, with the latter cycle number theoretically yielding $>10^{12}$ amplicons from one DNA molecule. Although greatly permitting the study of microorganisms without previous culturing, (Pepper et al. 2015) this method had its limitations as regards to giving the same amount of DNA, independently on the amount of input DNA templates, thus making quantifications difficult. The advent of quantitative PCR has however circumvented this problem. (Kubista et al. 2006)

1.6.1 Quantitative PCR

Quantitative PCR, hereby referred to as qPCR allows for the quantitative detection of products as they are made in real time. Detection is performed by fluorescence reporters, which can be non-specific and sequence specific. (Kubista et al. 2006) An example of the latter is the dual labelled TaqMan probe having a reporter dye in one end and a quencher molecule absorbing the fluorescence emission from the reporter at the other end. When the PCR nuclease degradation separates the molecules, fluorescence is released allowing for the detection of amplicons. Non-specific dyes such as SYBR Green and EvaGreen on the other hand, will emit fluorescence when bound to any dsDNA, but not in its free form. (Giulietti et al. 2001) Quantification of amplicons is enabled by the inclusion of a standard curve with different concentrations of target sequence. (Pepper et al. 2015) When the fluorescence reach a certain threshold for detection, a Ct-value representing the cycle number is registered, and can be used to determine the number of amplicons in the sample. (Bustin et al. 2005) Although non-specific dyes are cheaper than specific dyes, their binding to non-specific PCR products and primer dimers might serve a challenge due to the generation of false positives. (Kubista et al. 2006) QPCR-amplifications by use of these dyes are therefore often ensued by the inclusion of a melting curve where heat is applied in an increasing manner in order to separate all dsDNA in the sample. The following decreases in fluorescence at different temperatures will subsequently serve as indicators of the amount of target amplicons and non-specific products. (Pepper et al. 2015)

1.6.2 Quandaries associated with PCR of gut biopsies

When amplifying bacterial DNA from samples that might possess a high ratio of eukaryotic/prokaryotic DNA, such as gut biopsies, there are several possible complications affecting the outcome of the PCR reaction. First, if the PCR reaction embeds a high amount of non-target eukaryotic DNA, the diffusion of the Taq-polymerase might be hampered, thus impeding the synthesis of DNA. Second, an attempt to account for the low amounts of target DNA by increasing the number of cycles, might lead to an increase in the synthesis of nonspecific products (Kennedy & Oswald 2011) such as chimeras created from several parent sequences, which if undetected, could be regarded as a novel sequence in downstream analysis. (Nelson et al. 2014) Third, low amounts of target DNA are more prone to contamination of DNA degrading substances such as nucleases from skin. (Kennedy & Oswald 2011) Fourth, due to reports of several PCR inhibitors in fecal samples, such as complex polysaccharides (Monteiro et al. 1997) and bile acids (Lantz et al. 1997) it is reasonable to assume that biopsies from the GI tract might include similar inhibitors as well.

Finally, a possible cross-reactivity of prokaryotic primers with eukaryotic DNA might occur. Ideally, primers targeting the 16S rRNA gene result in amplification of prokaryotic DNA. Yet, cross-reactivity with eukaryotic18S rRNA gene has been shown due to the ancestrality of these genes. (Huys et al. 2008) However, modifications of annealing temperature has been proposed to improve the specificity of the primers. (Hwang et al. 2003) In addition, performing a nested approach to PCR, meaning in two consecutive reactions, has been proposed to increase the efficiency, sensitivity and specificity of the reactions. (Ekman 1999)

1.7 DNA sequencing

1.7.1 First generation sequencing

First generation sequencing by means of Sanger sequencing, has for decades been subject to several modifications. Its foundation involves the use of radioactively labelled ddNTP lacking the 3`OH-group, leading to termination of the template extension. This gives a mixture of fragments that when separated by electrophoresis, ultimately can be visualized by autoradiography. (Sanger et al. 1977) The method is considered to deliver readings of relatively good quality and length (1000-2000bp), (Zhang et al. 2011) but has its limitations in regards to being time consuming and yielding a relatively low throughput. The drawbacks of this first generation sequencing method has to some extent been circumvented by the advent of second generation sequencing.

1.7.2 Second generation sequencing

Second generation sequencing, also commonly referred to Next Generation Sequencing (NGS) briefly involves the sequencing of massive number of strands in a parallel fashion. Several NGS sequencing platforms exist, such as Roche 454 pyrosequencing systems, SOLiD, Ion Torrent and Illumina (Rizzo & Buck 2012) with the latter platform possibly comprising the leading platform in terms of usage.

It is often said that the era of NGS emerged with the advent of Roche 454 pyrosequencer and its novel approach to sequencing. In this platform, DNA is fragmented and flanked with adaptors for subsequent attachment to beads. This is succeeded by an emulsion PCR, giving beads covered with a multitude of copies of a single stranded fragment. The beads are then transferred to a plate containing a large amount of wells, and in a repetitive manner exposed to nucleotides, which emit a light following incorporation by the polymerase. This signal is subsequently used for sequence

determination, thus, emanating the principle behind the sequencing by synthesis approach (SBS). Sequencing by use of the SOLiD and Ion Torrent platform, involves the use of DNA binding beads in a manner similar to the Roche 454 pyrosequencer. Albeit the latter platform also utilizing an SBS approach, sequence determination is based on detected decreases in pH followed by nucleotide incorporation, and not emission of light. In a repetitive manner, wells are filled with a solution containing each of the four nucleotides. If the flow of the respective nucleotide results in incorporation, a release of hydrogen ions and a subsequent decrease of pH is detected. (Fisherscientific.com) The SOLiD platform however does not employ an SBS approach to sequencing. Here, beads are attached to a glass slide and exposed to fluorescently labelled probes which will emit fluorescence upon binding to template. This is repeated in a number of cycles, and used for sequence determination. (appliedbiosystems.com)

NGS by the Illumina platform

An Illumina sequencing usually begins with a library preparation, i.e the attachment of adapters flanking the fragments, giving an overhang on each side of the region of interest after PCR. The adapters contain forward or reverse primers, followed by different indices or barcodes enabling sample identification. (Illumina.com) Use of dual indexing reduce the probability of indexes being assigned to the wrong sample in downstream analysis. (Nelson et al. 2014) The distal region of the adapters include sequences complementary to flow-cell oligos. Once the fragment is loaded onto the chip and bound to the oligos, clusters of clonal fragments are made, thereby increasing sequencing depth. This is achieved by repeated amplifications of the fragments that are bound to the oligos in a bridge like manner, interspersed by denaturation of the newly made dsDNA. The reverse strands are washed away, giving clusters of only forward strands. The density of these clusters might affect several sequencing parameters, such as Q30 score, clusters passing filter score, run quality and data output. Obtaining the appropriate density is therefore of great importance to the sequencing results. (Illumina.com)

Reading of the strands are performed by a sequencing by synthesis (SBS) method where the fluorescence of labelled nucleotides are detected while being added to the growing chain. This is done in a parallel fashion for all bound sequences in all the generated clusters. The probability of false base calls is captured by a Q30 score representing the percentage of base calls with an accuracy >99,9%. The emission generated from each of the clusters, is captured between each incorporation and used for the designating the emission to a particular nucleotide based on its

wavelength and intensity. As this optics require diversity between each nucleotide incorporation, phiX (phage DNA) is normally sequenced simultaneously, where the amount depend on the expected nucleotide diversity. (Navas-Molina et al. 2013) The percentage of clear signals from each cluster is represented by a clustering passing filter score, indicating signal quality. The read product and index read generated from reading of forward strand is removed. Again, a bridge amplification is performed to generate a reverse strand so that sequencing of this strand can be performed in a manner similar to that of forward strand. A total of 300 bp is being read, each way. This is referred to as paired end sequencing, increasing the accuracy of the reads. The outcome of this method of sequencing is an immense amount of reads from both forward and reverse strands, which are designated into different groups depending on the combination of indices. Reads are then submitted to an appropriate pipeline for data analysis. (Illumina.com)

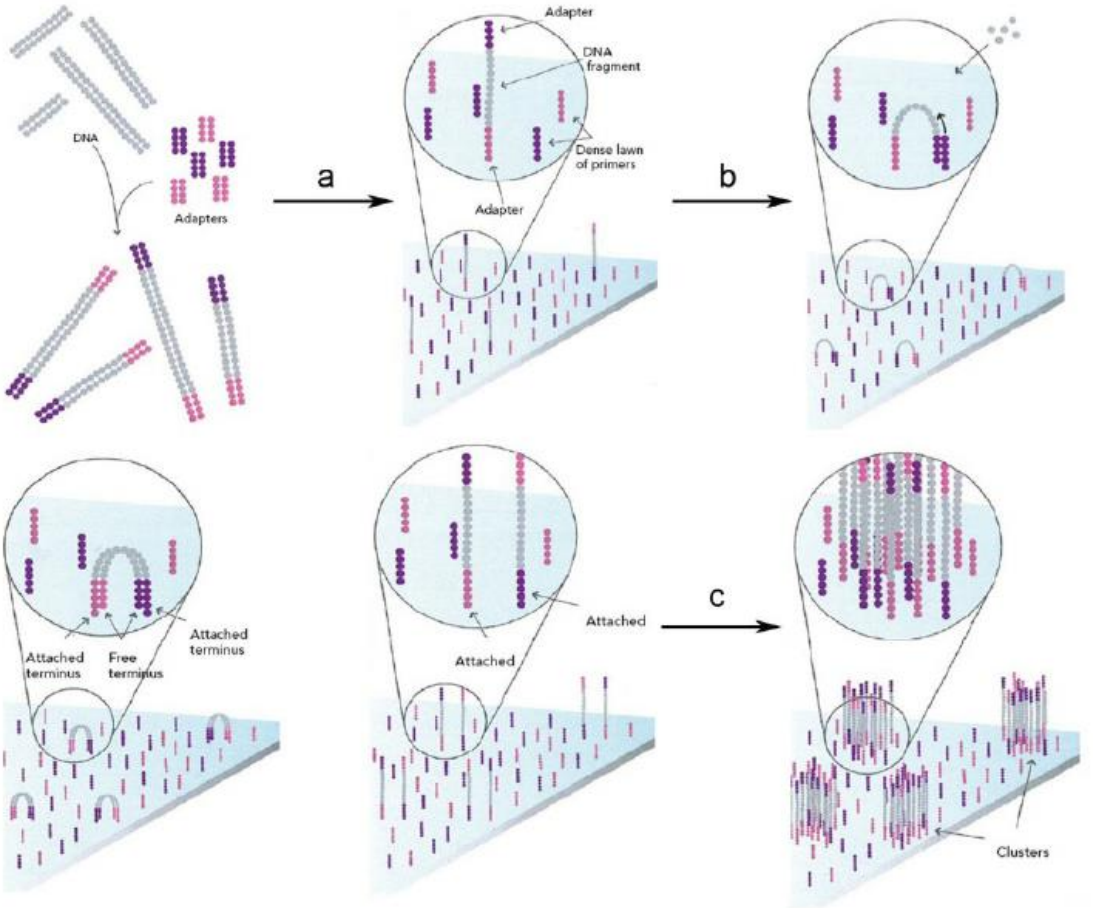


Figure 1.2: The figure illustrates the steps of bridge amplification and cluster generation during next generation sequencing by the Illumina platform. (researchgate.net)

Compared with traditional sequencing methods, NGS present itself with a higher throughput, overall lower sequencing costs and increased coverage per sample. (Zhang et al. 2011) The latter

merit allows for identification of genera that are otherwise low abundant in a community. (Claesson et al. 2010) Also, NGS has enabled more thorough analysis of structures and both taxonomic and metagenomics diversity of complex microbial communities such as the human gut. (Illumina.com) Despite these merits, there are some drawbacks associated with the NGS method, such as its immense requirement for computational power in order to drive the tracking and storage of data and its massive need for quality control. (Rizzo & Buck 2012) Also, the relative short read-lengths encumbers the performance of tasks with greater demands for longer sequence reads, such as de novo genome assembly. (Ferrarini et al. 2013)

1.7.3 Third generation sequencing

Although NGS still is considered as a relative new approach to sequencing, it will possibly be succeeded by approaches even more novel referred to as third generation sequencing. This includes methods such as nanopore sequencing and Pacbio-sequencing. Briefly, the first method involves the introduction of a voltage bias across a nanopore which consecutively give rise to detectable changes in the ionic current as molecules, such as a strand of nucleotides, are translocated through. (Branton et al. 2008) Pacbio sequencing on the other hand use DNA polymerases bound to 50nm wide structures on an array and fluorescently labelled nucleotides to synthesize DNA from a template. Owing to the immense amount of these structures on the same array, several templates are synthesized and sequenced simultaneously. Albeit the similarities in principles behind the SBS technology of Illumina and Pacbio platforms, there are some major differences in the resulting output. Sequencing by the Pacbio method produce significantly longer reads than by the Illumina method, with an average length of 2246 bp. However, the length of these reads seem to come at the expense of the accuracy of the readings. (Ferrarini et al. 2013)

1.8 Sequence analysis through QIIME

A popular bioinformatics pipeline for analysis of sequences is QIIME, which is an abbreviation for Quantitative Insights Into Microbial Ecology. A mapping file is normally required for data analysis, giving the program necessary information about the samples. Navas-Molinas et al. (2013) have proposed a rough division of QIIME workflow into an “upstream” and “downstream” analysis, each encompassing several steps managed by a series of commands.

1.8.1 Upstream analysis

Pre-processing of input data

The first step of analysis by this open-access tool, is a pre-processing step encompassing several events impacting downstream analysis. The first event involves the designation of sequences into their respective samples, based on the unique barcode attached at the end, also known as demultiplexing. Barcodes and primers are eventually removed. This is ensued by a quality filtration step, where sequences of low quality or with possible ambiguities are discarded according to a given set of parameters. This could include the minimum Q-score (q), percentage of consecutive base calls of high quality (p) and the maximum number of consecutive base calls of low quality (r) and ambiguous bases (n). (Navas-Molina et al. 2013) Often, a sub sampling of sequences of a given threshold (cut-off value) is implemented after the quality filtering, giving an even depth in all samples before downstream analysis. Thus, a number of sequences identical to this cut-off value are selected from each sample in a random manner. (Kuczynski et al. 2011) (Nelson et al. 2014)

OTU designation

An important step that potentially could pose a great impact on downstream analysis, is the designation of sequences into OTUs, which normally is performed with 97% sequence similarity. QIIME present three different approaches for this purpose: de novo, open reference based and closed reference based sequence clustering. The de novo based method encompass the designation of sequences into OTUs based on their resemblance to each other, without the use of known reference sequences. The reference based approaches on the other hand involves sequence clustering against references, thus giving a predefined set of possible OTUs. The main difference between these two reference based approaches is that the closed approach involves the exclusion of sequences that fail to be clustered against the reference. In open reference based approach however, these sequences are clustered de novo. Thus, each OTU comprise several related sequences.

In order to simplify downstream computer analysis, one representative sequence is given to each OTU which subsequently is given a taxonomic identity. The hierarchical level of taxonomic designation however, is dependent on the resolution of the representative sequence. This sequence could if needed, be submitted to an appropriate database such as BLAST (Basic Local Alignment Search Tool) for further taxonomic identification. (Kuczynski et al. 2011) (qiime.org) The OTUs are finally used to make an OTU-table and to create a phylogenetic tree in order to visualize the phylogenetic relationship between the identified OTUs. It has been argued that the creation of an

OTU-table should be ensued by a second quality filtration step to remove spurious OTUs of low abundance (Navas-Molina et al. 2013) which often are the results of chimera formation, PCR errors or sequencing errors. (Nelson et al. 2014)

1.8.2 Downstream analysis

Using the constructed OTU-table and the phylogenetic tree, QIIME provides the user with a number of different possibilities for downstream analysis, statistics and visualization. The relative abundance of different taxonomic levels, both within and between communities can be visualized through charts, and through a number of commands, several different metrics can be implemented for estimates of diversity estimates. (qiime.org) For simplicity, only a subset of metrics and visualization options will be presented.

Intragroup diversity analysis

Alpha diversity encompass the diversity within samples and is often presented as OTU-richness, although several other indices for alpha diversity has been developed, such as the Chao1, Shannon and Simpson indices. While the Simpson indices tries to estimate the relative abundance of the species in a sample, the Shannon metric also tries to identify the number of unique species. Chao1 on the other hand aspire to estimate the number of species present in a sample, if sampled exhaustedly. Regardless of method for alpha diversity estimates, QIIME allows for presentation through a rarefraction plot, thus making it possible to assess whether the cut-off value gave satisfactory coverage of the species present. This is usually determined by evaluating the extent of which the slopes present an asymptotic shape. (Pepper et al. 2015) (qiime.org)

Intergroup diversity analysis

Beta diversity metrics typically aspire to present degree of similarity in species composition and/or distribution between samples. Several indices for beta diversity exist with the Jaccard, Bray Curtis and Unifrac possibly comprising the most common approaches. While Jaccard only consider the presence and absence of species, their relative abundance is taken into consideration in Bray-Curtis. (Pepper et al. 2015) Unifrac however, aims at determining the difference between microbial communities by establishing their phylogenetic distance in terms of branch length. (Lozupone & Knight 2005) Thus, the extent of tree similarity between communities determines the beta diversity. (Pepper et al. 2015) Unifrac measurements can be unweighted or weighted, where the latter approach accommodate for potential differences in the relative abundance of taxa in the compared

communities, thus giving a qualitative measurement of beta diversity. (Lozupone et al. 2007) The unweighted approach on the other hand only interpret the absence/presence of OTUs. (Navas-Molina, 2013) QIIME permit visualization of the beta diversity through PCoA-plot (Principal Coordinates Analysis-plot) and hierarchical clustering. (Kuczynski et al. 2011) (qiime.org)

1.9 Aim of project

The findings from Aamodt et al (2008) and Frank et al (2007) as mentioned in section 1.3.2 initiated the establishment of this project. Albeit presenting interesting results with respect to a potential linkage between drinking water and IBD, an elucidation of an association with respect to the microbiota still remains untouched. Therefore, the main aim of this research is determining if the microbiota of biopsies retrieved from selected patients can be explained by the microbiota of the tap water taken from the same subjects under investigation. Thus, we aspire to contribute to the investigation of if and how tap water can serve as an environmental trigger in the development of IBD.

As this seem to be a relatively new area of investigation, our null hypothesis is that there is no association between the microbial communities of tap water and biopsies, and no involvement of tap water in the aetiology of IBD. The alternative hypothesis is that tap water may serve as an etiologic agent in the development of IBD and that there is an association between the microbiota of tap water and biopsies. If latter hypothesis is to apply, we suggest that this association can be explained by either direct or indirect means. An association by direct means refer to a possible direct transmission or colonization of tap water bacteria to the mucosa of subjects under investigation. Indirect means on the other hand may involve the production of substances or metabolites by tap water bacteria which potentially might alter the biochemical conditions of the ingested water with a subsequent influence on the microbial growth in the mucosa. Albeit similar hypotheses earlier have been proposed by Aamodt et al (2008) and Frank et al (2007), projects with aims comparable to the aim of this research has to my knowledge not been performed.

426 biopsies and 227 water samples from a selected Norwegian cohort consisting of IBD patients and healthy controls were used as study material and analysed by using culture independent techniques. The V3-V4 region of the 16S rRNA gene was amplified in a nested approach and followed by sequence determination using Next Generation Sequencing and the Illumina MiSeq-sequencer. QIIME-pipeline was employed for data analysis.

2.0 Materials and methods

2.1 Study material

Materials were collected between 2005-2007, originally as a part of the Inflammatory Bowel Disease in south-eastern Norway II (IBSEN II) project, aiming at investigating genetic, immunological and environmental factors implicated to participate in the aetiology of IBD. In this project, prospective patients from geographically restricted areas in south-eastern Norway, presenting several traits characteristic of IBD were invited to participate. Using colonoscopic examination, a final diagnosis was established based on a given set of criteria for IBD (Lennardjones 1989) with subsequent classification of the disease based on the Montreal classification. (Satsangi et al. 2006) Patients who met the given requirements were divided into UC, CD and Inflammatory bowel disease unclassified (IBDU). Patients with less pronounced findings were labelled as possible, while patients without any pathological findings were classified as non-IBD and included as controls.

227 water samples and 426 biopsies from 224 different patients were used for this research project, and includes samples retrieved from both adults and children <18 years. Patients returning for a follow up study and re-evaluation of diagnosis, 1-1,5 years after the initial diagnosis was determined are included in these numbers. Biopsies were retrieved from both non-inflamed (A) and inflamed (B) tissue and from tissue of unknown category. A total of 7 different locations within the gut served as origin. (Ileum=I, Caecum=II, Ascendes=III, Transversum= IV, Descendens=V, Sigmoideum=VI, Colon=VII) The distribution of patients and material for analysis with respect to essential parameters such as age group, diagnosis, sample location and tissue type is illustrated in figure 2.1. A complete list of all samples used for this research project, identification number, and origin with respect to patient, diagnosis, tissue type and location is given in appendix A.

Biopsies were stored in a freezer at Rikshospitalet for subsequent transportation on dry ice to NMBU for further storage at -60^{o}C. Prior to this project, DNA of water samples had been extracted, purified, quantified and stored in freezer.

An overview of the workflow implemented is shown in figure 2.2.

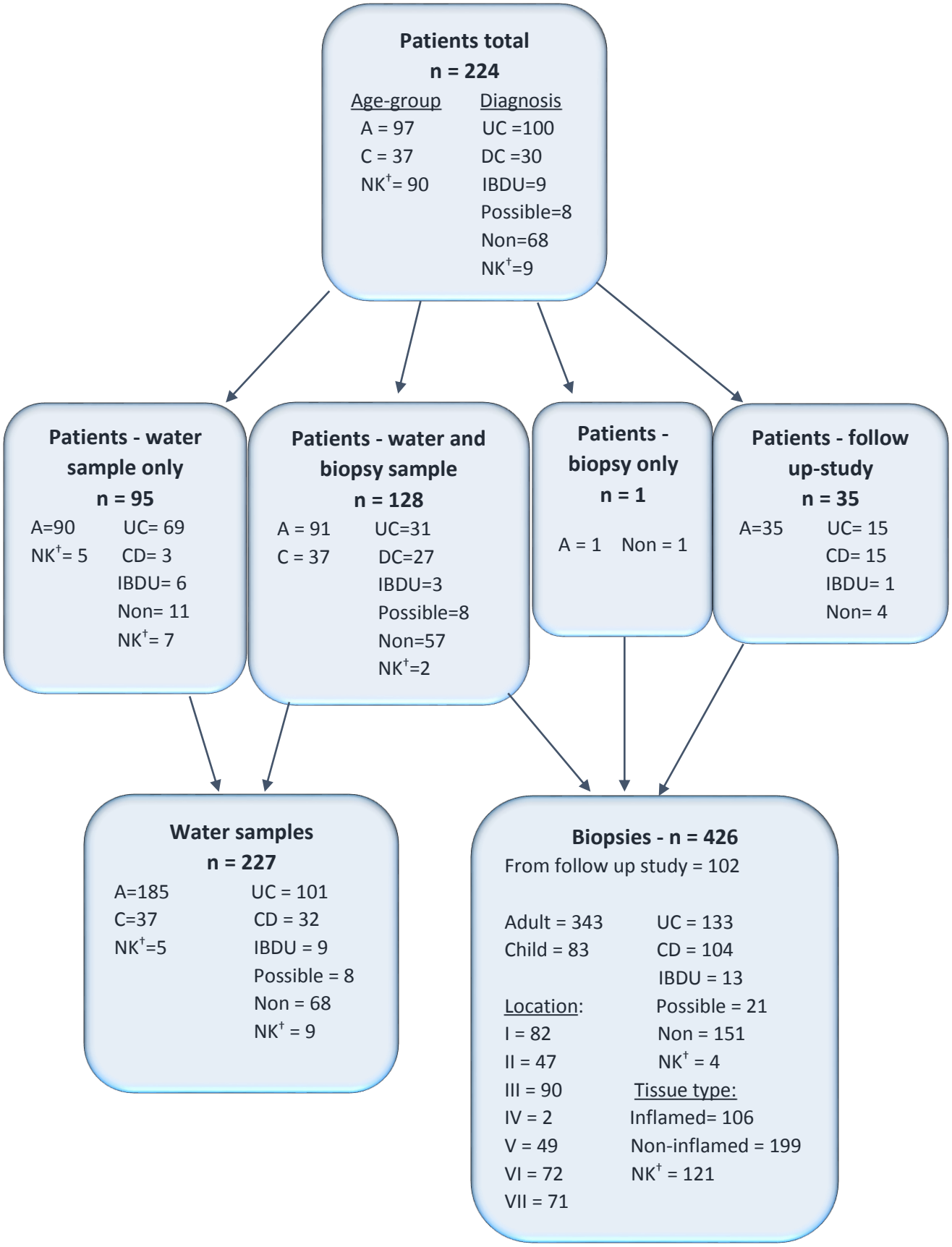


Figure 2.1: Shows the sample distribution with respect to patient category, age group and diagnosis for biopsies and water samples, and localization and tissue type for biopsies.
† NK = Not known

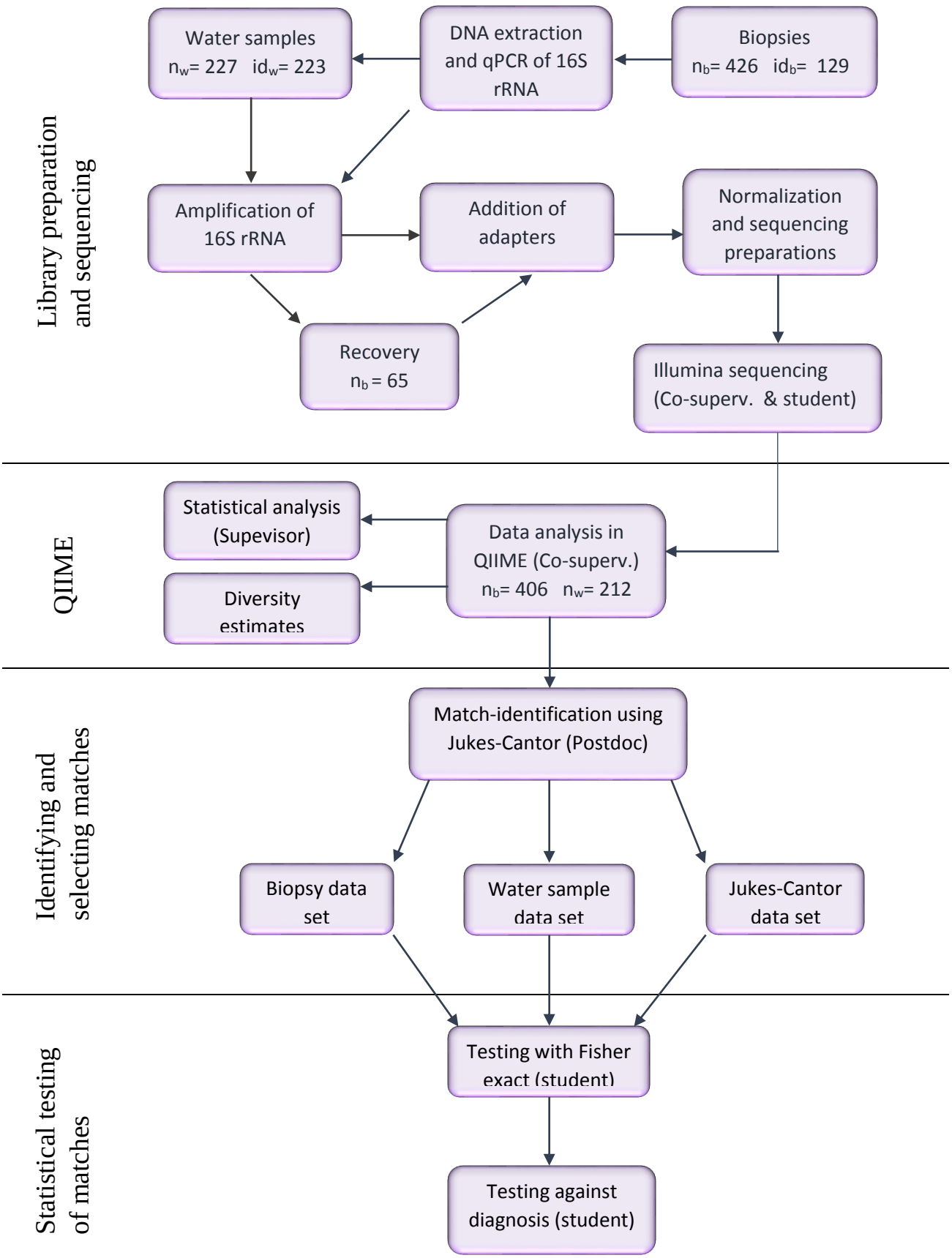


Figure 2.2: illustrates the workflow implemented during the research process with respect to both water samples and biopsies.

2.2 Obtainment of DNA and quality assurance

2.2.1 Cell lysis and extraction of DNA

Biopsies ranging from <1mm^3 to 6mm^3 in size were transferred to tubes containing approximately 0,25g acid washed glass beads (Sigma Aldrich, Steinheim, Germany) and 200µl S.T.A.R buffer (Stool Transport and Recovery, Roche, Basel, Switzerland) for preservation of DNA. Lysis of cells was performed twice with MagNA lyser instrument (Roche); 6500 rpm for 20s, with 1 min cooling between runs to avoid overheating and DNA degradation. During this pause, tubes were flicked to prevent biopsies to adhere to lid during lysis. Proximate to DNA extraction, samples were centrifuged at 13000 rpm for 5 min for separation of glass beads and cell matter from the DNA in supernatant.

Mag^{TM} mini kit (LGC, Middlesex, UK) was used for the extraction of DNA in gut biopsies. Lysis buffer and proteinase were added to supernatant, followed by a 55^o incubation for 10 minutes in order to degrade protein remnants. Ethanol and paramagnetic beads were added to the suspension containing DNA. Owing to the DNA binding capabilities of the latter, three successive washing steps with washing buffer BLM 1 and BLM 2 were permitted to remove impurities. The final step of extraction involved the release of the newly washed DNA from the beads with Elution buffer. The procedure was made automatized by the use of KingFisher^{TM} Flex Magnetic Particle Processor, (Thermo Scientific^{TM}, Waltham, USA) using the programs “ProteinaseLGC” and “MagMiniLGC” Negative control was included to secure that DNA contamination was avoided.

2.2.2 Quantification of prokaryotic DNA

Due to expectations of low bacterial quantity in biopsies, extracted DNA from all samples were subject to quantifications of prokaryotic DNA by quantitative PCR, hereby referred to as qPCR. The amount of amplicons of the16S rRNA gene as indicated by the resulting Ct-value, served as the main determinant for this purpose. Quantification was accomplished with LightCycler 480 II (Roche) by using 0,2µM of the16S forward and reverse rRNA primers PRK341F and PRK806R (Invitrogen, Thermo Scientific^{TM}). 5x HOT FIREPol$^{®}$ EvaGreen qPCRMix Plus (Solis BioDyne, Estonia) was used for fluorescence due to its non-specific binding properties, and diluted to a final concentration of 1x. 5µl DNA template was embedded in the total reaction volume of 20 µl.

Positive and negative controls (*E.coli* genomic DNA and mastermix respectively) were included. The following program was implemented: activation 95^{o}C for 15min, 40 cycles of 95^{o}C for 30s, 55^{o}C for 30 s and 72^{o}C for 45s. Owing to a relatively high amount of co-amplification of eukaryotic

DNA in a preceding test-run, annealing temperature was increased from 50°C in the original program to 55°C in order to increase primer specificity. Melting curve analysis was included to account for possible formation of nonspecific products. Assuming that each bacteria harbours a copy number of the 16S rRNA gene of 3, and that the amplification efficiency and detection threshold of the qPCR reaction equals1,6 and 10^{10} respectively, the following formula was implemented for theoretic estimates of bacterial counts in the qPCR reaction:

$$\frac{\left(\frac{10^{10}}{1{,}6^{\text{Ct}-\text{value}}}\right)}{3}$$

2.2.3 Quality assurance

In order to secure satisfactory results from DNA extraction, amplification and purification processes, samples were subject to quality assurance by quantitative or qualitative means. Both methods of DNA measurements were performed by the use of a 1:200 dilution of Qubit® dsDNA HS Reagent (Invitrogen) possessing DNA binding and fluorescent properties. For quantitative DNA measurements, the fluorescence and thus, the concentration of DNA was estimated by use of a Qubit™ fluorometer. (Invitrogen) Detection of fluorescence from qualitative DNA measurements on the other hand, was employed by the use of Cambrex FLx800cse machine (Cambrex, East Rutherford, USA)

Quality assurance also involved the application of gel electrophoresis, where samples including controls were validated on 1% agarose gel. PeqGreen RNA/DNA Dye (Peqlab, Erlangen, Germany) and100bp DNA ladder, (Solis BioDyne) were used for staining of DNA and comparison of band sizes respectively. An electric current of 80V for 30 minutes was applied and succeeded by visualization in Molecular Imager® Gel Doc™ XR Imaging System (Bio-Rad, Hercules, USA).

2.3 Amplicon library preparation

Amplification of V3 and V4 segments of the16S rRNA gene and the adjoining of adapters were performed in two separate steps/nested reactions in order to maximize the specificity of the primers. The latter reaction was not initiated until all samples reached the completion of the first PCR reaction. Positive and negative controls (*E.coli* genomic DNA and master mix, respectively) were included in all PCR reactions. The water samples were divided and processed in three batches, with duplicates of sample 4-41 in the last batch. The mucosal samples were divided into five batches,

with the last batch containing 6-7 duplicates from each of the four first batches. Comprising as much variation as possible in sample characteristics was considered to be of main importance when choosing the latter duplicates.

In addition to visualizing a subset of amplicons from all PCR reactions on gel, quantitatively and qualitatively measurements were applied, as explained under quality assurance. A subset of water samples from the first PCR reaction were quantitatively measured, while a qualitative measurement was applied on all mucosal samples owing to large variations in the Ct-values from the qPCR. A qualitative measurement was applied on both water samples and mucosal samples after the final PCR reaction. Fluorescence from negative sample was subtracted from the resulting fluorescence to account for excess nucleotides, primer dimers etc. Samples with a fluorescence equal to a non-detectable band were submitted to additional rounds of recovery with the rationale of generating a higher amount of amplicons. The subsequent adjustment of the conditions behind the two nested reactions primarily targeted the number of cycles and the amount of template DNA.

2.3.1 Nested PCR

Template DNA from both reactions was embedded in a 25µl reaction volume of 1x of HotFirePol® buffer B_2, 25mM of $MgCl_2$, (all Solis BioDyne) 200µM dNTP (Solis BioDyne). 1,25U concentration of HotFirePol® and FirePol® DNA polymerase were used in the first and second PCR reaction respectively. 5µl template DNA was used for the amplification of 16S rRNA, with DNA concentrations in the range of <0,5-1,1 ng/mL and 0,3-25 ng/mL from the water samples and gut biopsies respectively. For the adaptor adjoining 5-10µl template with DNA concentrations of 1-9 ng/mL and 0,1-3,1ng/mL of the respective sample types were used.

0,2 µM of PRK341F (5`-CCTACGGGRBGCASCAG-3`) and PRK806R (5`-GGACTACYVGGGTATCTAAT-3`) (Invitrogen) was included in the amplification of 16S rRNA, thus allowing for the amplification of V3, V4 and the conservative regions interspersing these variable regions. This reaction will be referred to as PRK PCR. For adjoining of adapters, 0,2 µM of the 16S rRNA forward and reverse indexing primers (Invitrogen) was employed. Indexing primers were added manually on the purified water sample PCR products, and made automatized on the mucosal samples by use of Eppendorf epMotion 5070 machine. (Eppendorf, Hamburg, Germany) 36 forward and 16 reverse primers, each with a unique barcode, were used in each set of samples. Primers were arranged in a manner giving each sample from each set a unique barcode

combination, allowing for the annotation of sequences to their respective sample after the final sequencing. This reaction will be referred to as indexing PCR.

The following PCR program was implemented for the amplification of 16S rRNA; activation 95ºC for 15 min, 25-30 cycles of 95ºC for 30s, 50-55ºC for 30s, 72ºC for 45s, and final elongation at 72ºC for 7 min. The adjoining of adapters on the other hand required the following program; 95ºC for 5 min, 10 cycles of 95ºC for 30s 55ºC for 1 min, 72ºC for 45s, and final elongation at 72ºC for 7min. 2720 Thermal Cycler (Applied Biosystems, Foster city, USA) served as the amplification instrument in both PCR reactions.

2.3.2 PCR product purification

The nested PCR reactions were interspersed by a purification of PCR products with Sera-Mag Magnetic Speed beads in order to remove unincorporated nucleotides, primer dimers, smaller fragments, etc that could pose an impact on the final sequencing process. A 1:1 ratio of PCR product and bead solution were mixed, allowing for DNA fragments over a certain size to bind to the magnetic beads. While on magnet, DNA was washed three times with fresh 80% ethanol, ensued by a release from the beads with nuclease-free water. The purification process was made automatized in Biomek® 3000 Workstation (Beckman Coulter Life Sciences, Indianapolis, USA) in concordance with manufacturers protocols. Bead solution was made from 0,1% carboxyl-modified Sera-Mag Magnetic Speed beads (Fisher Scientific, Thermo Scientific™) pre-washed with TE, 18% PEG, 1M NaCl, 10mM Tris-HCl pH 8, and 1mM EDTA pH8. After each purification, DNA was quantitatively measured with Qubit in a subset of samples as described under Quality assurance, to confirm the binding of DNA to the beads. A second post PCR purification process after the adapter adjoining was not considered necessary due to the low amounts of interfering products shown after gel electrophoresis on a subset of samples.

2.3.3 Sequencing preparations.

Ensuing adapter adjoining on all samples, a normalization process was performed in order to achieve that an equal amount of indexing PCR products was transferred to the Illumina sequencing chip, and to avoid an over or under representation of certain samples. This was performed manually on both sample types. The qualitative DNA measurements of the PCR indexing products, were used as a baseline for normalization. The fluorescence from the negative control was subtracted from the value to account for possible primer dimers, excess nucleotides etc. Owing to the even fluorescence

in water samples, an equal volume from each sample was transferred to a common pool. The higher dispersion of fluorescence in the biopsies made it more feasible to divide samples in groups of ten. An appropriate normalization volume was subsequently assigned to each group, ensuing a fairly equal amount of DNA to be transferred from all samples. Both pools were subsequently subject to a manual purification in order to remove excess nucleotides, primer dimers and non-specific smaller amplicons etc. Pools were mixed with Sera-Mag magnetic speed beads in a 1:0,8 ratio allowing for a removal of DNA under a certain fragment length. While on magnet, the attached DNA was washed twice with 80% fresh ethanol, and eluted with nuclease-free water. Pools from before and after this purification step was checked on 1% agarose gel as described under Quality assurance, in order to confirm the success of the clean-up.

Both cleansed pools, hereby referred to as libraries, were submitted to quantification with Perfecta® NGS Library Quantification Kit for Illumina® Sequencing Platforms, (Quanta BioSciences, Gaithersburg, USA) following manufacturers instructions and by the use of LightCycler 480 II. A 1:2000 and 1:20000 dilution of the libraries were included, together with five standards ranging from 0,0005pM to 5pM. All samples were run in triplicate reactions to increase the reliability and account for possible deviations. Negative control was also included. Based on the Ct-values from the included standards, an equation from the resulting calibration curve was made, and the corrected concentration in the amplicon libraries was estimated. Libraries were subsequently diluted to 4nM in order to generate a proper cluster density.

2.3.4 Library denaturation and Miseq sequencing

Amplicon libraries consisting of water samples and mucosal samples were sequenced separately. Before loading onto chip, diluted libraries were prepared and denatured according to Illumina Library Preparation guide and by use of Miseq reagent cartridge (Illumina, San Diego, USA). A denatured control of PhiX was included to serve as a contrast during the reading process and permit error rate calculations. 4nM of PhiX and amplicon library were prepared in a similar manner by first separately combining the samples with equal amounts of 0,5N NaOH, giving samples of 2nM. Samples were vortexed, centrifuged at 280g at 20^{o}C for 1 min (libraries only), followed by a 5 minute incubation at room temperature in order to separate the strands. Libraries and PhiX were further diluted to 6pM with HT1. Amplicon library of mucosal samples was spiked with15% PhiX according to manufacturers recommendations, while a 30% spike-level was used for the water samples due to recent technical problems with the Miseq machine. Spiked libraries were separately

applied onto the flow cell of an Illumina chip, and sequenced by use of a MiSeq sequencing platform (Illumina) All library denaturation and MiSeq sequencing steps were performed under the supervision of co-supervisor.

2.4 Analysis of sequencing data

2.4.1 Analysis in QIIME

Raw sequence data from the water samples and the biopsies were uploaded and processed separately in QIIME by co-supervisor. Sequences were initially demultiplexed and filtered to secure that only sequences of satisfactory quality were used for downstream analysis. For this, minimum sequence length and E-value was set to be 350 nucleotides and 0,2 respectively. For sequences from both biopsies and water samples, a cut-off value of 3000 sequences from each sample was set and served as the basis for the subsequent designation of sequences into OTUs. This was performed using usearch, the UPARSE algorithm, and a closed OTU-picking strategy. Ultimately, Greengenes database served as the reference system. Sequences were screened for potential chimeras using ChimeraSlayer. OTUs were then subject to several diversity estimates using the command core_diversity_analysis in QIIME. Phylogenetic diversity whole tree, Observed species, Shannon, Simpson and Chao1 served as indices for estimates of alpha diversity, while weighted and non-weighted UniFrac, Jaccard and Bray Curtis indices were implemented for estimates of beta diversity. This was visualized through rarefraction and PCoA-plots respectively. Graphics of charts showing relative abundance of taxonomic groups in the microbial communities were included

2.4.2 Statistical analysis of datasets

To test for differences of a given OTU within the groups of the biopsy and water sample data set, the command OTU_category_significance was incorporated to the QIIME workflow by co-supervisor, using statistical principles of Kruskal-Wallis test. Correction of the resulting p-value with the Bonferroni approach was included, to further reduce chances of getting false positives.

A Principal Component Analysis (PCA) including a score plot and loading plot was performed on the biopsy OTUs. Thus, the dataset was reduced to a smaller and more manageable pattern of data referred to as principal components, alleviating further downstream statistical analysis. To test for potential interactions between the independent variables and their impact on the dependent variables, ASCA Analysis of Variance (ANOVA) was employed as a statistical method using PLS Toolbox. (Eigenvector Inc, Washington, USA) A significance level of 5% were used for all

statistical tests. To further identify if potential intragroup differences were present in the OTUs implicated to be of significance for potential interactions between age and diagnosis, a Kruskal-Wallis test was performed using SYSTAT13 (Systat Software Inc, California, USA). As this method do not detect where potential intragroup differences occur, Conover-Inman test for pairwise comparisons was implemented as a statistical method as well on the median percentagewise prevalence of the OTU, also using of SYSTAT13. The latter analysis does however not announce the direction of significance in each pair. For this, the median values of the tested OTU were in each group of pair were compared. To test for potential significances of the prevalence of OTUs in different combinations of inflammation category and age, Kruskal-Wallis followed by Conover-Inman was implemented. This was performed on all of the enrolled patients. Tissue of unknown category was excluded from this analysis to prevent the introduction of possible biases. All of these analyses were performed by supervisor.

In situations were further identification of OTUs on a lower hierarchical level was of interest, the reference sequence generated by QIIME during the designation of sequences into the respective OTU was uploaded to BLAST by student. The16S ribosomal RNA sequence database was used for identification. Only suggested taxonomic annotations with the most suitable query cover, identity and E-value were presented and discussed.

2.4.3 Analysis of associations between OTUs in water and biopsies

Identifying and selecting matches

In order to unveil any potential transmissions of OTUs from water to mucosa, the reference sequence from each OTUs in the biopsy and water sample data set were first mapped against each other by a postdoc from the department using MATLAB®. (MathWorks, Natick) A threshold of ≥97% sequence similarity was employed for the identification of potential OTU matches using the following Jukes-Cantor model for sequence divergence estimates: $d = -{}^{3}/_{4}\, ln\left(1 - {}^{4}/_{3}\, p\right)$ where *d* represents the evolutionary distance between two sequences, and *p* is the proportion of substitutions across the sequence alignment, i.e. the sequence distance. (Xiong 2006) This was initially performed without taking the prevalence of the OTUs into consideration. Each match was then given a taxonomic identity.

Owing to the complexity of identifying potential associations on all matches, only selected matches from the water sample data set, the biopsy data set, and from the Jukes-Cantor data set were

submitted to further analysis by student. Only matches showing significant Bonferroni-corrected p-values with respect to diagnosis during the implemented statistical testing with Kruskal-Wallis in QIIME, were chosen for this purpose in the first to data sets. As the aim was searching for potential transmission of OTUs from water to mucosa, matches in the Jukes-Cantor dataset on the other hand, were selected based on the prevalence of the water OTUs.

To reduce chances of analysing water OTUs present by mere coincidence, matches were narrowed down to include those connected to the 50 most prevalent OTUs from the water sample data set. An overview of taxonomic belonging was made to evaluate if a further narrowing of the matches was needed prior to subsequent analysis. To account for the possibility that spurious OTUs still might comprise a part of the remaining matches between the datasets, water OTUs were plotted against its percentwise prevalence and a threshold was established where a change of decline could be observed. Thus, OTUs from water sample data set showing an average prevalence above this threshold were submitted to further analysis of potential transmission using Fisher exact.

Statistical testing with Fisher exact

Selected matches from the biopsy, water sample and Jukes-Cantor dataset were subject to statistical analysis of any plausible associations between OTU matches in water and biopsies. This was performed with the Fisher exact method by student, with the rationale that plausible associations potentially could be used for further evaluation of OTU transmission from water to mucosa. A match was considered to be present in both samples if $\geq$1 sequence(s) from each of the OTUs in the match could be detected in both water sample and in biopsy. In cases were a patient presented two water samples or more than one biopsy, of which only one of the respective samples contained the OTU of interest, the OTU was considered present. Level of significance was set to be 0,05. Characteristics of the samples such as age-group, diagnosis etc. was not taken into consideration, as the primary aim was searching for potential associations regardless of origin.

Matches presenting a Fisher exact value below the level of significance were submitted to an additional round of Fisher exact testing by student to see if possible associations could be attributed to certain diagnosis groups. Dataset of the OTU match of significance was decomposed into Non, IBD, CD and UC groups, where the IBD group encompassed patients from the latter two groups and IBDU. Patients having a status of diagnosis marked as possible or unknown were excluded from this final analysis to prevent the introduction of possible biases.

3.0 Results

3.1 Library preparation

The quantifications of DNA ranged from <0,5-1,1ng/mL for the water samples and 0,3-26ng/mL for the biopsies. The output from the qPCR measurements on the biopsies, gave a wide range of Ct-values from 33,78 to 16,97 with a median of 23,33. Thus, the theoretic amount of bacterias per µl eluted DNA ranged from from 85 to 229 038, with a median of 11 529, corresponding well to the variation of biopsy sizes. Ct-values of positive and negative controls of 13 and 35 respectively, confirmed a successful amplification of prokaryotic DNA with little sign of DNA contamination.

Samples were submitted to two nested PCR reactions, with subsequent qualitative measurement of fluorescence. Examples of amplicons in a subset of biopsies from the first and second PCR reaction are given in figure 3.1 and 3.2 respectively. Net fluorescence is included for comparison reasons.

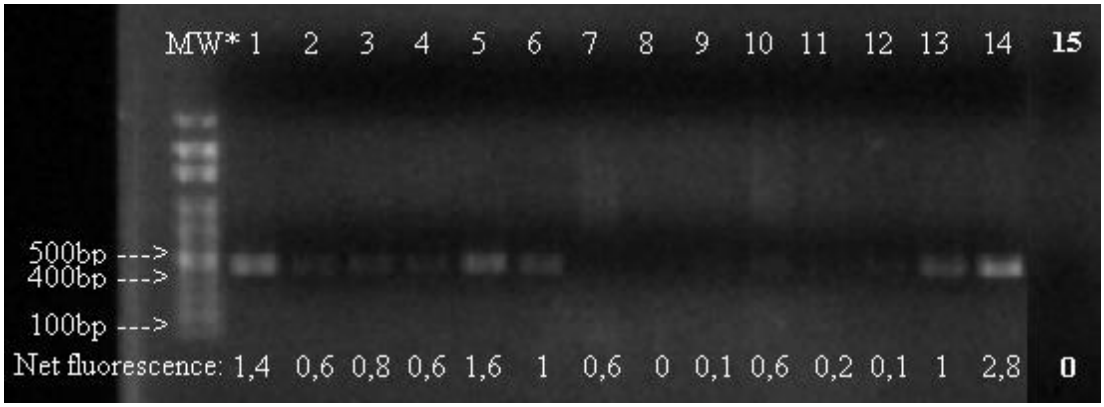


Figure 3.1: shows the resulting amplicon products in a subset of biopsies (lane 1-13) from the PRK PCR reaction. Positive and negative controls are shown in lane 14 and 15 respectively. Net fluorescence (thousands) is shown in the bottom of each lane.
*MW = Molecular weight, 100bp ladder.

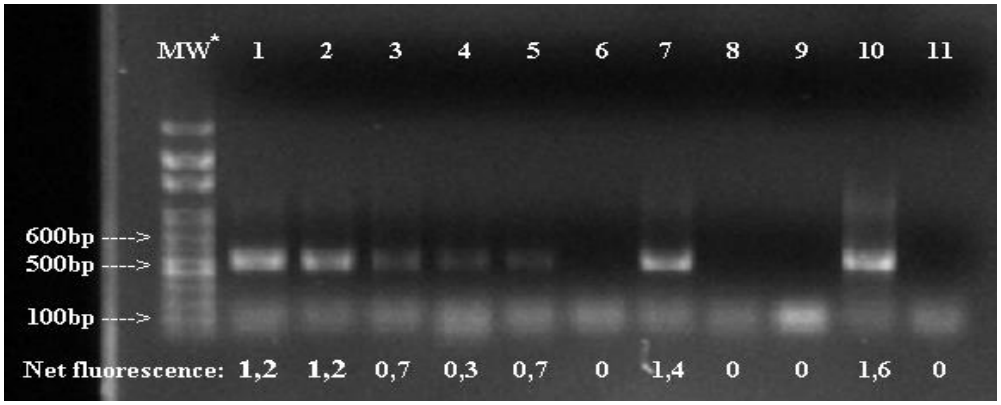


Figure 3.2: Shows the resulting amplicon products in a subset of biopsies (lane 1-9) from the indexing PCR reaction. Positive and negative controls are shown in lane 10 and 11 respectively. Net fluorescence (thousands) is included in the bottom of each lane.
*MW = Molecular weight, 100bp ladder.

Figure 3.1 and 3.2 shows the absence of a band in negative controls and a clear, visible band in both positive controls of the expected sizes of 450bp and 580bp in the PRK PCR and indexing PCR reaction respectively. The presence of expected product in the other samples is confirmed from bands of approximately equal length. With respect to the amount of amplicons, there seem to be large variations, corresponding to the large variations of bacterial counts. Both PCR reactions show sign of smear, and formation of primer dimers of approximately 100bp at the end of each line, although this observation seem to be more prominent in the indexing PCR.

When comparing fluorescence to band intensity, a probable association between net fluorescence and band strength become evident, where a fluorescence of 0,5 seem to be required for a band to become noticeable. Thus, 65 biopsy samples recalcitrant to give a fluorescence above 0,5 were submitted to 4 additional rounds of increased cycles and/or increased template DNA, as described in section 2.3 in material and methods. The baseline for normalization both before and after additional rounds, and sample distribution based on net fluorescence is illustrated in figure 3.3.

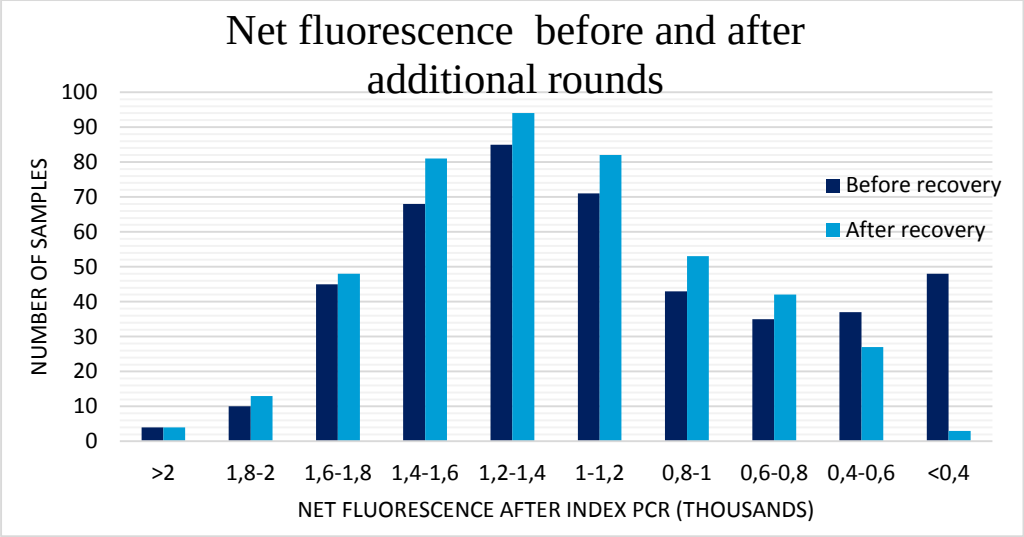


Figure 3.3: shows the total sample distribution based on net fluorescence after indexing PCR, both before and after additional rounds, and thus the baseline for normalization.

As presented in figure 3.3, a substantial amount of samples initially below the threshold of 0,5 returned a net fluorescence above 0,5 after the additional rounds of increased cycles and/or increased template DNA.

3.1.1 Illumina sequencing

The qPCR quantifications of the pool containing water samples resulted in concentrations of 21,08nM. Sequencing by Illumina resulted in cluster density of 577 K/mm^2 and a subsequent clustering passing filter of 95,72. Q30 was estimated to be 80,91. With respect to the biopsies, quantified pool returned a concentration of 21,3nM, while the Illumina-run resulted in a cluster density and passing filter of 920 K/mm^2 and 96,28. Q30-score was set to be 77,69.

3.2 Sequence processing

Three thousands sequences per sample was set to be the threshold for further downstream analysis for both datasets. Owing to the initial quality filtration step, 20 of 426 (4,7%) unique biopsy samples were excluded from the dataset. Thus, sequences from 406 biopsies served as material for further computer analysis. The number of sequences per sample in the biopsies ranged from 12 to 202 627, with a median of 13 238. With respect to the water samples, 15 of 227 (6,6%) of the unique water samples were removed, leaving 212 samples for analysis. Sequence number varied from 28 to 62 589 with a median of 18 145.

3.2.1 Intragroup diversity analysis

Alpha diversity estimates resulted in several rarefraction curves, of which one from each dataset is exemplified below in figure 3.4 and figure 3.5. This was amongst other things used to determine if the cut-off value of 3000 was set to a reasonable level. For simplicity, groups of unknown category or considered to be of minor significance are excluded from these plots. A more detailed description of the results of this analysis, is given in appendix B with all metrics employed for analysis and respective rates of error.

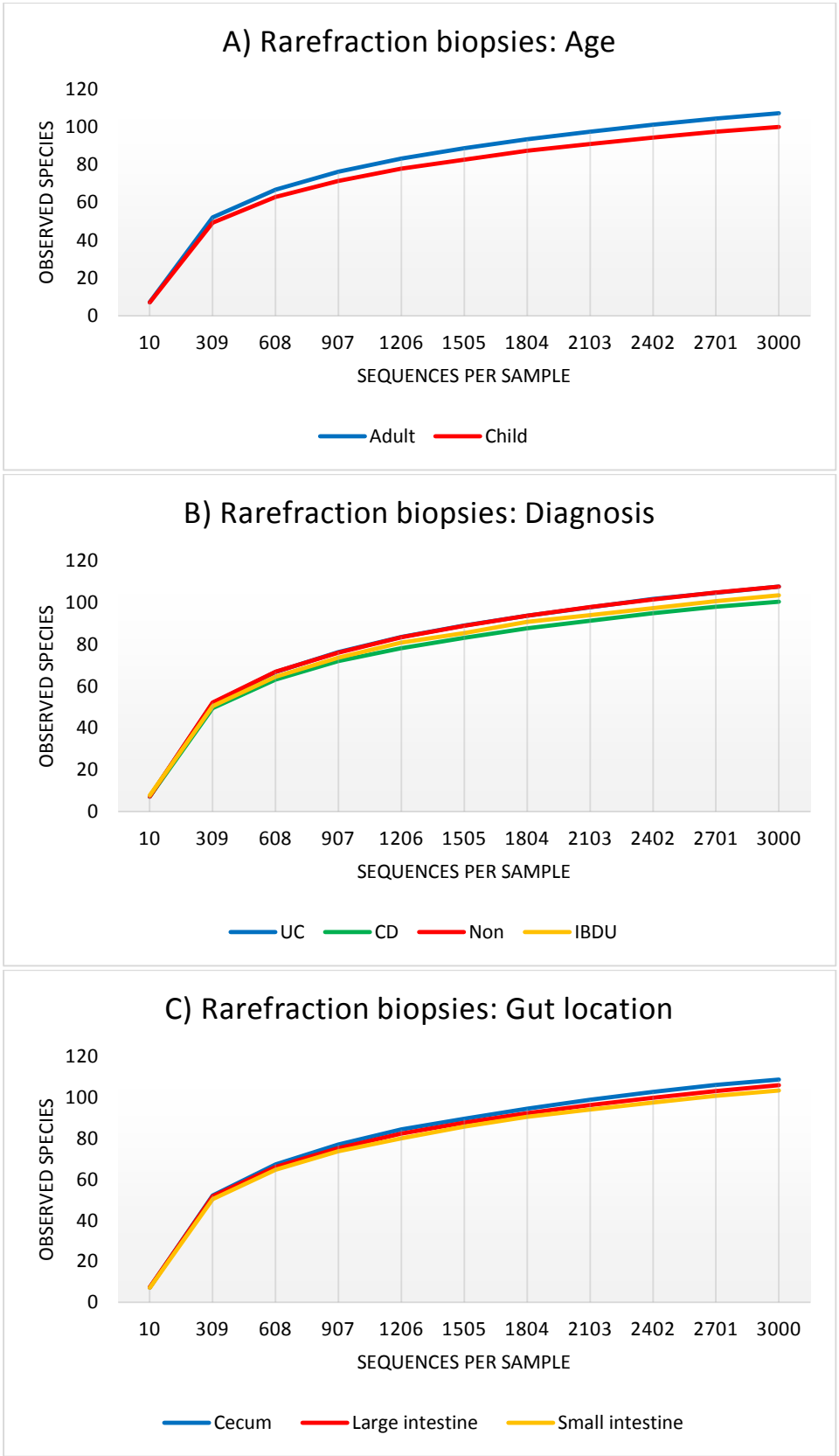


Figure 3.4: Rarefraction curve A, B and C illustrates the alpha diversity at 97% sequence similarity, using number of species as metrics. Curve A,B and C give the alpha diversity in biopsies with respect to age, diagnosis, and gut location respectively.

As displayed in figure 3.4, the rarefraction plot of the biopsies show a continuous flattening of the curves towards an asymptotic shape, as the amount of sequences increases. When performing alpha diversity estimates on the biopsies, all diversity metrics displayed a somewhat higher diversity in adults compared to children, as exemplified in curve A. With respect to diagnosis, all metrics disclosed CD and IBDU to have the lowest diversity estimates. UC and control group presented the highest diversity, and appeared to be equally diverse (Curve B) Concerning gut location, the metrics exhibited the lowest diversity in the small intestine. Large intestine and cecum on the other hand appeared to be equally diverse due to the lack of a consistent pattern between the metric. The latter location presented the highest estimates when observed species was employed as metrics, as shown in curve C. Furthermore, a consistent pattern of diversity in inflamed and non-inflamed tissue could also be observed with the latter demonstrating a higher diversity in all metrics. Using observed species as metric, at 3000 sequences non-inflamed tissue presented a higher alpha diversity compared to inflamed tissue, showing 110 and 98,5 species and errors of 34,4 and 33,1 respectively.

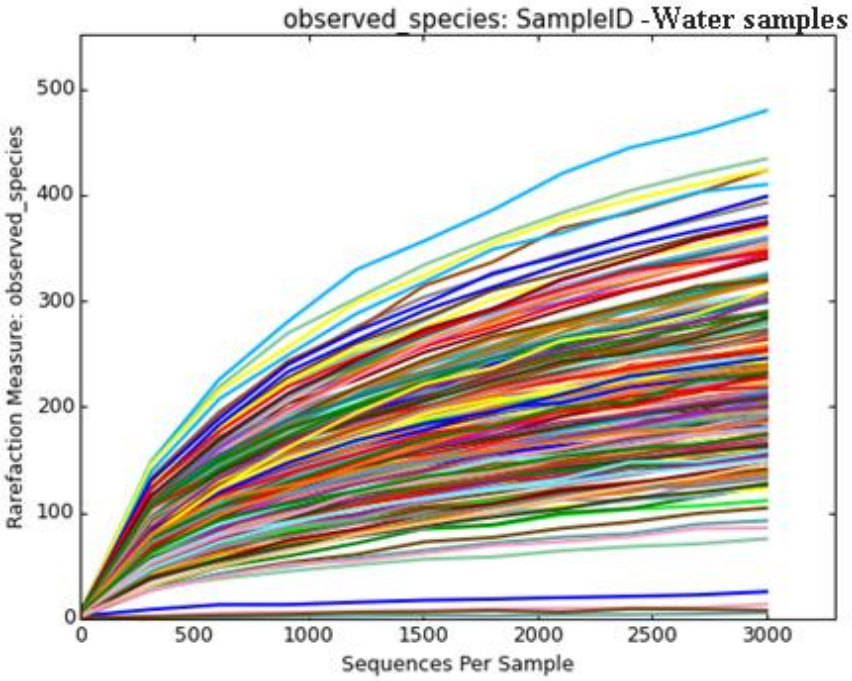


Figure 3.5: illustrates the alpha diversity in all water samples at different sequence amounts when number of species is employed as metric.

The rarefraction curve of the water samples also display a continuous flattening of the curve as the number of sequences increase, although somewhat less evident in the samples with the highest species number. When comparing alpha diversity estimates on water samples in figure 3.5 to those of biopsies, it becomes apparent that the number of species is considerably higher in most of the

water samples. All metrics displayed great variation between the water samples and as illustrated in figure 3.5, the number of species ranged from approximately 5 to 495 when the number of species was used as metric. Diversity estimates on water samples from combined groups of age and diagnosis showed a somewhat lessened congruency between the different diversity metrics. However, a pattern of water samples from pediatric CD patients displaying the lowest alpha diversity compared to Non_C appeared in all metrics.

Average values of the metrics employed in diversity analysis of both biopsies and water samples were however followed by relatively high errors obscuring potential significant conclusions.

3.3 Statistical testing of biopsy and water sample data set

The inclusion of the Kruskal-Wallis command in QIIME, resulted in several OTUs responsible for differences between the groups within the biopsy and water samples respectively. Significant OTUs having a Bonferroni corrected p-value <0,05 are enlisted in appendix G, for all groups tested.

3.3.1 Statistical analysis of biopsy data set

Kruskal-Wallis detected 23 OTUs implicated to be of significance for differences between the diagnosis groups, all presenting Bonferroni-corrected p-values <0,05. These OTUs stem from a variety of different phyla. Most, notably this includes members of the Firmicutes followed by Bacteroidetes, Proteobacteria and to a smaller extent Actinobacteria, Cyanobacteria and Tenericutes. Within the Firmicutes, members of the Clostridiales, such as *Lachnospiracheae* and *Ruminococcus* and *Erysipelotrichiae* seem to dominate. A more detailed description of the output of this analysis can be seen in appendix G, table 7.1.

Testing for interactions and group differences

Testing of interactions by ASCA-ANOVA showed the most significant interaction between age and diagnosis. For more details, see appendix L. By inspecting the loading plot, we found OTU 4 (*Enterobacteriaceae*) as the most important. Owing to the confined nature of this thesis, results of the biopsy dataset primarily connected to this OTU will be presented.

OTU 4 was submitted to further statistical analysis by Kruskal-Wallis to test for intragroup differences. The resulting p-value of 0,000 confirmed differences between amalgamated groups of age and diagnosis. Since Kruskal-Wallis do not detect where potential intragroup differences occur,

further statistical testing by the Conover-Inman method for pairwise comparisons was implemented. The results of the Kruskal-Wallis test and the significant results of the Conover-Inman test are illustrated in figure 3.6 below. For more detailed description of the output from these analyses, see appendix C and D respectively.

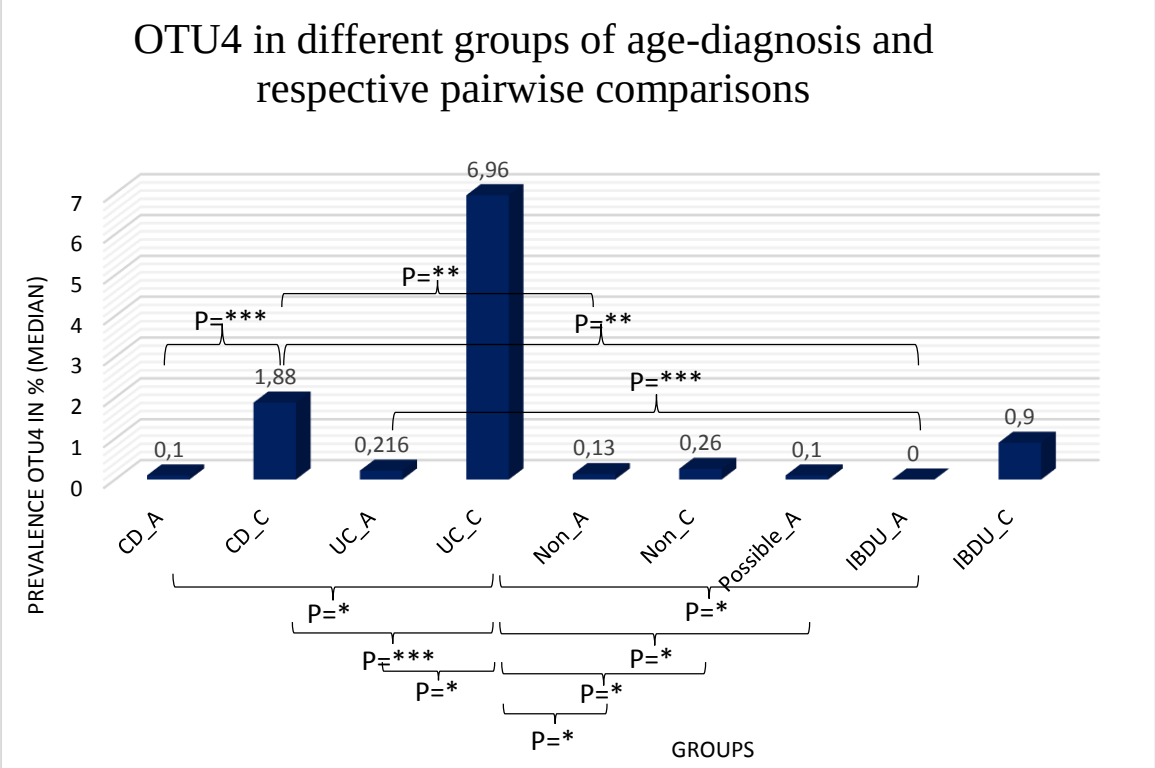


Figure 3.6: shows the median percentagewise prevalence of OTU4 in amalgamated groups of age and diagnosis, and the pairwise comparisons identified as significant from the Conover-Inman test. *P-value = 0,000 **P-value = <0,01 ***P-value = <0,05

The figure reveals a large variation with respect to OTU 4 prevalence in the groups, where UC_C present the highest prevalence, followed by CD_C and IBDU_C. The pairwise comparisons further disclose a significant increase of this OTU in UC_C compared to all other groups, with the exception of the IBDU_C cohort. The figure further illustrates an increased prevalence of OTU 4 in CD sufferers of the pediatric cohort compared to adult counterparts and adult controls. In summary, figure 3.6 show an increased prevalence of OTU 4 (*Enterobacteriaceae*) in the pediatric IBD cohort, especially UC sufferers, compared to the other groups included in this analysis.

Attempts to unveil potential taxonomic identification of OTU 4 a lower hierarchical level using BLAST, identified several potential matches affiliated to the *Escherichia/Shigella* genus. *E.*

fergusonii (n=3), *E. coli* (n=4), *S. sonnei* (n=2) and *S. flexneri* (n=1) all presented a query cover and identity of 98% and 99% respectively, and an E-value of 8e-168

OTU 4 in inflammation and age

Beta diversity output in QIIME using Bray-Curtis distance metric revealed that the relative abundance of *Enterobacteriaceae* was more than twice as large in inflamed tissue, as opposed to non-inflamed tissue, presenting a prevalence of 5,89% and 2,87% respectively. Further statistical testing was therefore performed on OTU 4. Testing for differences between amalgamated groups with respect to inflammation categories and age with Kruskal-Wallis, manifested significant differences at p=0,000. Conover-Inman test was further performed to uncover potential intragroup differences. The significant results can be seen in figure 3.7. A more detailed output of the Kruskal-Wallis and Conover-Inman analysis is given in appendix E and F respectively.

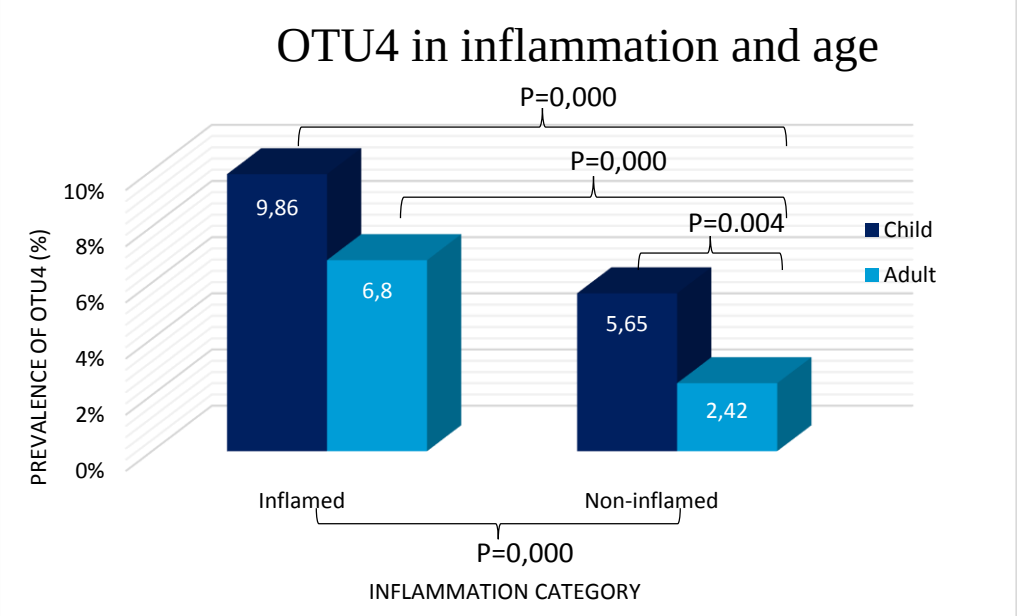


Figure 3.7:shows the percentagewise prevalence of OTU 4 in inflamed and non-inflamed tissue from both adults and children, and the respective p-values between the groups as measured with Conovan-Inman test.

As displayed in figure 3.7, Conover-Inman analysis revealed a difference between non-inflamed and inflamed tissue, where OTU 4 was significantly enhanced in the latter. (p=0,000). Children with inflamed tissue further displayed significantly more of this OTU than adults with non-inflamed tissue. With respect to non-inflamed tissue, at p=0,004 the pediatric cohort presented significantly more OTU 4 than adults. Differences between inflamed tissue of adults and children showed no significant results at a 5% level. Nor did analysis of differences in inflamed and non-inflamed tissue of children. However, at a p-value of 0,000, a difference in OTU 4 prevalence was detected between

the respective types of tissue in adults. Figure 3.7 further shows that OTU 4 constitute an abundant OTU in the categories enlisted as the percentages of this OTU is relatively high. Although not shown in figure, Conover-Inman also revealed a significant difference between adults and children at p=0,001.

3.3.2 Statistical analysis of water sample data set

Analysis by Kruskal-Wallis did not detect OTU 4 or any other OTUs of the *Enterobacteriaceae* to be significantly important for differences in any of the water sample groups. 18 OTUs were however detected to be of significance for differences between the diagnosis groups, all presenting Bonferroni-corrected p-values <0,05. Of these, 13 of 18 taxonomic groups are attributed to Proteobacteria, most notably the alpha and delta lineage.

3.4 Overlapping OTUs between water and biopsy data set

Analysis in MATLAB® detected 310 possible matches between OTUs from water sample and biopsy data set, where each match presented ≥ 97% sequence similarity. A complete list of all matches, their respective taxonomic annotation and distance in terms of Jukes-Cantor measurements, is given in appendix H.

Without taking OTU prevalence into consideration, the relative distribution on phyla level in all matches present itself as following: 50,6% Firmicutes, 27,7%, Proteobacteria 12,3%, Bacteroidetes, 5,5% Actinobacteria, and 3,9% of other phyla. The matches comprised 230 and 241 unique OTUs from the water sample and biopsy data set respectively, meaning that a single OTU potentially had several matches. The taxonomic identification of the 50 most abundant water sample OTUs (all presenting an average of ≥3 sequences and a prevalence ≥ 0,1%) holding a match can be seen in figure 3.8 below.

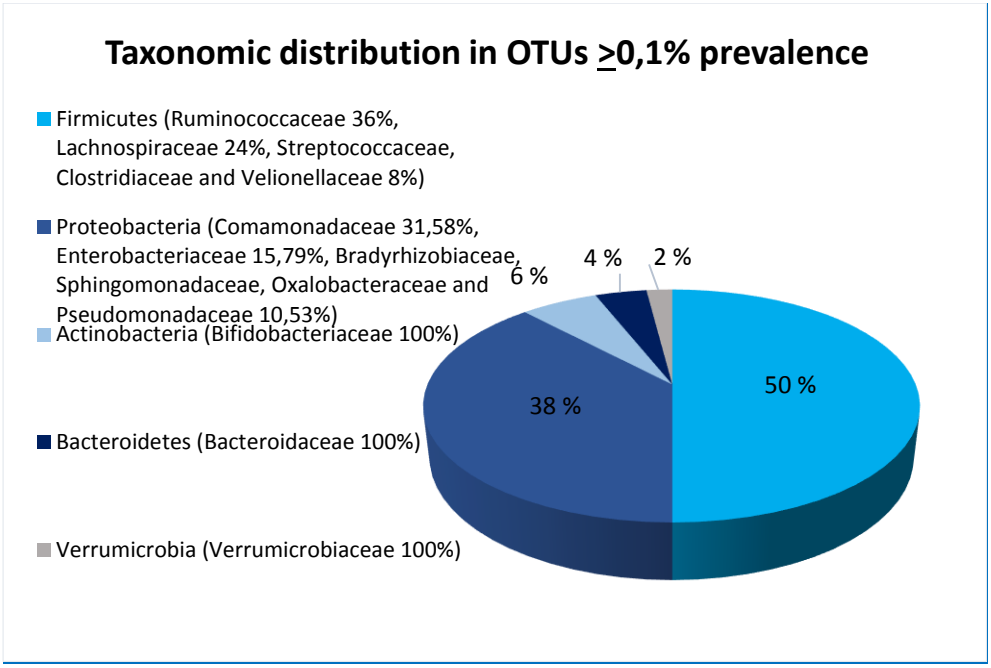


Figure 3.8: The figure shows the relative distribution on phyla level in the 50 most prevalent water sample OTUs (all ≥ 0,1%) holding a match. Distribution on family level within each phyla is given in parenthesis.

Of the Proteobacterial phylum, the distribution between Alpha, Beta and Gammaproteobacteria equaled 26,3%, 42,1% and 31,6% respectively. As the figure shows, there is a conspicuous dominance of bacteria belonging to Firmicutes, especially from the *Ruminococcaceae* and *Lachnospiraceae* family. Proteobacteria seem to exert a rather high dominance in the most abundant water OTUs as well presenting *Comamonadaceae* and *Enterobacteriaceae* as it most prevalent family members. Actinobacteria, Bacteroidetes and Verrumicrobia are also present, but in lower amounts.

3.5 Associations between water and biopsy OTUs

Owing to the complexity of performing exhaustive research on all identified matches, only selected matches from the water sample data set, the biopsy data set, and from the Jukes-Cantor data set were submitted to further analysis, as described in section 2.4.3 in materials and methods. Of 128 patients presenting both water and biopsy samples, information from 113 patients (88,3%) could be used for analysis of potential associations with Fisher exact, as some patients were removed during the initial quality filtration step in QIIME.

3.5.1 Matches determined by water and biopsy data set.

None of the OTUs implicated to be of significance for differences in the diagnosis groups within the water sample data set presented any matches with OTUs of the biopsy data set. Thus, further analysis to test for potential transmission of OTUs from water to mucosa was not performed on these OTUs. In the biopsy data set, six of the OTUs from the Kruskal-Wallis analysis held matches to the water sample data set. These matches are shown in table 3.1 below. Interestingly, OTU 4 from the biopsy data set presented a match to OTU 2 from the water sample data set. OTU match 4/2 is therefore included in further analysis as OTU 4 was involved in several findings of significance during the statistical analysis of the biopsy data set.

Table 3.1: The table shows OTUs identified as matches by the Jukes-Cantor method, from the biopsy OTUs implicated to be of significance in differences in diagnosis groups. Match 4/2 is included. OTU prevalence (%) is given in parenthesis. Standard deviation of biopsy OTU is also given along with taxonomic annotation

OTU-matches and prevalence (%)			
Biopsy	**Water**	**St.dev***	**Taxonomy†**
19 (1,349)	**1196** (0,039)	2,29	p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia(w‡)
56 (0,459)	**271** (0,179)	0,75	p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae
288 (0,07)	**1545** (0,012)	0,72	p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides
179 (0,039)	**179** (0,28) **217** (0,275)	0,11	p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis (w‡)
582 (0,035)	**271** (0,179)	0,14	p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae
572 (0,0005)	**1025** (0,0036)	0,007	p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae
4 (5,546)	**2** (0,1513)	13,8	p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae

*Standard deviation biopsy OTU with respect to percentagewise prevalence in all biopsies.
† k = kingdom, p = phyla, c = class, f = family, g = genus, s = species
‡ Taxonomic annotation on given level, is only attributed to water sample.

Four of the matches belong to the Firmicutes and to the order of Clostridiales and Erysipelotrichales respectively. All matches presents relatively low standard deviations, with the exception of match 4/2. Matches were further submitted to statistical analysis by Fisher exact method to test for possible associations between biopsies and water samples. As OTU 572 presented a very low prevalence and only could be identified in two biopsies, this match was excluded from further analysis. The result of this analysis is shown in figure 3.9 below. For details, see appendix J.

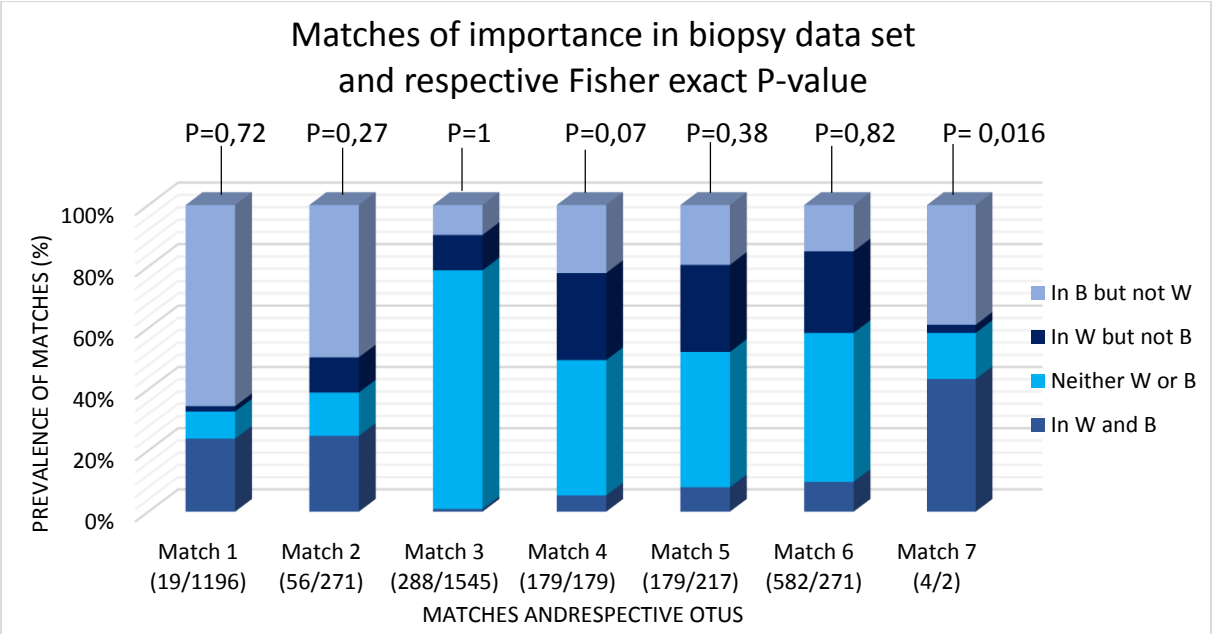


Figure 3.9: shows Fisher exact value of matches identified as being of importance from the biopsy data set, and percentage of which the respective OTU could be identified in none, both or one of the water and biopsy samples. OTU match is given in parenthesis.

In match 288/1545, 179/179, 179/217 and 582/271, the reference sequence of the respective OTUs could not be identified in either of the sample types in the majority of the patients. Only a minority of the patients presented matches in both sample types, with the exception of match 4/2 and to some extent 19/1196 and 56/271. A p-value above 0,5 for the six first matches indicate no significant associations between water samples and biopsies with respect to these OTUs. A p-value below 0,5 for match 4/2 however indicates a plausible association of *Enterobacteriaceae* in the113 patients encompassed by this analysis.

3.5.2 Matches determined by Jukes-Cantor data set.

To prevent spurious water OTUs from introducing potential biases in the subsequent analysis steps, the top 50 water OTUs presenting matches were plotted in decreasing order based on its percentagewise prevalence, and a threshold was determined. These OTUs are given in appendix K. As can be seen in figure 3.10, an OTU prevalence of 0,62% seem to mark the transition from a steep to a more continuous decline in prevalence, thus serving as a threshold.

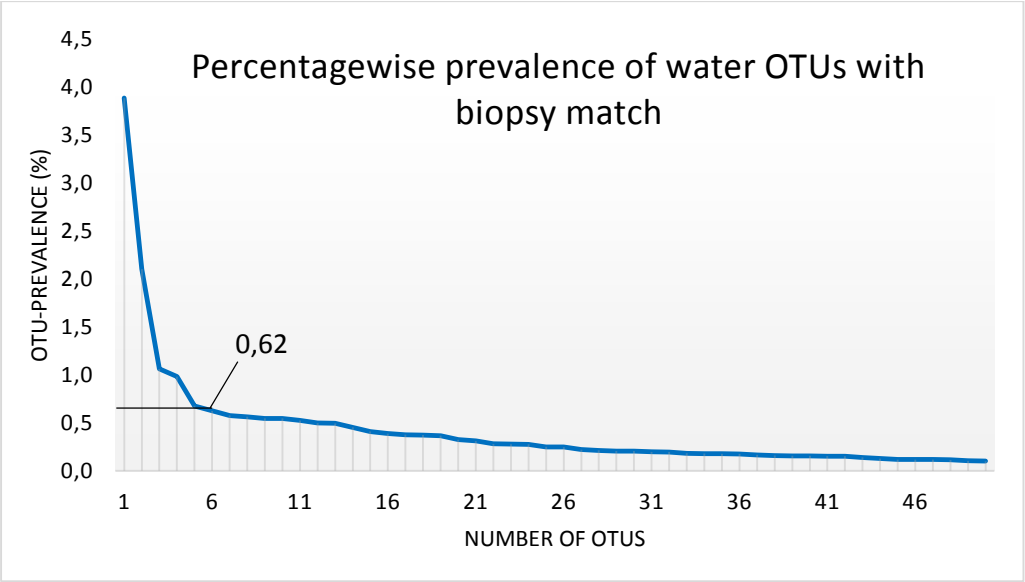


Figure 3.10: shows the percentwise prevalence of water OTUs presenting an abundance > 0,01% and having matching reference sequence to one or more biopsy OTUs.

Threshold determination resulted in seven OTU matches of six different taxonomic annotations implicated not to be of spurious origin. These are shown in table 3.2 below. In addition, conspicuous characteristics were seen amongst several of the OTU matches below threshold, owing to their relatively high prevalence in both water samples and biopsies. This includes OTUs from the family *Ruminococcaceae* (n=4), *Lachnospiracheae* (n=3), *Enterobacteriaceae* (n=1) and *Bacteroidacea*e (n=1), all presenting a prevalence of > 0,59% in biopsies and > 0,1% in water samples. The first two families encompass findings of *Faecalibacterium prausnitzii* (n=1), *Ruminococcus gnavus* (n=1), *Roseburia* (n=1) and *Blautia* (n=1) amongst others. Worthy of a comment is the observation that the match below threshold designated as *Enterobacteriaceae* is of the same OTU as previously tested.

Table 3.2: The table shows OTUs in biopsies and water identified as matches by the Jukes-Cantor method, when a prevalence threshold of >0,62% is employed for water OTUs. Respective OTU prevalence (%) is given in parenthesis. Standard deviation of water OTU is also given along with taxonomic annotation.

OTU-matches and prevalence (%)			
Biopsy	**Water**	**St.dev***	**Taxonomy†**
710 (0,004)	**6** (3,885)	10,0	p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Polaromonas (w‡)
778 (0,006)	**24** (2,104)	5,0	p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Bradyrhizobium (b‡)
461 (0,0013)	**112** (1,061)	4,3	p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b‡)
623 (0,001)	**22** (0,982)	3,7	p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Oxalobacter (b‡)

145 (0,049) **192** (0,02)	**40** (0,675)	1,0	p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
891 (0,0006)	**9** (0,625)	4,8	p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas;s__yabuuchiae

*Standard deviation water OTU with respect to percentagewise prevalence in all water samples.
† k = kingdom, p = phyla, c = class, f = family, g = genus, s = species
‡ Taxonomic annotation on given level, is only attributed to biopsy.

As displayed in table 3.2, five of the six most prominent OTUs belong to the phylum Proteobacteria while one belong to Firmicutes. Half of the OTUs are encompassed by the order Burkholderiales. The prevalence of the OTUs in each match belonging to the biopsy data set seem to be convincingly lower than what is observed in the water samples. Standard deviations for all water OTUs are relatively high, with the exception of the OTU 40, *Streptococcus*. Further attempts to identify OTU 40 on a lower taxonomic level in BLAST resulted in several potential matches, most notably *S. pseudoporcinus* and *S.suis*, each presenting an identity of 99% and 98% respectively. They further present E-values of 7e-164 and 3e-163respectively, and query covers of 99%.

To test for plausible associations, matches from table 3.2 were subject to statistical analysis by the Fisher exact method. The results are given in figure 3.11 below. For details, see appendix I.

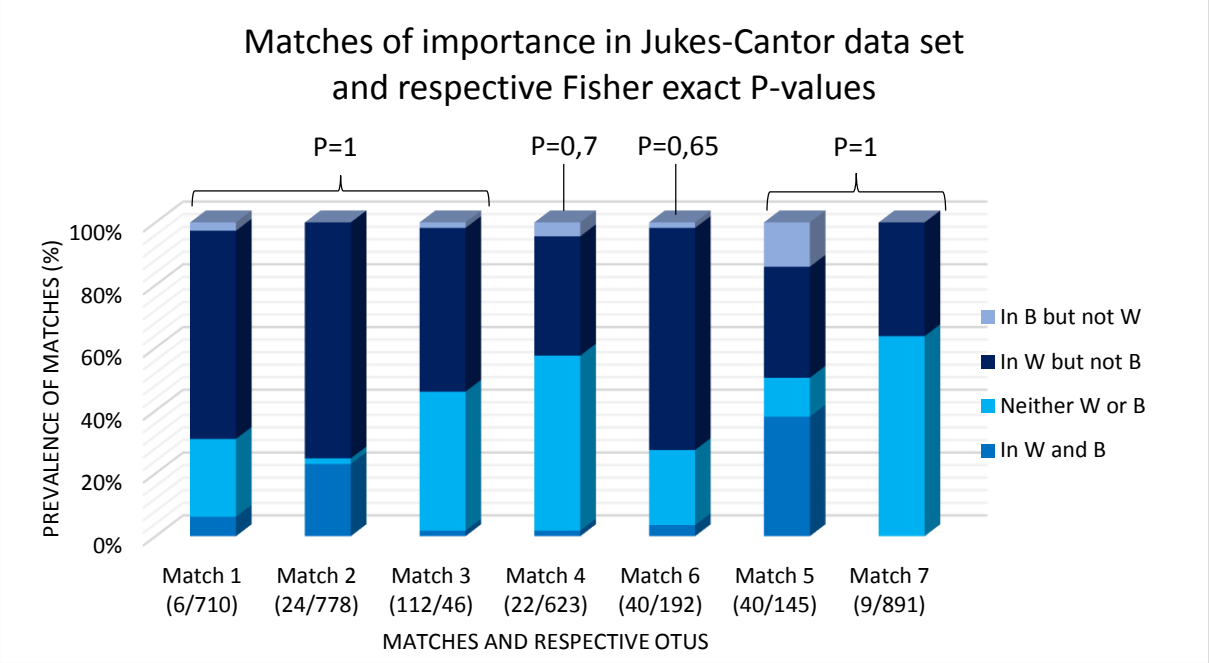


Figure 3.11: shows the Fisher exact value of the matches identified as being of importance from the Jukes-Cantor data set, and the percentage of which the respective OTU could be identified in none, both or one of water and biopsy samples. OTU match is given in parenthesis.

As figure 3.11 shows, the majority of patients presented reference sequences from the OTUs in each match in the water samples, but not in the biopsies. Only a small fraction of each OTU could be detected in both sample types, with the exception of match 40/145 and to some extent match 24/778 (*Streptococcus* and *Bradyrhizobium* respectively). Overall, it also seems like sequences were completely absent in both sample types in a large share of the patients within each match, with the exception of the two abovementioned matches. All of the Fisher exact p-values are >0,05, thus indicating no significant associations between the matching OTUs in water and biopsy samples.

3.5.3 Match 4/2 with respect to diagnosis.

To determine if the association with respect to match 4/2 from figure 3.9 could be attributed to a certain status of diagnosis, the match was decomposed into separate groups, and a Fisher exact test was performed. Of the 113 patients used for the previous Fisher exact testing, 101 was used for this purpose as 12 patients were excluded due to a status of diagnosis marked as unknown or possible. The result of this analysis can be seen in figure 3.12 below. The analysis of this OTU match from figure 3.9 is included for comparison reasons, as this analysis is performed on all diagnosis groups.

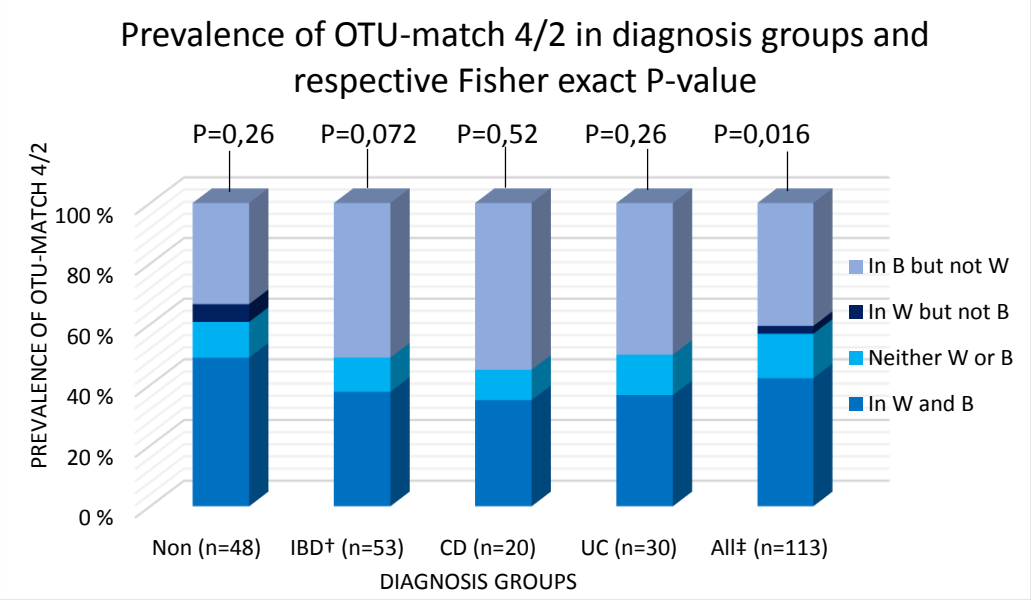


Figure 3.12: shows the Fisher exact value of the match 4/2 implicated to have an association between water and biopsy samples, and the percentage of which this OTU could be identified in none, both or one of water and biopsy samples in the different diagnosis groups.
† IBDU is included in this group together with CD and UC.
‡ Includes all diagnosis groups as given in figure 3.9.

As the figure shows, none of the diagnosis groups present significant Fisher exact values with respect to OTU match 4/2 when analysed separately. Compared to the group representing patients with all diagnoses, the fraction of the patients having OTU match 4/2 in both water and biopsies from the Non-group, is slightly higher. This observation does not apply to the three groups associated with IBD however as this fraction is somewhat lower. The overall distribution in the water and biopsy samples between the respective groups does however appear to commensurate. Although not below 0,05, the p-value of the IBD group is somewhat lower than that of the control group.

4.0 Discussion

4.1 Transmission of enterobacteria from water to mucosa?

Of great interest, is the observation that match 4/2 (*Enterobacteriaceae*) presents a Fisher exact p-value <0,05, suggesting a plausible association between water samples and biopsies with respect to this OTU, as shown in figure 3.9. Of further interest is the fact that this is the same OTU accountable for several compelling results revealed during analysis of the biopsies, especially with respect to its preeminence in pediatric UC patients, as will be discussed.

4.1.1 Tap water as a potential causative agent for the precedence of OTU4

If tap water is to serve as a potential causative agent for the observed precedence of OTU 4, the hypotheses regarding potential connections could primarily seek to explain this linkage by direct or indirect means, as mentioned in the introduction. Our findings of the OTU match 4/2 having a possible significant association in tap water and biopsies, as measured by the Fisher exact method, might give support to the theory of colonization by direct means. When considering this hypothesis, a few words with respect to the presence of this taxonomic family in drinking water should be mentioned.

Enterobacteria in drinking water

Enterobacteriaceae comprise a large family of the Gammaproteobacteria and are G-, non-spore forming, and facultative anaerobes with some exceptions. The members of this family can be of both nonpathogenic and pathogenic nature. (Mukhopadhya et al. 2012) (www.ilsi.org) As of today, microbial research with respect to the presence of this family in drinking water is scarce. An exceedingly large fraction of available research seem to encompass the characterization of the extended-spectrum beta-lactamases (ESBL) producing members. Albeit being of great significance to human health, this topic is somewhat beyond the scope of this thesis. Furthermore, as members of the *Enterobacteriaceae* are common inhabitants of the gut of animals and humans, studies detecting the presence of *E.coli* and coliforms of the species *Escherichia, Citrobacter, Enterobacter* and *Klebsiella,* is normally performed to assess the safety of drinking water. (Pepper et al. 2015) (www.who.int) The Norwegian drinking water act oblige the detection of *E.coli* and other coliforms to be 0/250ml water. (www.lovdata.no)

Of the few studies assessing enterobacterial identification in water, reservoirs of developing countries in Asia and Africa with different conditions from Norwegian reservoirs seem to be favored. However, a research project investigating the presence of *Enterobacteriaceae* in Polish ground waters, presented large variations with respect to the number of colonies/100ml water, ranging from 0 to several hundreds. In this study *S. marscescens, P. vulgaris, C. freundii* and *E.coli* comprised the most predominant enterobacterial species. (Golas et al. 2002) Although the majority (90%) of the Norwegian water source come from surface waters, and ground waters only represent the remaining 10%, (www.norskvann.no) it is likely to assume that members of this family might be present in Norwegian water sources as well. Both pathogenic and nonpathogenic members of *Enterobacteriaceae* can be found in water for domestic purposes, although members of the primary group are rare. This could be affiliated to the fact that several pathogens, such as those belonging to *Enterobacter, Helicobacter, Shigella* and *Klebsiella* are relatively sensitive to disinfectants and unstable in aquatic environments, although bacteria from the latter genus have shown an ability to create biofilms and grow in distribution systems. (www.who.int)

4.1.2 A direct transmission of OTU 4 from water to mucosa?

As touched upon in the introduction, the GI tract of IBD patients presents itself with lower stability and diversity, which is often accompanied by a diminished resilience to colonization by new bacteria. (Lozupone et al. 2012) Therefore, it might seem reasonable to consider the possibility that that bacteria from ingested food items, including tap water, more easily will colonize a GI tract characterized by a degraded microbiota by direct mechanisms. Furthermore, the possibility that apparently insignificant amounts of microbes such as *Enterobacteriaceae* ultimately may amount to large enough numbers to pose a threat on human health when consumed in considerable amounts over a long period of time, should not be excluded. This could especially apply for individuals presenting a microbiota easily prone to perturbations such as IBD sufferers. Thus, if *Enterobacteriaceae* in tap water is to be associated with IBD, it is not unlikely that a person holding a degraded microbiota and/or genetic susceptibility genes might be strongly influenced by tap water as an environmental factor.

4.2 Enterobacteria in IBD

4.2.1 Where do we stand so far?

As of today, researching efforts with respect to the microbiota in pediatric IBD seem to remain scarce, compared to adults. Findings regarding *Enterobacteriaceae* in the mucosa of children with CD seem to be deviating as both increased amounts (Gevers et al. 2014) and no significant increases have been reported. (Hansen et al. 2012) Regarding the prevalence of this taxonomic group in pediatric UC patients, even less is information seem to be available. In the more extensively studied microbiota of adults however, there seem to be a common acceptance that *Enterobacteriaceae* is increased in the mucosal samples of CD patients. (Chen et al. 2014) (Kabeerdoss et al. 2015) (Walker et al. 2011) (Willing et al. 2010) Although our results present enterobacterial increases, the lack of significance between the adult CD cohort compared to the healthy adults (figure 3.6) is unexpected, yet interesting.

4.2.2 Potential mechanisms for enterobacterial thrift in IBD

Whether the increased prevalence of bacteria within this family is a cause of the inflammatory response seen in IBD patients, or simply a result of the inflammatory milieu, remains to be elucidated. With respect to the latter point of view, it has been shown that inflammation of the GI tract has the ability to promote an increase of reactive nitrogen species (RNS) giving an accumulation of nitrate as a by-product. This substance have demonstrated to be utilized by *Enterobacteriaceae* for growth. (Winter et al. 2013) In addition, inflammation can also lead to an increase of oxygen to an otherwise oxygen-depleted environment. This could be explained by enhanced flow of water to the lumen as a result of diarrhea, leakage of oxygen-rich blood, or as a result of oxidative bursts such as the release of ROS by neutrophils. (Rigottier-Gois 2013) Furthermore, ROS have also shown to interact with other substances and create terminal electron acceptors that potentially could support the growth of pathogenic members of the *Enterobacteriaceae* family, such as Salmonella. (Winter et al. 2010) Thus, nitrogen and/or oxygen might possess the ability to promote a dysbiotic microbiota by suppressing the growth of obligate anaerobes whilst developing a niche where facultative anaerobes like *Enterobacteriaceae* could flourish.

4.2.3 Enterobacteria and its preeminence in pediatric IBD

A compelling observation is the precedence of OTU 4 in children with UC, and to a certain extent in children with CD, as the first group show significantly more of this OTU than all adult groups

and pediatric controls. (All p-values = 0,000) It is known that enterobacteria and other facultative anaerobes are amongst the first colonizers of the infant gut, and small children often present augmented levels of this family in comparison to adults. Thus, it might apparently seem feasible to explain our observed differences between children and adults with respect to OTU 4 by natural colonization. However, observations of the microbial profile of children resembling an adult profile after only a few years (Rodriguez 2015) and the fact that the children enrolled in this project are up to 18 years of age, makes it unlikely that natural succession present a significant influence on our results.

Phenotypically speaking, inflammatory bowel disease in children manifest itself in a somewhat different manner compared to adults, with more extensive intestinal involvement and increased severity of the disease being observed in several studies. (Langholz et al. 1997) (Limbergent et al. 2008) (Pigneur et al. 2010) With previous discussion of inflammation and ecological niches in mind, this disparity could potentially result in different environmental conditions within the GI tract of adults and children, where increased amounts of oxygen is more pronounced in the latter. Thus, a potential theory of the preeminence of OTU 4 in the pediatric IBD cohort might be that there has been a change of milieu driven by inflammation, with a subsequent development of a beneficial niche. Consequently, the GI tract of children with IBD might serve as a better habitat for the aerotolerant *Enterobacteriaceae*. On the other hand, there is a possibility that the gut microbiota of children is less recalcitrant to perturbations, such as a potential invasion from a pathogen, owing to a lower alpha diversity. (Rodriguez 2015) Thus, the fact that the precedence of OTU 4 might be a result of a bacterial invasion of members from this family should not be excluded, especially as the pediatric cohort displayed a decreased alpha diversity compared to the adult cohort.

4.2.4 Is there a connection to the extent of inflammation?

Production of large amounts of ROS by phagocytic cells such as macrophages and neutrophils is a natural response to pathogens as these oxidizing oxygen metabolites are toxic to infectious agents. (Mittal et al. 2014) If not properly controlled, these oxygen metabolites might also lead to damage on the host cells, as seen in inflamed tissue. The amelioration normally provided by antioxidants in the mucosa of healthy individuals, have shown to be impaired in inflammatory bowel diseases. (Kruidenier et al. 2003) Thus, it is reasonable to believe that the amount of ROS are increased in inflamed tissue as compared to non-inflamed tissue.

Although not investigated in this project, such observations could potentially apply to the biopsies used in this research as well. Some bacteria are equipped with different mechanisms for circumventing the toxic effects of these free oxygen radicals. For instance, it has been demonstrated that *E.coli*, when exposed to dissolved O_2, has the potential to express a manganese-superoxide dismutase (SOD) ameliorating the harmful effects of O_2^- by conversion to H_2O_2. (Baez & Shiloach 2013) Consequently, the theory regarding oxygen as an encourager of the growth of facultative anaerobes could potentially also relate to our findings of enterobacterial increases in inflamed tissue as opposed to non-inflamed tissue. (p=0,000) Although not displaying significant decreases, the alpha diversity of biopsies of inflamed category was found to be abated. Thus, a potential impaired recalcitrance to enterobacterial colonization might serve as a collateral explanation to the observed increases in this tissue.

Why the enterobacterial increases in inflamed tissue presents significance in adults (p = 0,000) and not in children is hard to tell. It has been demonstrated increased levels of SOD in adults in comparison to their younger counterparts during a GI infection with *S. flexneri*, (Raqib et al. 2000) potentially indicating a better ability of adults to circumvent the toxic effects of ROS. However, the complexity of immune responses and the seemingly lack of research with respect to differences in mucosal immune responses in the gut at different age groups, makes a proposition of an explanatory theory difficult. The enhanced levels of OTU 4 in both inflamed and non-inflamed tissue of the pediatric cohort as opposed to the non-inflamed tissue of the adult cohort is of interesting remark. With previous discussion in mind, it is possible that the inflamed tissue of children presents a more advantageous niche for the growth of *Enterobacteriaceae* compared to the non-inflamed region of adults. One plausible explanation for why the enhanced levels of OTU 4 extend to non-inflamed tissue of the pediatric cohort as well might be that these individuals present a microbiota of lower stability owing to their reduced alpha diversity. (Figure 3.4 A)

4.2.5 Could the precedence of OTU 4 be explained by AIEC?

Of the *Enterobacteriaceae*, *E.coli* seem to be the bacterium implicated to be associated with GI diseases such as IBD the most (Mukhopadhya et al. 2012) and elevated levels of antibodies against its O-antigen have been reported in IBD patients. (Tabaqchali et al. 1978) Observations that some strains possess invasive and proinflammatory properties, give support to theories embracing its role as an inducer of inflammation. Of most interest is the CD-associated AIEC, which has the ability to translocate across the mucosal barrier to the submucosa where it can invade and replicate within

macrophages. This bacterium has been detected in 29-36% of CD patients compared to 12-19% and 3-9% of UC-patients and controls respectively. (Mukhopadhya et al. 2012) Its presence lead to increased excretions of the proinflammatory cytokine TNF-α, thus provoking further immune responses. (Glasser et al. 2001) Other virulence factors include long polar fimbriae which it employs for stimulation of Peyer`s patches. (Chassaing et al. 2011)

Since the resolution of OTU 4 is only applicable for taxonomic assignment on family level, making assumptions of potential denotations on a lower hierarchical level is difficult. Although the finding of OTU 4 possibly being affiliated to species of the *Escherichia/Shigella* genus could be legitimate, this observation should be interpreted with caution as the polymerase used for this research project emanate from *E.coli*. Thus, there is a chance that amplification artifacts of DNA traces from this bacterium, might have posed an impact on the taxonomic annotation in BLAST. Further discussion of OTU 4, will therefore not take this finding into consideration.

Observations of OTU 4 being more predominant in pediatric UC patients than in pediatric and adult Crohn`s patients might imply that the enterobacterial precedence is not explained by the CD-associated pathogen AIEC. In addition, albeit not displaying a significant p-value, OTU 4 was slightly increased in adult controls compared to adults with CD, further suggesting the exclusion of this pathotype. *E.coli* associated with epithelial adherence have demonstrated to be significantly enhanced in the lamina propria of UC and CD patients. (Mylonaki et al. 2005) Thus, owing to the compromised nature of the mucosal barrier in IBD patients, it is also possible that the initiation of an immune response is caused a nonpathogenic member of the *Enterobacteriaceae* as well. However, estimating lysis intensity with respect to mucosal depth is difficult. In case of a situation where transmural lysis of profound bacteria deep within the mucosa has failed, there is a probability that analysis of potential invasive bacteria such as AIEC has been excluded from the subsequent steps.

4.2.6 Could a potential transmission of OTU 4 be attributed to IBD patients alone?

Our finding of an association between biopsies and tap water with respect to *Enterobacteriaceae* was only established when control patients and patients with IBD were viewed as an amalgamated unit. The fact that the association between water samples and biopsies with respect to OTU 4 did not apply to any of the diagnosis groups when analysed separately is somewhat surprising. Albeit being above the level of significance, the observation of the IBD group having a lower Fisher exact value than the Non-group is however of interesting remark. (Figure 3.12) Apparently, there might

seem as there is no association between biopsy and water sample with respect to *Enterobacteriaceae* in the different diagnosis groups. This dissension could however be explained by alterations in the dataset. Division of the 113 patients into different diagnosis groups led to each group being significantly smaller than when Fisher exact analysis was performed on all diagnosis groups combined. Thus, as the number of patients within each group decreased, the requirements of the observations became elevated in order to give a Fisher exact value below the level of significance.

4.3 Bacterial composition of OTU matches

The taxonomic distribution in the initial 310 OTU matches displayed a different profile than expected, much because the ratio of Firmicutes to Proteobacteria showed nearly a twofold magnitude of difference. This observation is most likely accredited to the fact that several of the OTUs comprised by the matches, could be present in spurious levels, thus not serving as legitimate representatives of the microbial community in tap water. As the requirements for OTU prevalence in the water samples increased to 0,1%, the taxonomic distribution converged towards a more anticipated profile, as OTUs of spurious origin theoretically were excluded.

4.3.1 Proteobacteria and its contributions to the microbiota of drinking water

Although the taxonomic composition of drinking water and water in distribution systems might differ according to initial water source, method of treatment, pH, availability of nutrients, dissolved oxygen and other biochemical compounds, there seem to be several conjectures applying to our results of the six most prevalent water OTUs. The observed dominance of Betaproteobacteria, echoes findings by other studies conducted with respect to the bacterial diversity of drinking water (Pinto et al. 2012) (Revetta et al. 2010) (Rudi et al. 2010) although some studies have reported a predominance of Alphaproteobacteria. (Lu et al. 2013) (Williams et al. 2004) Members of the α-lineage have been proposed to be less recalcitrant to disinfection compared to bacteria from the β-lineage, which to some extent might explain the predominance of the latter in this research.(Niemi et al. 2009) Somewhat surprising is the relative high prevalence of Gammaproteobacteria as this class has been proposed to be present in drinking water, but in modest amounts. (Rudi et al. 2010) (Vaz-Moreira et al. 2013) (Liu et al. 2014) This lineage has however been shown to predominate biofilms of distribution systems. (Douterelo et al. 2016) It should however be noted that the

methods employed for detection differed significantly between the respective researching projects used for comparison.

Observations of *Bradyrhizobiaceae* and *Sphingomonadaceae* of the α-lineage, Burkholderiales of the β-lineage, and *Pseudomonadaceae* of the μ-lineage being present in relatively high amounts, could to some extent have been expected, as members of these taxonomic division frequently have been detected in drinking water. (Berg et al. 2009) (Hwang et al. 2012) (Liu et al. 2014) (Martiny et al. 2002) (Vaz-Moreira et al. 2013) Thus, the fact that five of the six most prevalent water OTUs are affiliated to the first three divisions, as shown in table 3.2, could be considered reasonable findings. Our results of the *Polaromonas* of *Comamonadaceae* being the most prevalent OTUs, substantiate previous research on the microbiota of drinking water in distribution systems. In addition to being frequently isolated from granular activated carbon-filters employed for water treatment, (Magic-Knezev et al. 2009) it has been proposed that as much as 69% of the bacteria from the water in distribution systems can be affiliated to this genus. (Liu et al. 2014) A precedence of *Comamonadaceae* was also presented in a study by Martiny et al (2002) as they not only found this family to prevail the microbiota of biofilms, but the bulk water in distribution systems as well. As several members of this family are capable of denitrification, they constitute an important part of the microbiota of activated sludge. (Khan et al. 2002)

The sixth most prevalent OTU, *Sphingomonas*, seem to be another important member of drinking water as well as this genus has been reported to be of both dominating nature (Berg et al. 2009) and present in considerable amounts. (Martiny et al. 2002) (Liu et al. 2014) *Sphingomonas* has furthermore been reported to prevail the microbiota of biofilms, suspended solids and loose deposits in distribution systems. (Liu et al. 2014) This genus can be found in a wide spectrum of environments (Berg et al. 2009), and its survival in low nutrient, oligotrophic environments such as distribution systems, can be attributed to its enhanced uptake system. (Liu et al. 2014) The second and fourth most prevalent OTU, *Bradyrhizobium* and *Oxalobacteraceae*, are common inhabitants of the environment. Members of the latter family thrive in anaerobic environments such as fresh lakes and sediments, but can also be found in the rumen of several animals like sheep and cattle. (Garrity et al. 2004) The *Bradyrhizobiaceae* family can been detected in both drinking water (Vaz-Moreira et al. 2013) and BAC-filters (Niemi et al. 2009) while species of *Bradyrhizobium* also can be found in root nodules where they perform nitrogen fixation. (Garrity et al. 2004)

It is surprising that one of the six most prevalent water OTUs belong to Streptococcus, as this is a human associated bacterium known to include several members of pathogenic nature. *S. pseudoporcinus,* annotated by BLAST as the most likely strain, is considered to be relatively rare although it has been isolated from urine, skin, vaginal and rectal specimens. (Stoner et al. 2011) *S. suis* on the other hand is a common pig pathogen, normally residing in the upper respiratory tract of pigs. Although having the potential to serve as a human pathogen, this occurs with low frequency in Western countries. (Goyette-Desjardins et al. 2014) *Streptococcus* being a natural inhabitant of drinking water is unlikely, although other genera such as *Staphylococcus, Mycobacterium* and *Nocardia*, known to include pathogens, have been isolated from drinking water as well. (Berg et al. 2009) Still, its prevalence might be accredited to mechanisms of contamination. If the source of contamination is sewage leakage or transmission from patient to sample during the sampling process, one could probably expect a high variation and standard deviation between the samples. OTU 40 however only exhibited a standard deviation of 1%, indicating an equal distribution in the samples. Thus, the possibility of contamination during sample processing should not be excluded.

4.3.2 The water microbiota in relation to previous research

On a phylum level, our findings of Verrumicrobia, Actinobacteria and Bacteroidetes (figure 3.8) are in concordance with previous research as these phyla have been detected in drinking water. (Lu et al. 2013) (Vaz-Moreira et al. 2013) Albeit a further narrowing of the OTUs resulted in a profile in more compliance with the abovementioned expectations, there is a relatively high presence of several bacterial groups more common of the gut microbiota. In particular, this includes the detection of *Ruminococcaceae, Lachnospiracheae* and *Streptococcaceae* of the Firmicutes, *Bifidobacteriaceae* of the Actinobacteria, and *Bacteroidaceae* of the Bacteroidetes.

Although several of the taxonomic annotations of the OTU matches allegedly not are implicated to be associated with the natural microbiota of tap water, several gut associated bacteria have been found in filters of drinking water distribution systems. This includes *Ruminococcus, Lachnospira, Blautia, Roseburia* and *Faecalibacterium* found in Chinese drinking water. Interestingly, the amount of Firmicutes has also shown to be positively correlated with nitrite (Wu et al. 2015) thus illuminating the potential influence biochemical conditions might exert on microbial growth in water systems. Despite the likely dissimilarities in conditions affecting the drinking water microbiota in China and Norway, the possibility that such bacteria could be present in Norwegian distribution systems, should not be excluded. The fact that 50% of the 50 most prevalent OTUs are

affiliated with Firmicutes might however indicate that there are potential mechanisms of contamination involved as well.

4.4 Possible transmission of other OTUs from water to mucosa?

4.4.1 OTU matches from Jukes-Cantor dataset

The observed lack of significance between water and biopsies with respect to each OTU match, as measured with the Fisher exact method, might imply that there has not been a transmission of these bacteria from water to mucosa. The lower prevalence of each OTU match in the biopsies, as can be seen in table 3.2, is not surprising as the matches were sorted with respect to water sample prevalence. Furthermore, as the rationale was searching for potential transmission of OTUs from water to mucosa, it is reasonable to assume that only a small fraction of the transferred bacteria, are capable of adhering to and survive in the GI environment. The low prevalence of the OTU matches in the biopsies, is to some extent reflected in the results from figure 3.11, where relatively few of the reference sequences could be found in the biopsies. The strikingly high fraction of patients showing an absence of the reference sequence in both sample types, might indicate that the prevalence of the water OTUs are not equally distributed between the samples. Thus, it is possible that the presence of these OTUs are affiliated to only a few samples, especially since the respective standard deviations were relatively high. The possibility of a potential transmission of matches below the threshold of 0,62% should not be excluded as the OTUs above might be subject to bias.

4.4.2 OTU matches from biopsy data set

As figure 3.9 shows, the Fisher exact p-value of six of the seven matches defined as important in the biopsy dataset are above 0,05. Thus, none of these matches can be regarded as significant contributors to the theory of a possible association between water samples and biopsies. This can partly be explained by the fact that these matches to a relatively little extent can be recaptured in the water samples under investigation. As the taxonomic groups encompassed by these matches are highly typical of the gut microbiota and are most likely adapted to thrive in the gut where abiotic and biotic factors are expected to differ significantly from distribution system waters, this could have been expected. With this in mind, one could expect the OTU matches and its reference sequences to be present in biopsies to a larger extent than what can be observed from table 3.1 and figure 3.9 respectively. However, the allegedly low recapture of these matches in the biopsies can be explained by the fact that the OTU matches under investigation initially was defined by testing

with the Kruskal-Wallis method which do not take relative OTU prevalence into consideration when seeking to explain the variation within the diagnosis group. Thus, albeit being typical of the human gut, the prevalence of these OTUs do not reflect the actual prevalence of the respective taxonomic groups, as there presumably are several OTUs annotated to the same taxonomy. Furthermore, the OTUs implicated to be of importance in the biopsy data set was chosen based on the Bonferroni p-value, which is a more stringent correction of the p-value than for instance the FDR. Thus, employment of the latter correction could possibly have led to the annotation of more OTUs of significance and consequently more OTUs being subject to further testing with Fisher exact.

4.4.3 Could tap water introduce perturbations to the gut microbiota by indirect means?

The gut is an extremely complex ecosystem, with an immense amount of reactions and interactions in a subtle and delicate balance. In individuals where this balance is frail or easily altered, such as IBD patients, there could be a possibility that biotic and/or abiotic components of ingested items such as tap water, indirectly could disturb this equilibrium. This includes hypotheses of tap water bacteria producing metabolites or other substances affecting microbial growth directly or by indirectly creating an ecological niche for the thrift of certain microbial groups, such as members of the *Enterobacteriaceae*. The fact that tap water bacteria and their different metabolic processes could change the biochemical conditions of tap water such as levels of minerals like iron, magnesium and sulfur or their chemical form, should also be considered as a possible hypothesis as this could affect microbial growth as well. Albeit international databases such as the KEGG PATHWAY are under constant development, our current understanding of the interplay between metabolic pathways and how substances impinge upon microbes, is generally scarce.

4.5 Analysis in QIIME

4.5.1 Intragroup diversity analyses

Several of the findings from the alpha diversity analysis seem to be in corroboration with previous research, thus strengthening its reliability. Concerning estimates of the biopsy data set, the result of adults presenting a higher diversity than children is as expected as this is in concordance with the general belief. (Rodriguez 2015) As is our findings of the enhanced diversity in healthy controls, and UC biopsies as compared to CD biopsies. (Walker et al. 2011) (Bibiloni et al. 2006) Our diversity estimates of the anatomical sites is echoed by the general assumption of the cecum and

large intestine presenting a more stable environment for bacterial growth compared to the small intestine, where growth is impaired by gastric juices. (Lu et al. 2014) Alpha diversity in terms of species richness must not be confused with bacterial concentrations as these observations have shown different results in inflamed and non-inflamed tissues of IBD patients. While studies have presented augmented bacterial concentrations in inflamed tissue, (Swidsinski et al. 2009) the species richness in our and other studies, is found to be decreased. (Sepehri et al. 2007)

Interestingly, the increased species estimates of the majority of tap water samples compared to the biopsies, reflect the high diversity of the relatively undiscovered microbiota of water. The variation with respect to species number in the water samples can probably be accredited to the different conditions of the tap waters under investigation. As mentioned initially, the microbiota of tap water is influenced by several factors of the distribution system (Pepper et al. 2015) frequency of tap usage (Rudi et al. 2009) and also if the water originates from ground waters or surface waters. (Douterelo et al. 2016) The fact that water samples from CD patients present the lowest diversity is a finding deserving of a comment, as the mucosal diversity of these patients have shown to be depleted as mentioned. Although implying a potential link between these observations is a rather bold remark, the possibility of this being concomitant observations should not be excluded. Owing to difficulties of drawing significant conclusions from the diversity analyses, these results should be interpreted with caution.

The flattening of the rarefraction curves indicates that the cut-off values of 3000 sequences embrace most of the species from both the biopsies and water samples. The continued increase in the rarefraction curve of some of the water samples do however indicate that there are more species left to be captured in these samples. Although a higher cut-off value possibly would have given a more asymptotic shape of these curves, thus indicating an inclusion of nearly all species present, this would also have led to an exclusion of more samples from the sequencing analysis steps. Furthermore, 3000 sequences seem to be a decent amount as 1000 sequences has been suggested to be the appropriate minimum threshold to circumvent too much influence from different issues of quality. (Navas-Molina et al. 2013)

4.6 Library preparation and sequencing

As the qubit quantifications and qPCR values indicate, there are large variations of DNA and bacterial counts between the samples. The deviating amounts of DNA in water samples is most

likely explained by variations in the tap water distribution systems of these patients, as this can pose a great impact on bacterial growth. (Pepper et al. 2015)

The higher Q30 value of the water sample run (80,91) compared to the biopsy run (77,69) could be attributed to the lower cluster density of this run (577 K/mm^2) compared to 920 K/mm^2 in biopsy library. This is partially because a smaller space between the clusters increases the risk of a wrong base call due to overlapping signals, resulting in a lower Q30 score. However, as all quality parameters generated from the two Illumina runs in this project are within acceptable levels, the process of sequencing could could generally be regarded as being successful.

4.7 Critical appraisals and possible artifacts

Performing a research project without the introduction of possible biases has proven to be difficult as artifacts easily is introduced. With respect to this project, there are a few critical appraisals that should be raised owing to their potential to introduce biases.

4.7.1 Technical issues

Although mentioned briefly in the introduction, the potential repercussions of the modifications implemented in order to increase the yield of the nested PCR are worthy of a second reiteration. First, increased template DNA could have led to an impediment of the polymerase. Second, the increased number of cycles implemented on both water samples and biopsies might have generated non-specific products, which possibly could have led to the designation of these sequences to a novel OTUs. (Kennedy & Oswald 2011) (Nelson et al. 2014) Although not likely to affect the outcome of the most dominating species, these artificial OTUs might have led to other biases such as overestimations of alpha diversity. Third, albeit increased annealing temperature resulted in better primer specificity, there is a risk that the preclusion of eukaryotic sequences in the biopsies was achieved at the expense of the primers being too little sensitive. Thus, poorly characterized sequences often from species typical of environments outside the human body, might have failed to be amplified. Water sample bacteria potentially colonizing the mucosa of our patients could therefore have gone undetected or below a noticeable level during downstream analysis.

It should be mentioned that since the foundation for the analysis performed in this research project is DNA, there is a chance that several of the sequences designated into OTUs originate from dead,

non-viable bacteria, or remnants of DNA. In fact, it has been proposed that as much as 99% of the bacterial diversity in disinfected waters stem from non-viable or non-culturable bacteria. (Vaz-Moreira et al. 2013) As one of the main aims of this thesis was searching for a potential transmission of bacteria from water to mucosa, there is a chance that a part of the OTUs under investigation to some extent stemmed from non-viable bacteria.

Our current limited knowledge of bacterial species from environments outside the human body might also have posed an impact during the taxonomic designation of the analysed sequences, as relatively little information seem to be available in genome databases as of today. As Greengenes, the database used for this researching purpose, mostly comprise microorganisms associated with human health, the taxonomic designation of less known environmental bacteria might unfortunately have been subject to bias. Problems of taxonomic annotation of drinking water bacteria became particularly evident in a study where 57,6% of the partial 16S rRNA gene sequences could not be classified when using the ribosomal Database Project Classifier. (Revetta et al. 2010) Furthermore, the use of a closed OTU-picking strategy might have led to a failure to identify novel species as sequences not presenting a similarity $\geq$ 97% to those of the Greengenes database, were excluded from analysis. (Rideout et al. 2014) Consequently, species of novel nature, and possible important drinking water bacteria, might have been precluded from subsequent analysis steps.

4.7.2 Research design

Antecedent to this research, plausible artifacts could have been introduced already during the recruitment of patients. Albeit not presenting any pathological traits of the GI tract, the patients regarded as controls in this research project were initially enrolled to the IBSEN II study due to suspicions of inflammatory conditions. Although disproven to have IBD, there is a possibility that these patients might have other concealed GI illnesses potentially associated with dysbiosis, thus obscuring their value as healthy controls with normobiosis.

A few words with respect to the Jukes-Cantor model should also be mentioned, as this served as the method for identification of matches. As this model does not take into consideration whether the substitution occurs in variable or conserved regions of the sequence, all substitutions are treated equally. Theoretically, if two different alignments present an equal number of substitutions, but where the majority of these substitutions are located in variable and conserved regions respectively, both alignments will obtain the same evolutionary distance by the Jukes-Cantor method. As a result, it is possible that the evolutionary distance of the latter alignment might be slightly underestimated

compared to the distance of the first alignment where an overestimation might occur. Consequently, the identification of matches might have been subject to biases associated with distance measurements.

OTU matches under investigation in the Fisher exact analyses were submitted to the parameters given in section 2.5.2 in materials and methods. Thus, it should be noted that a change of these parameters such as increases in the requirements to the number of sequences present in a sample, could lead to different results of the statistical analysis.

4.7.3 Mechanisms of contamination

As to plausible mechanisms of contamination seeking to explain the relatively high presence of bacteria associated with human gut in the matches, several possibilities exist. Prior to this research, the distribution systems of several of the patients could have been subject to contamination from sewage. Contamination could also have occurred at the sampling step, as patients took their own samples without supervision of anyone with knowledge of sterile sampling techniques. Furthermore, as DNA from the water samples already had been extracted and purified before this project, the possibility that contamination of any kind could have occurred during this pre-processing should not be excluded. Furthermore, processing of materials took place in a lab where microbial research on fecal specimens frequently is performed, and where PCR is extensively used. Thus, DNA originating from bacteria normally residing within the gut could potentially have been present on benches, equipment, in dust etc and contaminated the samples as most work was performed on a working station in an open environment. Regardless of mechanism of contamination, as the number of cycles in the first PCR reaction was increased to 30 cycles it is possible that even the smallest traces of contamination could have had an impact on subsequent analysis steps, especially because the amount of DNA in the water samples initially were so low.

4.8 Concluding remarks

This primary aim of this research project was investigating the microbiota of tap water and biopsies, with the rationale of uncovering potential associations between the two habitats. We unveiled a significant association between tap water and biopsies with respect to OTU 4 of the *Enterobacteriaceae*. We further disclosed highly significant increases of this OTU in pediatric IBD sufferers, especially of the UC cohort. Further analysis also revealed augmented levels of this taxonomic group in biopsies of inflamed origin.

Whether our findings suggest that there has been a direct transmission of *Enterobacteriaceae* from drinking water to mucosa, remains a question in need of more research. The clear association between water and biopsies with respect to OTU 4, is however worthy of further attention, especially since this taxonomic group has been implied to be of significance in IBD in our and other studies. Hopefully, our findings of will serve as important contributors to further research within environmental and microbial aspects of IBD aetiology. Furthermore, our results of enterobacterial increases in the pediatric cohort will hopefully provide novel insights into the field of mucosal microbiota of younger IBD patients, especially UC sufferers, as this seem to be a relatively untouched area of investigation.

4.9 Future research

The extensive researching efforts implemented to reveal the cause of IBD have disclosed a disease of many faces. Still, the precise aetiology remains unknown. As advances in researching technologies proceed, one should expect a continuous unveiling of the mysteries behind this complex disease.

Future research aspiring for further understanding of the aspects of IBD and the factors contributing to its commencement, should account for several considerations. Discrimination with respect to IBD diagnosis should be performed, as UC and CD should be regarded as two distinct diseases with possible deviating aetiology and/or microbial key players. Further investigations should also seek to differentiate between adults and children as different age groups plausibly might display different aetiology or pathophysiology. Microbial research using DNA should include methods for distinguishing viable bacteria from non-viable if possible, especially if investigating potential transmission of bacteria. With respect to a broader context, future focus within the field of

microbiology should focus on incremented knowledge of how abiotic compounds such as metabolites and chemical substances affect microorganisms. Accommodation of such potential indirect impacts on the gut microbiota is possibly of great importance if environmental triggers of IBD is to be understood. Collectively, one should also seek to bridge the gap between the numerous bacterial species in these environments and the available information with respect to their genome sequences. In addition to being of importance to addressing several environmental questions, expanded information about these bacteria is most likely of great importance if questions regarding its implications in IBD and human health in general is to be ascertained

References

Aamodt, G., Bukholm, G., Jahnsen, J., Moum, B., Vatn, M. H. & Grp, I. S. (2008). The Association Between Water Supply and Inflammatory Bowel Disease Based on a 1990-1993 Cohort Study in Southeastern Norway. *American Journal of Epidemiology*, 168 (9): 1065-1072.

Alkanani, A. K., Hara, N., Lien, E., Ir, D., Kotter, C. V., Robertson, C. E., Wagner, B. D., Frank, D. N. & Zipris, D. (2014). Induction of Diabetes in the RIP-B7.1 Mouse Model Is Critically Dependent on TLR3 and MyD88 Pathways and Is Associated With Alterations in the Intestinal Microbiome. *Diabetes*, 63 (2): 619-631.

Allen, M. J., Edberg, S. C. & Reasoner, D. J. (2004). Heterotrophic plate count bacteria - what is their significance in drinking water? *International Journal of Food Microbiology*, 92 (3): 265-274.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., et al. (2011). Enterotypes of the human gut microbiome (vol 473, pg 174, 2011). *Nature*, 474 (7353).

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., Fukuda, S., Saito, T., Narushima, S., Hase, K., et al. (2013). T-reg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature*, 500 (7461): 232-+.

Avershina, E. & Rudi, K. (2013). Is it who you are or what you do that is important in the human gut? *Beneficial Microbes*, 4 (3): 219-222.

Avershina, E., Storro, O., Oien, T., Johnsen, R., Pope, P. & Rudi, K. (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *Fems Microbiology Ecology*, 87 (1): 280-290.

Avershina, E. & Rudi, K. (2015). Confusion about the species richness in human gut microbiota. *Beneficial Microbes*, In press.

Baez, A. & Shiloach, J. (2013). Escherichia coli avoids high dissolved oxygen stress by activation of SoxRS and manganese-superoxide dismutase. *Microbial Cell Factories*, 12.

Barreiro-de Acosta, M., Alvarez Castro, A., Souto, R., Iglesias, M., Lorenzo, A. & Dominguez-Munoz, J. E. (2011). Emigration to western industrialized countries: A risk factor for developing inflammatory bowel disease. *Journal of Crohns & Colitis*, 5 (6): 566-569.

Baumgart, D. C. & Carding, S. R. (2007). Gastroenterology 1 - Inflammatory bowel disease: cause and immunobiology. *Lancet*, 369 (9573): 1627-1640.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., Yantiss, R., Orsi, R. H., Wiedmann, M., McDonough, P., Kim, S. G., et al. (2007). Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *Isme Journal*, 1 (5): 403-418.

Berg, K. A., Lyra, C., Sivonen, K., Paulin, L., Suomalainen, S., Tuomi, P. & Rapala, J. (2009). High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. *Isme Journal*, 3 (3): 314-325.

Berry, D. & Reinisch, W. (2013). Intestinal microbiota: A source of novel biomarkers in inflammatory bowel diseases? *Best Practice & Research in Clinical Gastroenterology*, 27 (1): 47-58.

Bibiloni, R., Mangold, M., Madsen, K. L., Fedorak, R. N. & Tannock, G. W. (2006). The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. *Journal of Medical Microbiology*, 55 (8): 1141-1149.

Biedermann, L., Zeitz, J., Mwinyi, J., Sutter-Minder, E., Rehman, A., Ott, S. J., Steurer-Stey, C., Frei, A., Frei, P., Scharl, M., et al. (2013). Smoking Cessation Induces Profound Changes in the Composition of the Intestinal Microbiota in Humans. *Plos One*, 8 (3).

Biedermann, L. & Rogler, G. (2015). The intestinal microbiota: its role in health and disease. *European Journal of Pediatrics*, 174 (2): 151-167.

Bordin, M., D'Atri, F., Guillemot, L. & Citi, S. (2004). Histone deacetylase inhibitors up-regulate the expression of tight junction proteins. *Molecular Cancer Research*, 2 (12): 692-701.

Bouskra, D., Brezillon, C., Berard, M., Werts, C., Varona, R., Boneca, I. G. & Eberl, G. (2008). Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature*, 456 (7221): 507-U34.

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X. H., et al. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology*, 26 (10): 1146-1153.

Burisch, J. & Munkholm, P. (2015). The epidemiology of inflammatory bowel disease. *Scandinavian Journal of Gastroenterology*, 50 (8): 942-951.

Bustin, S. A., Benes, V., Nolan, T. & Pfaffl, M. W. (2005). Quantitative real-time RT-PCR - a perspective. *Journal of Molecular Endocrinology*, 34 (3): 597-601.

Cammarota, G., Ianiro, G., Cianci, R., Sibbo, S., Gasbarrini, A. & Curro, D. (2015). The involvement of gut microbiota in inflammatory bowel disease pathogenesis: Potential for therapy. *Pharmacology & Therapeutics*, 149: 191-212.

Carbonero, F., Nava, G. M., Benefiel, A. C., Greenberg, E. & Gaskins, H. R. (2011). Microbial DNA extraction from intestinal biopsies is improved by avoiding mechanical cell disruption. *Journal of Microbiological Methods*, 87 (1): 125-127.

Casen, C., Vebo, H. C., Sekelja, M., Hegge, F. T., Karlsson, M. K., Ciemniejewska, E., Dzankovic, S., Froyland, C., Nestestog, R., Engstrand, L., et al. (2015). Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Alimentary Pharmacology & Therapeutics*, 42 (1): 71-83.

Chassaing, B., Rolhion, N., de Vallee, A., Salim, S. Y., Prorok-Hamon, M., Neut, C., Campbell, B. J., Soderholm, J. D., Hugot, J. P., Colombel, J. F., et al. (2011). Crohn disease-associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. *Journal of Clinical Investigation*, 121 (3): 966-975.

Chen, L. P., Wang, W., Zhou, R., Ng, S. C., Li, J., Huang, M. F., Zhou, F., Wang, X., Shen, B., Kamm, M. A., et al. (2014). Characteristics of Fecal and Mucosa-Associated Microbiota in Chinese Patients With Inflammatory Bowel Disease. *Medicine*, 93 (8).

Claesson, M. J., Wang, Q. O., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P. & O'Toole, P. W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research*, 38 (22).

Claesson, M. J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J. R., Falush, D., Dinan, T., Fitzgerald, G., et al. (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences of the United States of America*, 108: 4586-4591.

Clarke, S. F., Murphy, E. F., O'Sullivan, O., Lucey, A. J., Humphreys, M., Hogan, A., Hayes, P., O'Reilly, M., Jeffery, I. B., Wood-Martin, R., et al. (2014). Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*, 63 (12): 1913-1920.

Coates, M. E. (1973). Gnotobiotic Animals in Nutrition Research. *Proceedings of the Nutrition Society*, 32 (2): 53-58.

Cuiv, P. O., de Carcer, D. A., Jones, M., Klaassens, E. S., Worthley, D. L., Whitehall, V. L. J., Kang, S., McSweeney, C. S., Leggett, B. A. & Morrison, M. (2011). The Effects from DNA Extraction Methods on the Evaluation of Microbial Diversity Associated with Human Colonic Tissue. *Microbial Ecology*, 61 (2): 353-362.

Davies, J. M. & Abreu, M. T. (2015). The innate immune system and inflammatory bowel disease. *Scandinavian Journal of Gastroenterology*, 50 (1): 24-33.

Dinan, T. G. & Cryan, J. F. (2013). Melancholic microbes: a link between gut microbiota and depression? *Neurogastroenterology and Motility*, 25 (9): 713-719.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (26): 11971-11975.

Douterelo, I., Jackson, M., Solomon, C. & Boxall, J. (2016). Microbial analysis of in situ biofilm formation in drinking water distribution systems: implications for monitoring and control of drinking water quality. *Applied Microbiology and Biotechnology*, 100 (7): 3301-3311.

Ekman, S. (1999). PCR optimization and troubleshooting, with special reference to the amplification of ribosomal DNA in lichenized fungi. *Lichenologist*, 31: 517-531.

Elson, C. O. & Cong, Y. (2012). Host-mmicrobiota interactions in inflammatory bowel disease. *Gut Microbes*, 3 (4): 332-44.

Emerenciano, V. P., Militao, J. S. L. T., Campos, C. C., Romoff, P., Kaplan, M. A. C., Zambon, M. & Brant, A. J. C. (2001). Flavonoids as chemotaxonomic markers for Asteraceae. *Biochemical Systematics and Ecology*, 29 (9): 947-957.

Fava, F. & Danese, S. (2011). Intestinal microbiota in inflammatory bowel disease: Friend of foe? *World Journal of Gastroenterology*, 17 (5): 557-566.

Feller, M., Huwiler, K., Stephan, R., Altpeter, E., Shang, A., Furrer, H., Pfyffer, G. E., Jemmi, T., Baumgartner, A. & Egger, M. (2007). Mycobacterium avium subspecies paratuberculosis and Crohn's disease: a systematic review and meta-analysis. *Lancet Infectious Diseases*, 7 (9): 607-613.

Ferrarini, M., Moretto, M., Ward, J. A., Surbanovski, N., Stevanovic, V., Giongo, L., Viola, R., Cavalieri, D., Velasco, R., Cestaro, A., et al. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *Bmc Genomics*, 14.

Fiedler, F. & Schaffler, M. J. (1987). Teichoic-Acids in Cell-Walls of Strains of the Nicotianae Group of Arthrobacter - a Chemotaxonomic Marker. *Systematic and Applied Microbiology*, 9 (1-2): 16-21.

Frank, D. N., Amand, A. L. S., Feldman, R. A., Boedeker, E. C., Harpaz, N. & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (34): 13780-13785.

Garrity, G. M., Bell, J. A. & Lilburn, T. G. (2004). *Taxonomic outline of the prokaryotes. Bergey`s manual og systematic bacteriology.* 2 ed. New York: Springer.

Gevers, D., Kugathasan, S., Denson, L. A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B. Y., Schwager, E., Knights, D., Song, S. J., Yassour, M., et al. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe*, 15 (3): 382-392.

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R. & Mathieu, C. (2001). An overview of real-time quantitative PCR: Applications to quantify cytokine gene expression. *Methods*, 25 (4): 386-401.

Glasser, A. L., Boudeau, G., Barnich, N., Perruchot, M. H., Colombel, J. F. & Darfeuille-Michaud, A. (2001). Adherent invasive Escherichia coli strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death. *Infection and Immunity*, 69 (9): 5529-5537.

Golas, I., Filipkowska, Z., Lewandowska, D. & Zmyslowska, I. (2002). Potentially pathogenic bacteria from the family Enterobacteriaceae, Pseudomonas sp and Aeromonas sp in waters designated for drinking and household purposes. *Polish Journal of Environmental Studies*, 11 (4): 325-330.

Goll, R. & Granlund, A. V. (2015). Intestinal barrier homeostasis in inflammatory bowel disease. *Scandinavian Journal of Gastroenterology*, 50 (1): 3-12.

Goyette-Desjardins, G., Auger, J. P., Xu, J., Segura, M. & Gottschalk, M. (2014). Streptococcus suis, an important pig pathogen and emerging zoonotic agent-an update on the worldwide distribution based on serotyping and sequence typing. *Emerging Microbes & Infections*, 3.

Halfvarson, J., Bodin, L., Tysk, C., Lindeberg, E. & Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin-cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology*, 124 (7).

Hamady, M. & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, 19 (7): 1141-1152.

Hansen, R., Russell, R. K., Reiff, C., Louis, P., McIntosh, F., Berry, S. H., Mukhopadhya, I., Bisset, W. M., Barclay, A. R., Bishop, J., et al. (2012). Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn's But Not in Ulcerative Colitis. *American Journal of Gastroenterology*, 107 (12): 1913-1922.

Hirsch, P. R., Mauchline, T. H. & Clark, I. M. (2010). Culture-independent molecular techniques for soil microbial ecology. *Soil Biology & Biochemistry*, 42 (6): 878-887.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H. Z., Wu, G. D., Lewis, J. D. & Bushman, F. D. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *Plos One*, 8 (6).

Hooper, L. V., Wong, M. H., Thelin, A., Hansson, L., Falk, P. C. & Gordon, J. I. (2001). Molecular analysis of commensal host-microbial relations hips in the intestine. *Science*, 291 (5505): 881-884.

Huys, G., Vanhoutte, T., Joossens, M., Mahious, A. S., De Brandt, E., Vermeire, S. & Swings, J. (2008). Coamplification of eukaryotic DNA with 16S rRNA gene-based PCR primers: Possible consequences for population fingerprinting of complex microbial communities. *Current Microbiology*, 56 (6): 553-557.

Hwang, C., Ling, F. Q., Andersen, G. L., LeChevallier, M. W. & Liu, W. T. (2012). Microbial Community Dynamics of an Urban Drinking Water Distribution System Subjected to Phases of Chloramination and Chlorination Treatments. *Applied and Environmental Microbiology*, 78 (22): 7856-7865.

Hwang, I. T., Kim, Y. J., Kim, S. H., Kwak, C. I., Gu, Y. Y. & Chun, J. Y. (2003). Annealing control primer system for improving specificity of PCR amplification. *Biotechniques*, 35 (6): 1180-1184.

Jimenez, E., Marin, M. L., Martin, R., Odriozola, J. M., Olivares, M., Xaus, J., Fernandez, L. & Rodriguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology*, 159 (3): 187-193.

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, L. P., Sharma, Y., Anderson, C. A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 491 (7422): 119-124.

Kabeerdoss, J., Jayakanthan, P., Pugazhendhi, S. & Ramakrishna, B. S. (2015). Alterations of mucosal microbiota in the colon of patients with inflammatory bowel disease revealed by real time polymerase chain reaction amplification of 16S ribosomal ribonucleic acid. *Indian Journal of Medical Research*, 142: 23-32.

Kalmbach, S., Manz, W., Bendinger, B. & Szewzyk, U. (2000). In situ probing reveals Aquabacterium commune as a widespread and highly abundant bacterial species in drinking water biofilms. *Water Research*, 34 (2): 575-581.

Kennedy, S. & Oswald, N. (2011). *PCR Troubleshooting and Optimization: The Essential guide*. Norfolk: Caister Academic Press.

Khan, S. T., Horiba, Y., Yamamoto, M. & Hiraishi, A. (2002). Members of the family Comamonadaceae as primary poly(3-hydroxybutyrate-co-3-hydroxyvalerate)-degrading denitrifiers in activated sludge as revealed by a polyphasic approach. *Applied and Environmental Microbiology*, 68 (7): 3206-3214.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T. & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 108: 4578-4585.

Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 361 (1475): 1929-1940.

Kruidenier, L., Kulper, I., Lamers, C. B. H. W. & Verspaget, H. W. (2003). Intestinal oxidative damage in inflammatory bowel disease: semi-quantification, localization, and association with mucosal antioxidants. *Journal of Pathology*, 201 (1): 28-36.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogren, B., Strombom, L., et al. (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27: 95-125.

Kuczynski, J., Stombaugh, J., Walters, W. A., Gonzalez, A., Caporaso, J. G. & Knight, R. (2011). Using QIIME to analyze the 16S rRNA gene sequences from Microbial Communities. *Current Protocols in Bioinformatics*, 36:10.7:10.7.1–10.7.20.

Langholz, E., Munkholm, P., Krasilnikoff, P. A. & Binder, V. (1997). Inflammatory bowel diseases with onset in childhood - Clinical features, morbidity, and mortality in a regional cohort. *Scandinavian Journal of Gastroenterology*, 32 (2): 139-147.

Lantz, P. G., Matsson, M., Wadstrom, T. & Radstrom, P. (1997). Removal of PCR inhibitors from human faecal samples through the use of an aqueous two-phase system for sample preparation prior to PCR. *Journal of Microbiological Methods*, 28 (3): 159-167.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D. & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology*, 24 (2): 160-168.

Lennardjones, J. E. (1989). Classification of Inflammatory Bowel-Disease. *Scandinavian Journal of Gastroenterology*, 24: 2-6.

Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Dore, J. & Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut*, 57 (3).

Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. (2006). Microbial ecology - Human gut microbes associated with obesity. *Nature*, 444 (7122): 1022-1023.

Limbergent, V., Russel, J., Drummond, R., Aldhous, H., Round, M., Nimmo, N., Smith, E., Gillett, L., McGrogan, P., Weaver, P., et al. (2008). Definition of Phenotypic Characteristics of Childhood-Onset Inflammatory Bowel Disease. *Gastroenterology*, 135 (4): 1114-1122.

Liu, G., Bakker, G. L., Li, S., Vreeburg, J. H. G., Verberk, J. Q. J. C., Medema, G. J., Liu, W. T. & Van Dijk, J. C. (2014). Pyrosequencing Reveals Bacterial Communities in Unchlorinated Drinking Water Distribution System: An Integral Study of Bulk Water, Suspended Solids, Loose Deposits, and Pipe Wall Biofilm. *Environmental Science & Technology*, 48 (10): 5467-5476.

Liverani, E., Scaioli, E., Cardamone, C., Dal Monte, P. & Belluzzi, A. (2014). Mycobacterium avium subspecies paratuberculosis in the etiology of Crohn's disease, cause or epiphenomenon? *World Journal of Gastroenterology*, 20 (36): 13060-13070.

Lopetuso, L. R., Scaldaferri, F., Petito, V. & Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathogens*, 5.

Loubinoux, J., Bronowicki, J. P., Pereira, I. A. C., Mougenel, J. L. & Le Faou, A. E. (2002). Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *Fems Microbiology Ecology*, 40 (2): 107-112.

Lozupone, C. & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71 (12): 8228-8235.

Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, 73 (5): 1576-1585.

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, 489 (7415): 220-230.

Lu, H. P., Lai, Y. C., Huang, S. W., Chen, H. C., Hsieh, C. H. & Yu, H. T. (2014). Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. *Scientific Reports*, 4.

Lu, P. P., Chen, C., Wang, Q. F., Wang, Z., Zhang, X. J. & Xie, S. G. (2013). Phylogenetic diversity of microbial communities in real drinking water distribution systems. *Biotechnology and Bioprocess Engineering*, 18 (1): 119-124.

Lupton, J. R. (2004). Microbial degradation products influence colon cancer risk: the butyrate controversy. *Journal of Nutrition*, 134 (2): 479-482.

Magic-Knezev, A., Wullings, B. & Van der Kooij, D. (2009). Polaromonas and Hydrogenophaga species are the predominant bacteria cultured from granular activated carbon filters in water treatment. *Journal of Applied Microbiology*, 107 (5): 1457-1467.

Martiny, A. C., Nielsen, A. T., Arvin, E., Molin, S. & Albrechtsen, H. J. (2002). In situ examination of microbial populations in a model drinking water distribution system. *2nd World Water Congress: Water and Health-Microbiology, Monitoring and Disinfection*, 2 (3): 283-288.

Maynard, C. L., Elson, C. O., Hatton, R. D. & Weaver, C. T. (2012). Reciprocal interactions of the intestinal microbiota and immune system. *Nature*, 489 (7415): 231-241.

Mittal, M., Siddiqui, M. R., Tran, K., Reddy, S. P. & Malik, A. B. (2014). Reactive Oxygen Species in Inflammation and Tissue Injury. *Antioxidants & Redox Signaling*, 20 (7): 1126-1167.

Mizrahi-Man, O., Davenport, E. R. & Gilad, Y. (2013). Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *Plos One*, 8 (1).

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K. G., Bonnet, J., Vidal, R., Cabrita, J. & Megraud, F. (1997). Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. *Journal of Clinical Microbiology*, 35 (4): 995-998.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., Reyes, J. A., Shah, A. H., LeLeiko, N., Snapper, S. B., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.*, 13 (9).

Mukhopadhya, I., Hansen, R., El-Omar, E. M. & Hold, G. L. (2012). IBD-what role do Proteobacteria play? *Nature Reviews Gastroenterology & Hepatology*, 9 (4): 219-230.

Mylonaki, M., Rayment, N. B., Rampton, D. S., Hudspith, B. N. & Brostoff, J. (2005). Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. *Inflammatory Bowel Diseases*, 11 (5): 481-487.

Nannipieri, P. & Smalla, K. (2006). Nucleic Acids and Proteins in Soil. In vol. 8 *Nucleic Acids and Proteins in Soli*. Heidelberg: Springer.

Naser, S. A., Sagramsingh, S. R., Naser, A. S. & Thanigachalam, S. (2014). Mycobacterium avium subspecies paratuberculosis causes Crohn's disease in some inflammatory bowel disease patients. *World Journal of Gastroenterology*, 20 (23): 7403-7415.

Navas-Molina, J. A., Peralta-Sanchez, J. M., Gonzalez, A., McMurdie, P. J., Vazquez-Baeza, Y., Xu, Z. J., Ursell, L. K., Lauber, C., Zhou, H. W., Song, S. J., et al. (2013). Advancing Our Understanding of the Human Microbiome Using QIIME. *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*, 531: 371-444.

Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L. & Graf, J. (2014). Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. *Plos One*, 9 (4).

Ng, S. C., Bernstein, C. N., Vatn, M. H., Lakatos, P. L., LoftusJr, E. V., Tysk, C., O'Morain, C., Moum, B., Colombel, J. F. & Force, E. N. H. T. (2013). Geographical variability and environmental risk factors in inflammatory bowel disease. *Gut*, 62 (4): 630-649.

Niemi, R. M., Heiskanen, I., Heine, R. & Rapala, J. (2009). Previously uncultured beta-Proteobacteria dominate in biologically active granular activated carbon (BAC) filters. *Water Research*, 43 (20): 5075-5086.

Ott, S. J., Musfeldt, M., Wenderoth, D. F., Hampe, J., Brant, O., Folsch, U. R., Timmins, K. N. & Schreiber, S. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut*, 53 (5): 685-693.

Ott, S. J., Kuhbacher, T., Musfeldt, M., Rosenstiel, P., Hellmig, S., Rehman, A., Drews, O., Weichert, W., Timmis, K. N. & Schreiber, S. (2008). Fungi and inflammatory bowel diseases: Alterations of composition and diversity. *Scandinavian Journal of Gastroenterology*, 43 (7): 831-841.

Payment, P., Franco, E., Richardson, L. & Siemiatycki, J. (1991). Gastrointestinal Health-Effects Associated with the Consumption of Drinking-Water Produced by Point-of-Use Domestic Reverse-Osmosis Filtration Units. *Applied and Environmental Microbiology*, 57 (4): 945-948.

Payment, P., Coffin, E. & Paquette, G. (1994). Blood Agar to Detect Virulence Factors in Tap Water Heterotrophic Bacteria. *Applied and Environmental Microbiology*, 60 (4): 1179-1183.

Pepper, I. L., Gerba, C. P. & Gentry, T. J. (2015). *Environmental Microbiology*. 3 ed. Waltham: Academic Press.

Perez-Cobas, A. E., Gosalbes, M. J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K., Otto, W., Rojo, D., Bargiela, R., von Bergen, M., et al. (2013). Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut*, 62 (11): 1591-1601.

Perl, D. P., Fogarty, U., Harpaz, N. & Sachar, D. B. (2004). Bacterial-metal interactions: The potential role of aluminum and other trace elements in the etiology of Crohn's disease. *Inflammatory Bowel Diseases*, 10 (6): 881-883.

Pigneur, B., Seksik, P., Viola, S., Viala, J., Beaugerie, L., Girardet, J. P., Ruemmele, F. M. & Cosnes, J. (2010). Natural History of Crohn's Disease: Comparison Between Childhood- and Adult-Onset Disease. *Inflammatory Bowel Diseases*, 16 (6): 953-961.

Pinto, A. J., Xi, C. W. & Raskin, L. (2012). Bacterial Community Structure in the Drinking Water Microbiome Is Governed by Filtration Processes. *Environmental Science & Technology*, 46 (16): 8851-8859.

Qin, J. J., Li, R. Q., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464 (7285): 59-U70.

Rajendhran, J. & Gunasekaran, P. (2011). Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research*, 166 (2): 99-110.

Rajilic-Stojanovic, M., Shanahan, F., Guarner, F. & de Vos, W. M. (2013). Phylogenetic Analysis of Dysbiosis in Ulcerative Colitis During Remission. *Inflammatory Bowel Diseases*, 19 (3): 481-488.

Raqib, R., Mia, S. M. S., Qadri, F., Alam, T. I., Alam, N. H., Chowdhury, A. K., Mathan, M. M. & Andersson, J. (2000). Innate immune responses in children and adults with shigellosis. *Infection and Immunity*, 68 (6): 3620-3629.

Revetta, R. P., Pemberton, A., Lamendella, R., Iker, B. & Domingo, J. W. S. (2010). Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. *Water Research*, 44 (5): 1353-1360.

Rideout, J. R., He, Y., Navas-Molina, J. A., Walters, W. A., Ursell, L. K., Gibbons, S. M., Chase, J., McDonald, D., Gonzalez, A., Robbins-Pianka, A., et al. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *Peerj*, 2.

Rigottier-Gois, L. (2013). Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. *Isme Journal*, 7 (7): 1256-1261.

Rizzo, J. M. & Buck, M. J. (2012). Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing. *Cancer Prevention Research*, 5 (7): 887-900.

Rodriguez, J. M. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis*, 26 (10).

Rolhion, N. & Darfeuille-Michaud, A. (2007). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Inflammatory Bowel Diseases*, 13 (10): 1277-1283.

Romano, I., Bellitti, M. R., Nicolaus, B., Lama, L., Manca, M. C., Pagnotta, E. & Gambacorta, A. (2000). Lipid profile: a useful chemotaxonomic marker for classification of a new cyanobacterium in Spirulina genus. *Phytochemistry*, 54 (3): 289-294.

Rowan, F., Docherty, N. G., Murphy, M., Murphy, B., Coffey, J. C. & O'Connell, P. R. (2010). Desulfovibrio Bacterial Species Are Increased in Ulcerative Colitis. *Diseases of the Colon & Rectum*, 53 (11): 1530-1536.

Rudi, K., Tannaes, T. & Vatn, M. (2009). Temporal and Spatial Diversity of the Tap Water Microbiota in a Norwegian Hospital. *Applied and Environmental Microbiology*, 75 (24): 7855-7857.

Rudi, K., Berg, F., Gaustad, E., Tannes, T. & Vatn, M. (2010). Ratios between Alpha-, Beta- and Gamma-proteobacteria in tap water determined by the ProteoQuant assay. *Letters in Applied Microbiology*, 50 (1): 1-6.

Salonen, A., Nikkila, J., Jalanka-Tuovinen, J., Immonen, O., Rajilic-Stojanovic, M., Kekkonen, R. A., Palva, A. & de Vos, W. M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *Journal of Microbiological Methods*, 81 (2): 127-134.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74 (12): 5463-5467.

Sartor, R. B. (2006). Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nature Clinical Practice Gastroenterology & Hepatology*, 3 (7): 390-407.

Sartor, R. B. & Mazmanian, S. K. (2012). Intestinal Microbes in Inflammatory Bowel Diseases. *The American journal of Gastroenterology Supplements*, 1: 15-21.

Sartor, R. B. (2015). Optimal sampling of the intestinal microbiota for research. *Nature Reviews Gastroenterology & Hepatology*, 12 (5): 253-254.

Satokari, R. (2015). Contentious host-microbiota relationship in inflammatory bowel disease - can foes become friends again? *Scandinavian Journal of Gastroenterology*, 50 (1): 34-42.

Satsangi, J., Silverberg, M. S., Vermeire, S. & Colombel, J. F. (2006). The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut*, 55 (6): 749-753.

Schwartz, T., Kalmbach, S., Hoffmann, S., Szewzyk, U. & Obst, U. (1998). PCR-based detection of mycobacteria in biofilms from a drinking water distribution system. *Journal of Microbiological Methods*, 34 (2): 113-123.

Sekelja, M., Berget, I., Naes, T. & Rudi, K. (2011). Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *Isme Journal*, 5 (3): 519-531.

Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. (2010). Gut Microbiota in Health and Disease. *Physiological Reviews*, 90 (3): 859-904.

Sepehri, S., Kotlowski, R., Bernstein, C. N. & Krouse, D. O. (2007). Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. *Inflammatory Bowel Diseases*, 13 (6): 675-683.

Sheehan, D., Moran, C. & Shanahan, F. (2015). The microbiota in inflammatory bowel disease. *Journal of Gastroenterology*, 50 (5): 495-507.

Sommer, F. & Backhed, F. (2013). The gut microbiota - masters of host development and physiology. *Nature Reviews Microbiology*, 11 (4): 227-238.

Stackebrandt, E. & Goebel, B. M. (1994). A Place for DNA-DNA Reassociation and 16s Ribosomal-Rna Sequence-Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology*, 44 (4): 846-849.

Stoner, K. A., Rabe, L. K., Austin, M. N., Meyn, L. A. & Hillier, S. L. (2011). Incidence and Epidemiology of Streptococcus pseudoporcinus in the Genital Tract. *Journal of Clinical Microbiology*, 49 (3): 883-886.

Strauss, J., Kaplan, G. G., Beck, P. L., Rioux, K., Panaccione, R., DeVinney, R., Lynch, T. & Allen-Vercoe, E. (2011). Invasive Potential of Gut Mucosa-derived Fusobacterium nucleatum Positively Correlates with IBD Status of the Host. *Inflammatory Bowel Diseases*, 17 (9): 1971-1978.

Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D. & Dore, J. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Applied and Environmental Microbiology*, 65 (11): 4799-4807.

Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P. & Lochs, H. (2005). Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *Journal of Clinical Microbiology*, 43 (7): 3380-3389.

Swidsinski, A., Loening-Baucke, V. & Herber, A. (2009). Mucosal Flora in Crohn's Disease and Ulcerative Colitis - an Overview. *Journal of Physiology and Pharmacology*, 60: 61-71.

Szewzyk, U., Szewzyk, R., Manz, W. & Schleifer, K. H. (2000). Microbiological safety of drinking water. *Annual Review of Microbiology*, 54: 81-127.

Tabaqchali, S., Odonoghue, D. P. & Bettelheim, K. A. (1978). Escherichia-Coli Antibodies in Patients with Inflammatory Bowel-Disease. *Gut*, 19 (2): 108-113.

Tedelind, S., Westberg, F., Kjerrulf, M. & Vidal, A. (2007). Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: A study with relevance to inflammatory bowel disease. *World Journal of Gastroenterology*, 13 (20): 2826-2832.

Thibault, R., Blachier, F., Darcy-Vrillon, B., de Coppet, P., Bourreille, A. & Segain, J. P. (2010). Butyrate Utilization by the Colonic Mucosa in Inflammatory Bowel Diseases: A Transport Deficiency. *Inflammatory Bowel Diseases*, 16 (4): 684-695.

Tortora, G. J., Funke, B. R. & Case, C. L. (2010). *Microbiology, An introduction*. 10 ed. San Fransisco: Pearson Benjamin Cummings.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., et al. (2009). A core gut microbiome in obese and lean twins. *Nature*, 457 (7228): 480-U7.

Vaz-Moreira, I., Egas, C., Nunes, O. C. & Manaia, C. M. (2013). Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. *Fems Microbiology Ecology*, 83 (2): 361-374.

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G. & Petrovska, L. (2011). High-throughput clone library analysis of the mucosa-

associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *Bmc Microbiology*, 11.

Willey, M. J., Sherwood, L. M. & Wolverton, C. J. (2009). *Prescott`s Principles of Microbiology*. International edition ed. New York: Mc-Graw Hill education.

Williams, M. M., Domingo, J. W. S., Meckes, M. C., Kelty, C. A. & Rochon, H. S. (2004). Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *Journal of Applied Microbiology*, 96 (5): 954-964.

Willing, B. P., Dicksved, J., Halfvarson, J., Andersson, A. F., Lucio, M., Zheng, Z., Jarnerot, G., Tysk, C., Jansson, J. K. & Engstrand, L. (2010). A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes. *Gastroenterology*, 139 (6): 1844-U105.

Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., et al. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, 467 (7314): 426-429.

Winter, S. E., Winter, M. G., Xavier, M. N., Thiennimitr, P., Poon, V., Keestra, A. M., Laughlin, R. C., Gomez, G., Wu, J., Lawhon, S. D., et al. (2013). Host-Derived Nitrate Boosts Growth of E. coli in the Inflamed Gut. *Science*, 339 (6120): 708-711.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., et al. (2011). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*, 334 (6052): 105-108.

Wu, H. T., Zhang, J. X., Mi, Z. L., Xie, S. G., Chen, C. & Zhang, X. J. (2015). Biofilm bacterial communities in urban drinking water distribution systems transporting waters with different purification strategies. *Applied Microbiology and Biotechnology*, 99 (4): 1947-1955.

Xavier, R. J. & Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, 448 (7152): 427-434.

Xiong, J. (2006). *Essential Bioinformatics*. New York: Cambridge University Press.

Zhang, J., Chiodini, R., Badr, A. & Zhang, G. F. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38 (3): 95-109.

Zoetendal, E. G., Ben-Amor, K., Akkermans, A. D. L., Abee, T. & de Vos, W. M. (2001). DNA isolation Protocols affect the detection limit of PCR approaches of bacteria in samples from the human gastrointestinal tract. *Systematic and Applied Microbiology*, 24 (3): 405-410.

Zoetendal, E. G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. L. & de Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Applied and Environmental Microbiology*, 68 (7): 3401-3407.

Online references

454.com. *454 sequencing*, *The technology*. [ONLINE] Available at http://454.com/products/technology.asp [Accessed 31th Oct 2015]

Applied Biosystems. 2008. SOLiD™ System barcoding. [ONLINE] Available at http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html [Accessed 31th Oct 2015]

Illumina, Inc. 2015. *An Introduction to Next-Generation Sequencing Technology*. [ONLINE] Available at: http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf [Accessed 28th Aug 2015]

Illumina, Inc. 2015. *Diagnosing and Preventing Flow Cell Overclustering on the MiSeq® System*. [ONLINE] Available at https://support.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf [Accessed 20th Nov 2015]

Illumina, Inc. 2011. *Quality Scores for Next-Generation Sequencing*. [ONLINE] Available at: http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf [Accessed 20th Nov 2015]

Ilsi.org. 2011. *The Enterobacteriaceae and their significance to the food industry*. [ONLINE] Available at: http://www.ilsi.org/Europe/Documents/EP%20Enterobacteriaceae.pdf [Accessed 11th April 2016]

Lovdata.no. 2001. *Forskrift om vannforsyning of drikkevann (Drikkevannforskriften).* [ONLINE] Available at: https://lovdata.no/dokument/SF/forskrift/2001-12-04-1372 [Accessed 29th March 2016]

Norskvann.no. 2013. *Vannkilder*. [ONLINE] Available at http://www.norskvann.no/index.php/vann/vannkilde [Accessed 18thApril 2016]

Researchgate.net. 2012. *Figure cluster generation and bridge amplification by Illumina*. [ONLINE] Available at: https://www.researchgate.net/figure/259468015_fig6_Figure-6-Cluster-generation-by-bridge-amplification-of-Illumina-a-Randomly-fragmented [Accessed 25th Jan 2016]

Thermo Fisher Scientific Inc. *Introduction to Ion Torrent™ Next-Generation Sequencing Technology*. [ONLINE] Available at https://www.thermofisher.com/no/en/home/life-science/sequencing/sequencing-education.html# [Accessed 31th Oct 2015]

QIIME. 454 Overview Tutorial: denovo OTU picking and diversity analyses using 454 data [ONLINE] Available at: http://qiime.org/tutorials/tutorial.html [Accessed 7th Dec 2015]

WHO.int. 2008. *Guidelines for Drinking-water Quality.* [ONLINE] Available at: http://apps.who.int/iris/bitstream/10665/204411/1/9789241547611_eng.pdf?ua=1 [Accessed 11thApril 2016]

WHO.int. 2003. *Relationship between common water bacteria and pathogens in drinking-water*. [ONLINE] Available at: http://www.who.int/water_sanitation_health/dwq/HPC6.pdf [Accessed 25th April 2016

Appendix A – Overview of samples and patients

Table 1: The table shows all samples analyzed for this research project, their respective ID number and origin with respect to patient number, diagnosis and age group. For biopsy samples, category with respect to inflammation category and location for retrieval is given. All samples are sorted with respect to patient number.

Patient number[†]	Sample ID water	Sample ID biopsy	Biopsi category[‡]	Diagnosis	Age group
8	162	301	AI	Non	Child
8	162	302	AII	Non	Child
32	44	303	AIII	Non	Child
32	44	304	AVI	Non	Child
35	66	305	AIII	Non	Child
35	66	306	AVI	Non	Child
36	81	307	AIII	CD	Adult
36	81	308	AVI	CD	Adult
K0036	-	105	AIII	CD	Adult
K0036	-	106	AVI	CD	Adult
37	82 + 182	309	II	CD	Adult
37	82 + 182	310	AVI	CD	Adult
37	82 + 182	311	BVII	CD	Adult
38	87	312	II	CD	Child
38	87	313	AV	CD	Child
38	87	314	BVII	CD	Child
K0038	-	107	I	CD	Adult
K0038	-	108	AV	CD	Adult
K0038	-	109	AVII	CD	Adult
39	45	315	II	CD	Adult
39	45	316	BV	CD	Adult
39	45	317	BVII	CD	Adult
K0039	-	110	I	CD	Adult
K0039	-	111	II	CD	Adult
K0039	-	112	AV	CD	Adult
K0039	-	113	AVII	CD	Adult
40	78	318	AVI	Non	Adult
41	7	319	AIII	Non	Adult
41	7	320	AVI	Non	Adult
42	16	321	AIII	Non	Adult
42	16	322	AVI	Non	Adult
K0042	-	141	I	Non	Adult
K0042	-	142	AIII	Non	Adult
K0042	-	143	AVI	Non	Adult
43	6	323	AIII	Non	Adult
43	6	324	AVI	Non	Adult
44	123	325	II	UC	Adult
44	123	326	BV	UC	Adult
44	123	327	BVII	UC	Adult
K0044	-	118	I	UC	Adult
K0044	-	119	II	UC	Adult
K0044	-	120	BVI	UC	Adult
K0044	-	121	BVII	UC	Adult
46	4	328	AIII	CD	Adult
46	4	329	BI	CD	Adult
K0046	-	122	VII	CD	Adult
K0046	-	123	AIII	CD	Adult
K0046	-	124	BI	CD	Adult

47	80	330	AVI	Non	Child
47	80	331	BVI	Non	Child
48	49	332	AI	Possible	Adult
48	49	333	AVI	Possible	Adult
48	49	334	BVII	Possible	Adult
49	132	335	VII	CD	Adult
49	132	336	AIII	CD	Adult
K0049	-	138	VII	CD	Adult
K0049	-	139	AIII	CD	Adult
K0049	-	140	BI	CD	Adult
51	52	337	I	CD	Adult
51	52	338	AIII	CD	Adult
55	43	340	AI	Non	Adult
55	43	341	AIII	Non	Adult
55	43	342	AVI	Non	Adult
58	67	343	AIII	Non	Child
58	67	344	AVI	Non	Child
K0058	-	115	I	Non	Adult
K0058	-	116	AIII	Non	Adult
K0058	-	117	AVI	Non	Adult
59	46	345	AI	UC	Adult
59	46	346	AIII	UC	Adult
59	46	347	BIV	UC	Adult
59	46	348	BVII	UC	Adult
K0059	-	98	I	UC	Adult
K0059	-	99	AIII	UC	Adult
K0059	-	100	BVII	UC	Adult
60	38	349	AI	Non	Adult
60	38	350	AIII	Non	Adult
60	38	351	AVI	Non	Adult
61	17	352	BVI	UC	Adult

62	35	353	AIII	Non	Adult
62	35	354	AVII	Non	Adult
63	71	355	AIII	Non	Adult
63	71	356	AVI	Non	Adult
64	100	357	I	IBDU	Child
64	100	358	AIII	IBDU	Child
64	100	359	AVI	IBDU	Child
K0064	-	125	I	IBDU	Adult
K0064	-	126	AIII	IBDU	Adult
K0064	-	127	AVI	IBDU	Adult
65	97	408	AVI	Possible	Adult
65	97	409	BVII	Possible	Adult
66	109	410	AIII	CD	Adult
66	109	411	BI	CD	Adult
67	69	412	AIII	Non	Adult
67	69	413	AVI	Non	Adult
69	40	415	AVI	Possible	Adult
70	27	416	AIII	Non	Adult
70	27	417	BI	Non	Adult
71	91	424	BIII	CD	Child
71	91	425	BVI	CD	Child
K0071	-	171	I	CD	Adult
K0071	-	172	AIII	CD	Adult
K0071	-	173	AVI	CD	Adult
72	119	360	AV	Non	Child
73	103	300	AIII	CD	Adult
73	103	339	BI	CD	Adult
74	39	418	I	Non	Adult
74	39	419	AIII	Non	Adult
74	39	420	AVI	Non	Adult
75	68	421	I	Possible	Adult

75	68	422	AVI	Possible	Adult
75	68	423	BVII	Possible	Adult
76	42	214	I	Non	Adult
76	42	215	AV	Non	Adult
76	42	216	BVII	Non	Adult
77	175	217	AIII	Non	Child
77	175	218	AVI	Non	Child
K0077	-	254	I	Non	Adult
K0077	-	255	AIII	Non	Adult
K0077	-	256	AVI	Non	Adult
78	131	219	I	Possible	Adult
78	131	220	AIII	Possible	Adult
78	131	221	AVI	Possible	Adult
79	20	222	AVI	UC	Child
79	20	223	BVII	UC	Child
K0079	-	101	I	UC	Adult
K0079	-	103	AVI	UC	Adult
K0079	-	104	BVII	UC	Adult
K0079	-	114	II	UC	Adult
81	223 + 33	224	II	CD	Adult
81	223 + 33	225	AVI	CD	Adult
81	223 + 33	226	BVI	CD	Adult
K0081	-	147	I	CD	Adult
K0081	-	148	AIII	CD	Adult
K0081	-	149	BVI	CD	Adult
83	189	227	II	UC	Adult
83	189	228	AV	UC	Adult
83	189	229	BVII	UC	Adult
K0083	-	150	AVI	UC	Adult
K0083	-	151	AVII	UC	Adult
85	193	230	I	Non	Adult
85	193	231	AIII	Non	Adult
85	193	232	AVI	Non	Adult
86	147	233	II	UC	Child
86	147	234	BV	UC	Child
86	147	235	BVII	UC	Child
K0086	-	152	BV	UC	Adult
K0086	-	153	BVII	UC	Adult
87	211	236	AIII	Non	Child
87	211	237	AV	Non	Child
88	190	238	AV	Non	Child
90	140	239	I	Non	Adult
90	140	240	AIII	Non	Adult
90	140	241	AVI	Non	Adult
93	194	242	II	UC	Adult
93	194	243	BV	UC	Adult
93	194	244	BVII	UC	Adult
K0093	-	167	I	UC	Adult
K0093	-	168	II	UC	Adult
K0093	-	169	BV	UC	Adult
K0093	-	170	BVII	UC	Adult
95	205 + 222	245	AVII	UC	Adult
95	205 + 222	246	BV	UC	Adult
97	217	247	AIII	Non	Adult
97	217	248	AVI	Non	Adult
K0097	-	144	II	Non	Adult
K0097	-	145	BI	Non	Adult

K0097	-	146	BVII	Non	Adult
98	149	249	AIII	Non	Adult
98	149	250	AVI	Non	Adult
101	155	251	VI	Possible	Adult
101	155	252	AIII	Possible	Adult
101	155	253	BI	Possible	Adult
102	208	426	II	Non	Adult
102	208	427	AVI	Non	Adult
102	208	428	BVII	Non	Adult
103	138	362	AIII	Non	Adult
103	138	365	AVI	Non	Adult
104	135	366	I	CD	Child
104	135	367	BIII	CD	Child
104	135	368	BVII	CD	Child
K0104	-	174	I	CD	Adult
K0104	-	175	AIII	CD	Adult
K0104	-	176	AV	CD	Adult
105	152	369	II	UC	Adult
105	152	370	AVI	UC	Adult
105	152	371	BVII	UC	Adult
K0105	-	177	AV	UC	Adult
K0105	-	178	AVII	UC	Adult
106	196	372	VII	Non	Adult
106	196	373	AIII	Non	Adult
106	196	374	BI	Non	Adult
107	154	375	I	CD	Adult
107	154	376	AIII	CD	Adult
107	154	377	BI	CD	Adult
107	154	378	BV	CD	Adult
108	133	379	AIII	Non	Adult
108	133	380	AVI	Non	Adult

109	191	381	BVI	UC	Adult
K0109	-	179	I	UC	Adult
K0109	-	180	AIII	UC	Adult
K0109	-	93	AVI	UC	Adult
110	188	382	VII	Non	Adult
110	188	383	BI	Non	Adult
110	188	414	AIII	Non	Adult
112	179	384	AI	CD	Adult
112	179	385	BVI	CD	Adult
112	179	386	BVII	CD	Adult
K0112	-	262	I	CD	Adult
K0112	-	263	II	CD	Adult
K0112	-	264	AVI	CD	Adult
K0112	-	265	AVII	CD	Adult
113	160	387	I	Non	Adult
113	160	388	AIII	Non	Adult
113	160	389	AVI	Non	Adult
116	150	390	II	UC	Adult
116	150	391	BV	UC	Adult
116	150	392	BVII	UC	Adult
116	150	393	BI	UC	Adult
K0116	-	266	II	UC	Adult
K0116	-	267	BV	UC	Adult
K0116	-	268	BVII	UC	Adult
117	184	394	I	Possible	Adult
117	184	395	AIII	Possible	Adult
117	184	396	AVI	Possible	Adult
120	207	200	III	CD	Adult
120	207	201	V	CD	Adult
K0120	-	269	VII	CD	Adult
K0120	-	270	AV	CD	Adult

122	192	397	AIII	Non	Adult
122	192	398	AVI	Non	Adult
123	-	399	I	Non	Adult
123	-	400	AV	Non	Adult
123	-	401	BVII	Non	Adult
125	163	402	BV	UC	Adult
125	163	403	BVII	UC	Adult
K0125	-	271	I	UC	Adult
K0125	-	272	II	UC	Adult
K0125	-	273	AV	UC	Adult
K0125	-	274	AVII	UC	Adult
132	214	211	II	UC	Adult
132	214	212	BV	UC	Adult
132	214	213	BVII	UC	Adult
K0132	-	259	II	UC	Adult
K0132	-	260	AV	UC	Adult
K0132	-	261	AVII	UC	Adult
1009	58	294	II	UC	Adult
1009	58	295	AV	UC	Adult
1009	58	296	BVII	UC	Adult
1010	61	291	VI	CD	Adult
1010	61	292	AIII	CD	Adult
1010	61	293	BIII	CD	Adult
1011	22	288	I	Non	Adult
1011	22	289	AIII	Non	Adult
1011	22	290	AVI	Non	Adult
1012	115	208	II	UC	Adult
1012	115	209	AVI	UC	Adult
1012	115	210	BVI	UC	Adult

1013	112	202	II	UC	Child
1013	112	203	BV	UC	Child
1013	112	204	BVII	UC	Child
1014	31	361	I	Non	Adult
1014	31	363	AVI	Non	Adult
1014	31	364	AIII	Non	Adult
1015	62	184	I	Non	Adult
1015	62	185	AV	Non	Adult
1015	62	186	BVII	Non	Adult
1016	74	190	AIII	CD	Adult
1016	74	191	BIII	CD	Adult
1017	124	192	AVII	UC	Adult
1017	124	193	BVII	UC	Adult
1018	12	187	I	Non	Adult
1018	12	188	AIII	Non	Adult
1018	12	189	AVI	Non	Adult
1019	37	183	I	Non	Adult
1020	37	136	AIII	Non	Adult
1020	37	137	AVI	Non	Adult
1021	199	297	II	UC	Adult
1021	199	298	BV	UC	Adult
1021	199	299	BVII	UC	Adult
2004	101	284	I	UC	Adult
2004	101	285	II	UC	Adult
2004	101	286	AV	UC	Adult
2004	101	287	BVII	UC	Adult
2005	77	128	I	Non	Adult
2005	77	129	AIII	Non	Adult
2005	77	130	AVI	Non	Adult

2006	53	131	I	Non	Adult
2006	53	132	AIII	Non	Adult
2006	53	133	AVI	Non	Adult
2007	89	275	I	Non	Adult
2007	89	276	AIII	Non	Adult
2007	89	277	AVI	Non	Adult
2009	104	278	AVI	Non	Adult
2009	104	279	I	Non	Adult
2009	104	280	AIII	Non	Adult
2010	85	205	I	Non	Adult
2010	85	206	AIII	Non	Adult
2010	85	207	AVI	Non	Adult
2011	41	3	I	Non	Adult
2011	41	4	AIII	Non	Adult
2011	41	5	AVI	Non	Adult
2012	79	6	I	UC	Adult
2012	79	7	II	UC	Adult
2012	79	8	AV	UC	Adult
2012	79	9	BVII	UC	Adult
2013	5	10	I	Non	Adult
2013	5	11	AIII	Non	Adult
2013	5	13	AVI	Non	Adult
2014	99	21	AIII	Non?	Adult
2014	99	22	BI	Non?	Adult
2014	99	23	BVII	Non?	Adult
2015	125	24	III	CD	Adult
2015	125	25	?VI	CD	Adult
2016	128	26	I	UC	Adult
2016	128	27	II	UC	Adult
2017	75	28	AIII	?	Adult
2017	75	29	BI	?	Adult

2020	204	49	II	UC	Adult
2020	204	50	AV	UC	Adult
2020	204	51	BVII	UC	Adult
2021	170	52	I	UC	Adult
2021	170	53	II	UC	Adult
2021	170	54	BV	UC	Adult
2021	174	55	BVII	UC	Adult
2022	174	56	II	UC	Adult
2022	174	57	BV	UC	Adult
2022	174	58	BVII	UC	Adult
2023	173	59	AV	UC	Adult
2023	173	60	BVII	UC	Adult
2024	144	76	VII	CD	Child
2024	144	77	AIII	CD	Child
2024	144	78	BVII	CD	Child
5001	114 og 25	404	I	IBDU	Adult
5001	114 og 25	405	II	IBDU	Adult
5001	114 og 25	406	AV	IBDU	Adult
5001	114 og 25	407	BVII	IBDU	Adult
5004	108	194	I	Non	Adult
5004	108	195	AIII	Non	Adult
5004	108	196	AVI	Non	Adult
5005	57	197	I	Non	Adult
5005	57	198	AIII	Non	Adult
5005	57	199	AVI	Non	Adult
5007	60	14	VII	Possible	Adult
5007	60	15	AIII	Possible	Adult
5007	60	16	BI	Possible	Adult
5008	117	17	I	IBDU	Adult
5008	117	18	VII	IBDU	Adult
5008	117	19	AIII	IBDU	Adult

5009	95	20	AVI	?	Adult
5009	95	91	AIII	?	Adult
5010	218	30	II	Non	Adult
5010	218	31	AVI	Non	Adult
5011	148	61	AIII	Non	Adult
5011	148	62	AVI	Non	Adult
5013	177	79	I	Non	Adult
5013	177	80	AIII	Non	Adult
5013	177	81	AVI	Non	Adult
6001	151	281	I	UC	Child
6001	151	282	BIII	UC	Child
6001	151	283	BVI	UC	Child
K6001	-	88	I	UC	Adult
K6001	-	89	II	UC	Adult
K6001	-	90	BV	UC	Adult
K6001	-	92	BVII	UC	Adult
6002	176	134	BIII	CD	Child
6002	176	135	BV	CD	Child
K6002	-	94	BII	CD	Adult
6003	187	1	AIII	CD	Child
6003	187	2	AVI	CD	Child
6005	18	32	I	UC	Child
6005	18	33	II	UC	Child
6005	18	34	BV	UC	Child
6005	18	35	BVII	UC	Child
6006	64	36	VII	Non?	Child
6006	64	37	AIII	Non?	Child
6006	64	38	BI	Non?	Child
6007	14	39	I	UC	Child
6007	14	40	AIII	UC	Child
6007	14	41	BVII	UC	Child
K6007	-	95	II	UC	Adult
K6007	-	96	AIII	UC	Adult
K6007	-	97	BVII	UC	Adult
6008	102	42	VI	Non	Child
6008	102	43	AIII	Non	Child
6009	51	44	AIII	Non	Child
6009	51	45	AVI	Non	Child
6010	90	46	AVII	CD	Child
K6010	-	154	I	CD	Adult
K6010	-	155	II	CD	Adult
K6010	-	156	AV	CD	Adult
K6010	-	157	BVI	CD	Adult
6011	19	47	AII	UC	Child
6011	19	48	AVI	UC	Child
6013	28	181	BIII	CD	Child
6013	28	182	BV	CD	Child
K6013	-	257	AIII	CD	Adult
K6013	-	258	AV	CD	Adult
6014	213	63	I	Non	Child
6014	213	64	AII	Non	Child
6014	213	65	AV	Non	Child
6015	136	66	AII	Non	Child
6015	136	67	AV	Non	Child
6016	216	68	AII	Non	Child
6017	227	69	I	UC	Child
6017	227	70	II	UC	Child
6017	227	71	BV	UC	Child

6017	227	72	BVII	UC	Child
K6017	-	158	I	UC	Adult
K6017	-	159	?VI	UC	Adult
K6017	-	160	?VII	UC	Adult
6018	224	73	AVI	CD	Child
K6018	-	161	VI	CD	Adult
K6018	-	162	AIII	CD	Adult
6019	228	74	VII	CD	Child
6019	228	75	AII	CD	Child
K6019	-	163	I	CD	Adult
K6019	-	164	VII	CD	Adult
K6019	-	165	AII	CD	Adult
K6019	-	166	BV	CD	Adult
6020	180	82	BVII	CD	Child
6022	226	83	AIII	CD	Child
6022	226	84	AV	CD	Child
6023	178	85	I	UC	Child
6023	178	86	II	UC	Child
6023	178	87	AIV	UC	Child
25	116	-	-	-	-
52	36	-	-	UC	Adult
1022	229	-	-	UC	Adult
1023	165	-	-	Non	Adult
2018	59	-	-	IBDU	Adult
5002	83	-	-	IBDU	Adult
9006	56	-	-	-	-
AIFO 30.11.45	181	-	-	-	-
ESG 210453	166	-	-	-	-
SV 120168	23	-	-	-	-

4001	9	-	-	CD	Adult
4008	24	-	-	UC	Adult
4015	126	-	-	UC	Adult
4016	107	-	-	Non	Adult
4017	47	-	-	UC	Adult
4021	8	-	-	UC	Adult
4034	88	-	-	UC	Adult
4038	105	-	-	Non	Adult
4053	48	-	-	UC	Adult
4003	113	-	-	-	Adult
4004	29	-	-	UC	Adult
4005	111	-	-	UC	Adult
4007	70	-	-	UC	Adult
4009	72	-	-	UC	Adult
4012	63	-	-	UC	Adult
4013	92	-	-	UC	Adult
4018	34	-	-	UC	Adult
4019	15	-	-	UC	Adult
4020	130	-	-	UC	Adult
4022	10	-	-	UC	Adult
4023	118	-	-	UC	Adult
4024	110	-	-	UC	Adult
4025	121	-	-	UC	Adult
4026	13	-	-	Non	Adult
4027	50	-	-	UC	Adult
4028	120	-	-	UC	Adult
4029	127	-	-	UC	Adult
4030	32	-	-	UC	Adult
4031	96	-	-	UC	Adult
4032	93	-	-	UC	Adult
4033	76	-	-	UC	Adult

4036	94	-	-	UC	Adult
4037	129	-	-	Non	Adult
4039	65	-	-	IBDU	Adult
4040	55	-	-	UC	Adult
4041	106	-	-	IBDU	Adult
4042	54	-	-	UC	Adult
4043	98	-	-	-	Adult
4044	21	-	-	UC	Adult
4045	73	-	-	Non	Adult
4047	26	-	-	Non	Adult
4048	11	-	-	UC	Adult
4049	122	-	-	UC	Adult
4051	30	-	-	UC	Adult
4054	86	-	-	UC	Adult
4055	156	-	-	UC	Adult
4056	137	-	-	UC	Adult
4057	198	-	-	UC	Adult
4058	219	-	-	UC	Adult
4059	220	-	-	UC	Adult
4060	195	-	-	IBDU	Adult
4061	209	-	-	UC	Adult
4062	202	-	-	UC	Adult
4063	146	-	-	UC	Adult
4064	145	-	-	CD	Adult
4065	200	-	-	UC	Adult
4066	167	-	-	UC	Adult
4067	141	-	-	IBDU	Adult
4068	225	-	-	UC	Adult
4069	164	-	-	UC	Adult
4070	159	-	-	UC	Adult
4072	153	-	-	CD	Adult
4073	215	-	-	UC	Adult
4074	157	-	-	UC	Adult
4075	171	-	-	UC	Adult
4076	183	-	-	UC	Adult
4078	186	-	-	UC	Adult
4079	212	-	-	UC	Adult
4080	185	-	-	UC	Adult
4081	169	-	-	UC	Adult
4082	158	-	-	UC	Adult
4083	203	-	-	UC	Adult
4084	168	-	-	Non	Adult
4085	210	-	-	UC	Adult
4086	139	-	-	Non	Adult
4089	172	-	-	UC	Adult
4091	143	-	-	UC	Adult
4092	134	-	-	UC	Adult
4093	197	-	-	UC	Adult
4094	206	-	-	Non	Adult
4095	161	-	-	UC	Adult
4097	230	-	-	Non	Adult
4098	142	-	-	UC	Adult
4100	231	-	-	UC	Adult
4103	201	-	-	UC	Adult

† Patients marked with K are from follow-up study

‡ A =Non inflamed, B = inflamed, Neither A or B = Inflammation category not known. Ileum=I, Caecum=II, Ascendens=III, Transversum= IV, Descendens=V, Sigmoideum=VI, Colon=VII

Appendix B – Results of alpha diversity estimates in QIIME

Table 2: The table shows the result of alpha diversity analysis implemented in QIIME. Owing to the magnitude of output produced by this analysis, only a few selected groups from each data set is chosen, and the diversity is only given at 3000 sequences. All metrics employed for analysis is presented, including their respective errors.

Alpha diversity analysis at 3000 sequences

Biopsies	PD whole tree	PD whole tree Error.	Chao1	Chao1 Error.	Observed species	Observed species Error.	Shannon	Shannon Error.	Simpson	Simpson Error.
Age										
Adult	11.210	3.201	139.298	45.511	107.328	35.927	4.591	0.769	0.902	0.071
Child	10.581	2.318	130.340	36.004	100.088	27.659	4.446	0.709	0.894	0.063
Inflammation										
Inflamed	10.562	2.785	127.639	41.106	98.576	33.168	4.473	0.789	0.899	0.068
Non-inflamed	11.396	3.132	142.513	43.811	110.029	34.414	4.654	0.734	0.906	0.069
Diagnosis										
CD	10.718	2.903	129.609	37.732	100.392	30.968	4.508	0.854	0.896	0.095
UC	11.182	3.554	139.249	54.580	107.619	42.496	4.558	0.848	0.897	0.071
Non	11.260	2.774	140.505	38.399	107.636	30.521	4.580	0.611	0.904	0.048
IBDU	10.301	2.509	133.669	36.214	103.467	25.184	4.365	0.666	0.886	0.045
Possible	11.457	2.421	145.947	34.249	111.389	24.841	4.770	0.620	0.920	0.039
Gut part										
Large intestine	11.092	3.062	137.721	43.857	105.978	34.546	4.585	0.749	0.903	0.068
Small intestine	10.891	2.671	134.718	37.699	103.407	29.419	4.477	0.742	0.893	0.072
Cecum	11.295	3.532	140.013	52.646	108.696	41.430	4.547	0.830	0.897	0.071
Water samples										
Age-diagnosis										
A_CD	22.271	5.360	316.664	92.106	207.290	61.147	4.972	0.960	0.895	0.083
C_CD	23.642	6.348	356.311	100.707	221.627	68.403	4.933	0.862	0.900	0.063
A_UC	25.785	7.394	367.517	110.381	240.562	79.970	5.116	1.206	0.884	0.121
C_UC	26.440	3.742	358.445	59.730	249.227	41.781	5.266	1.130	0.886	0.128
A_Non	24.096	5.432	348.037	91.105	225.632	56.365	5.046	0.777	0.898	0.063
C_Non	27.253	6.072	400.288	113.301	263.943	80.592	5.475	0.909	0.916	0.071
A_IBDU	31.059	6.192	447.047	98.535	302.525	83.584	5.851	1.150	0.918	0.071
C_IBDU	24.856	-	397.276	-	219.500	-	4.407	-	0.857	-

Appendix C – Output of Kruskal-Wallis test on age-diagnosis

Table 3: The table shows the output of the Kruskal-Wallis test performed on combined groups of age and diagnosis from the biopsy data set.

Group	N	Median*	Average Rank	Z-value	Test-statistic (H)	P-value
CD_A	79	3,0	187,5	-1,57		
CD_C	26	56,5	250,3	1,94		
IBDU_A	9	0,0	122,3	-2,15		
IBDU_C	3	27,0	260,2	0,78	37,92	0,000
Non_A	121	4,0	185,8	-2,27	38,61†	0,000†
Non_C	27	8,0	200,1	-0,29		
Possible_A	18	3,0	194,9	-0,42		
UC_A	100	6,5	214,2	0,74		
UC_C	29	209,0	312,5	4,97		

*Median value- number of sequences from OTU4 from each group

† When adjusted for ties

Appendix D - Output of Conover-Inman test on age-diagnosis

Table 4: The table shows the resulting p-value of the Conover-Inman analysis implemented for pairwise comparisons of diagnosis-age groups in the biopsy data set. The test statistic is given in parenthesis for each of the comparisons.

	CD_A	CD_C	UC_A	UC_C	Non_A	Non_C	IBDU_A	IBDU_C
CD_C	0,015 (2,446)	-	-	-	-	-	-	-
UC_A	0,118 (1,567)	0,150 (1,441)	-	-	-	-	-	-
UC_C	0,000 (5,411)	0,026 (2,232)	0,000 (4,431)	-	-	-	-	-
Non_A	0,916 (0,106)	0,009 (2,630)	0,064 (1,859)	0,000 (5,773)	-	-	-	-
Non_C	0,621 (0,494)	0,108 (1,612)	0,562 (0,580)	0,000 (3,938)	0,556 (0,590)	-	-	-
IBDU_A	0,102 (1,637)	0,004 (2,919)	0,020 (2,332)	0,000 (4,549)	0,106 (1,622)	0,075 (1,782)	-	-
IBDU_C	0,278 (1,087)	0,887 (0,142)	0,491 (0,689)	0,402 (0,839)	0,263 (1,120)	0,385 (0,870)	0,069 (1,823)	-
Possible_A	0,799 (0,255)	0,113 (1,587)	0,509 (0,661)	0,000 (3,645)	0,746 (0,324)	0,886 (0,143)	0,116 (1,574)	0,359 (0,919)

Appendix E - Output of Kruskal-Wallis test on age-inflammation

Table 5: the table shows the output of the Kruskal-Wallis test performed on amalgamated groups of age and inflammation status of the biopsy data set with respect to OTU4.

Group	N	Rank sum	Test statistic	P-value
Adult_Non-inflamed	158	19 924,5		
Adult_Inflamed	73	12 724,5	28,114	0,000
Child_Non-inflamed	42	7 052,0		
Child_Inflamed	26	5149,0		

Appendix F – Output of Conover-Inman test on age-inflammation

Table 6: the table shows the output of the Conover-Inman analysis performed on different groups of inflammation and age in the biopsy data set, with respect to OTU4.

Groups compared		Test statistic	P-value
Adult	Child	3,271	0,001
Non-inflamed	Inflamed	4,468	0,000
Adult_Non-inflamed	Adult_Inflamed	4,152	0,000
Adult_Non-inflamed	Child_Non-inflamed	2,935	0,004
Adult_Non-inflamed	Child_Inflamed	4,143	0,000
Adult_Inflamed	Child_Noninflamed	0,403	0,687
Adult_Inflamed	Child_Inflamed	1,267	0,206
Child_Non-inflamed	Child_Inflamed	1,472	0,142

Appendix G – Significant results Kruskal-Wallis test on OTUs from biopsy and water sample dataset

Table 7: The table shows all OTUs and their taxonomic denotation, detected by Kruskal-Wallis test as being significant in explaining the variations between subgroups within the groups of the biopsy and water sample data set respectively. Numbers are given as the percentage of detected sequences from each subgroup belonging to the respective OTUs and reduced to three decimals.

7.1) OTUs explaining differences in diagnosis (Biopsies)

OTU	P-value*	Non	NA†	Possible	CD	IBDU	UC	Taxonomy‡
570	2,43 E-13	0	3	0	0	0	0	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
774	4,33 E-10	0	0	0,222	0	0	0	k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__;s__
539	4,33 E-10	0	0	0,556	0	0	0	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
400	4,33 E-10	0	0	1,056	0	0	0	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__;s__
531	5,02 E-09	0	0	0	0,057	3,083	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Providencia;s__
525	3,95 E-08	0	0	0,944	0	0,167	0	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella;s__
728	1,74 E-06	0,007	0	0,389	0	0	0	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__copri
19	2,05 E-06	63,960	0,6	36,056	18,229	98,583	29,406	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
684	3,36 E-05	0,007	0	0	0	0,583	0,070	k__Bacteria;p__Cyanobacteria;c__4C0d-2;o__YS2;f__;g__;s__
572	7,89 E-05	0,020	0,6	0	0	0	0	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae;g__;s__
582	0,002	1,691	0	5,333	0,276	0,083	0,547	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
288	0,002	0,067	5,6	41,444	0,067	5,333	0,234	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__

56	0,003	11,141	50,4	7,833	14,590	21,333	15,266	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
179	0,012	0,530	3	0,056	1,419	1,5	1,797	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__
628	0,012	0	0	0	0	0	0,242	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];g__;s__
82	0,018	6,154	14,4	0,167	1,533	11,917	10,539	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
793	0,021	0	0,2	0	0	0	0,055	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
374	0,030	0,020	0	1,222	0	0	0,047	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__p-1630-c5
813	0,032	0,275	0,4	0,833	0,133	0,5	0,195	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
569	0,032	0,121	0	0	0,324	0	0,117	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Odoribacteraceae];g__Butyricimonas;s__
232	0,037	0,040	0,8	0,278	0,629	0	1,242	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus;s__anaerobius
149	0,043	5,114	0	4,333	0,133	0,583	0,477	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
112	0,048	0,490	0	0,333	0	0	1,852	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella;s__

7.2 OTUs explaining differences in age-group (Biopsy)

OTU	**P-value***	**Adult**	**Child**	**Taxonomy‡**
643	9,73E-28	0,003	0,023	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Oribacterium;s__
192	2,36E-05	0,570	0,023	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
830	2,49E-05	0,024	0,069	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella;s__

11	3,62E-05	12,282	0,011	k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__
184	0,000	0,882	0,279	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium;s__
61	0,001	0,897	5,372	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megamonas;s__
508	0,011	0,112	0,465	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
399	0,012	0,239	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus;s__
4	0,018	113,185	235,140	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
19	0,027	46,603	19,337	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
103	0,042	1,518	0,919	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus;s__

7.3 OTUs explaining differences in inflammation category (Biopsies)

OTU	P-value*	Non-infl.	NA†	Infl.	Taxonomy‡
91	0,008	11,462	7,992	28,62	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus;s__parainfluenzae
845	0,012	0,116	0,195	1,02	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Peptoniphilus;s__
567	0,029	2,528	1,949	0,79	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__

7.4) OTUs explaining differences in gut-location (Biopsies)

OTU	P-value*	I	VI	VII	IV	II	V	III	Taxonomy‡
718	4,73E-08	0	0	0	0,5	0,022	0,087	0	Unassigned
401	4,31E-06	0,138	0,081	0,714	0	0,022	0,348	0	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Finegoldia;s__
845	3,17E-05	0,275	0,221	1,444	0	0	0,326	0,011	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Peptoniphilus;s__

321	0,000	0,113	0,279	2,159	0	0,044	0	0,011	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__WAL_1855D;s__
784	0,001	0	0,012	0,349	0	0	0	0	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;s__
729	0,001	0,025	0	0,206	0	0	0	0	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__
792	0,002	0,088	0,023	0,048	1,5	0,156	0	0	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Acidaminococcus;s__

7.5) OTUs explaining differences in diagnosis (Water samples)

OTU	P-value*	Non	IBDU	UC	CD	Taxonomy‡
2337	3,57E-08	0	1	0	0,065	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__;g__;s__
2837	9,04E-07	0	0,444	0	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__;g__;s__
2934	2,38E-05	0,039	1,111	0,114	0,032	k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales;f__;g__;s__
2598	0,001	0	0,222	0	0,032	k__Bacteria;p__Bacteroidetes;c__Cytophagia;o__Cytophagales;f__Cytophagaceae;g__Cytophaga;s__
2398	0,002	0,013	1,444	0	0	k__Bacteria;p__Proteobacteria;c__;o__;f__;g__;s__
819	0,003	0	1,333	0,010	0	k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales;f__Caldilineaceae;g__;s__
1438	0,005	0,145	0,778	0,057	0,032	k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria];o__[Chthoniobacterales];f__[Chthoniobacteraceae];g__Candidatus Xiphinematobacter;s__
491	0,009	0,237	1,889	0,152	0,387	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__;s__
599	0,011	0,158	3	0,086	0	k__Bacteria;p__Proteobacteria;c__TA18;o__PHOS-HD29;f__;g__;s__
2966	0,011	0,013	0,444	0,029	0,032	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Stramenopiles;f__;g__;s__
2239	0,032	0,066	0,889	0,048	0,065	k__Bacteria;p__Verrucomicrobia;c__[Pedosphaerae];o__[Pedosphaerales];f__R4-41B;g__;s__
3068	0,044	0,092	0,778	0,019	0	k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria];o__[Chthoniobacterales];f__[Chthoniobacteraceae];g__Candidatus Xiphinematobacter;s__

7.6) OTU explaining difference in age (Water samples)

OTU	P-value *	Adult	Child	Taxonomy‡
2021	0,0216	0	0,135	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__0319-6G20;g__;s__

7.7) OTU`s explaining differences in age-diagnosis combined (Water samples)

OTU	P-value *	AdultIBDU	ChildCD	AdultUC	AdultCD	AdultNon	ChildNon	ChildUC	C-IBDU	Taxonomy‡
1201	2,01E-40	0	0	0	0	0	0	0	3	Unassigned
1958	1,20E-17	0	0	0	0	0,016	0	0	2	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__;g__;s__
2234	1,92E-17	0	0	0	0	0,016	0	0	1	Unassigned
1214	2,60E-10	0	0	0	0	0,032	0	0	1	Unassigned
1697	1,31E-09	0	0	0,01	0	0,048	0	0	1	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__;s__
2346	1,32E-09	0	0	0,07	0	0,016	0	0	1	k__Bacteria;p__TM6;c__SJA-4;o__;f__;g__;s__
2337	1,90E-07	1,13	0	0	0,1	0	0	0	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__;g__;s__
3001	5,52E-07	0	0	0	0	0,065	0	0	1	Unassigned

2493	8,96E-07	0	0,091	0,011	0	0,016	0	0	1	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__;g__;s__
928	1,27E-06	0	0	0	0	0	0	0,364	0	k__Bacteria;p__TM7;c__TM7-1;o__;f__;g__;s__
2245	2,28E-06	0	0	0,043	0	0,048	0,286	0	3	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Procabacteriales;f__Procabacteriaceae;g__;s__
2837	9,04E-06	0,5	0	0	0	0	0	0	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__;g__;s__
537	4,83E-05	0,125	0	0,011	0	0,016	0,357	0	16	k__Bacteria;p__TM7;c__TM7-1;o__;f__;g__;s__
1156	5,29E-05	0	0,273	0,032	0	0	0,071	0	2	k__Bacteria;p__Proteobacteria;c__TA18;o__PHOS-HD29;f__;g__;s__
2204	0,0002	0	0	0,053	0	0,016	0	0,909	1	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Perlucidibaca;s__
1249	0,0002	0	0	0,011	0	0,065	0	0	1	Unassigned
2934	0,0005	1,125	0	0,117	0,05	0,032	0,071	0,091	1	k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales;f__;g__;s__
2398	0,0044	1,625	0	0	0	0	0,071	0	0	k__Bacteria;p__Proteobacteria;c__;o__;f__;g__;s__
2673	0,0079	0	0	0	0	0	0	0,182	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__0319-6G20;g__;s__
1546	0,0079	0	0,182	0	0	0	0	0	0	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Stramenopiles;f__;g__;s__
2165	0,0079	0	0,364	0	0	0	0	0	0	Unassigned
2159	0,0079	0	0	0	0	0	0	0,545	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__;g__;s__
2811	0,0079	0	0	0	0	0	0	0,182	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__;s__
2753	0,0079	0	0	0	0	0	0	0,364	0	k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales;f__Caldilineaceae;g__;s__

1085	0,0079	0	0	0	0	0	0	0,364	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__;g__;s__
2216	0,0079	0	0	0	0	0	0	0,182	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__Aquicella;s__
2606	0,0079	0	0	0	0	0	0	0,182	0	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Procabacteriales;f__Procabacteriaceae;g__;s__
1108	0,0079	0	0	0	0	0	0	2,273	0	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Vogesella;s__
557	0,0079	0	0	0	0	0	0	2,545	0	k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales;f__Caldilineaceae;g__;s__
2485	0,0079	0	0	0	0	0	0	0,455	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__;s__
1271	0,0079	0	0	0	0	0	0	0,636	0	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae
1434	0,0079	0	0	0	0	0	0	0,273	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter;s__
2074	0,0079	0	0	0	0	0	0	0,636	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Crenotrichaceae;g__Crenothrix;s__
2096	0,0079	0	0	0	0	0	0	0,273	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__;s__
1643	0,0079	0	0	0	0	0	0	0,273	0	k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__;g__;s__
2286	0,0079	0	0	0	0	0	0	0,545	0	k__Bacteria;p__Chlamydiae;c__Chlamydiia;o__Chlamydiales;f__Parachlamydiaceae;g__Parachlamydia;s__
1088	0,0079	0	0	0	0	0	0	1	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__;s__
1999	0,0079	0	0	0	0	0	0	0,545	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__;g__;s__
979	0,0079	0	0	0	0	0	0	0,727	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__MND4;g__;s__
972	0,0079	0	0	0	0	0	0	0,636	0	k__Bacteria;p__Actinobacteria;c__Thermoleophilia;o__Solirubrobacterales;f__;g__;s__
2616	0,0079	0	0	0	0	0	0	0,273	0	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__;s__
1855	0,0083	0	0	0,053	0	0,016	0	0	1	k__Bacteria;p__Chloroflexi;c__S085;o__;f__;g__;s__

2598	0,0098	0,25	0	0	0,05	0	0	0	0	k__Bacteria;p__Bacteroidetes;c__Cytophagia;o__Cytophagales;f__Cytophagaceae;g__Cytophaga;s__
491	0,0109	2,125	0,455	0,160	0,35	0,210	0,357	0,091	0	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__;s__
2688	0,0141	0,125	0	0,053	0	0,016	0	0	3	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__;s__
2021	0,0152	0	0,182	0	0	0	0,214	0	0	k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__0319-6G20;g__;s__
1438	0,0180	0,75	0	0,064	0,05	0,081	0,429	0	1	k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria];o__[Chthoniobacterales];f__[Chthoniobacteraceae];g__Candidatus Xiphinematobacter;s__
2966	0,0213	0,5	0,091	0,021	0	0,016	0	0,091	0	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Stramenopiles;f__;g__;s__
819	0,0327	1,5	0	0,011	0	0	0	0	0	k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales;f__Caldilineaceae;g__;s__
599	0,0346	3,375	0	0,074	0	0,161	0,143	0,182	0	k__Bacteria;p__Proteobacteria;c__TA18;o__PHOS-HD29;f__;g__;s__
2832	0,0369	0,125	0	0,032	0,05	0,016	0,071	0	1	k__Bacteria;p__Chlamydiae;c__Chlamydiia;o__Chlamydiales;f__Rhabdochlamydiaceae;g__Candidatus Rhabdochlamydia;s__
1454	0,0443	0	0	0,085	0,05	0,065	0,214	0,818	3	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__;g__;s__

*Bonferroni-corrected P-value is given.

† NA = Marked as not applicable during data analysis, as no group category is available.

‡ k = kingdom, p = phyla, c = class, f = family, g = genus, s = species

Appendix H – Results of OTU matching using Jukes-Cantor

Table 8: The table shows all OTUs from biopsy and water samples that could be aligned with ≥ 97% identity in Matlab, and the respective taxonomy of the maches. Taxonomic levels that could only be assigned to one of the datasets are marked as (b) or (w) for biopsy or water sample set respectively.

OTU-matches		Distance	Taxonomy[†]
Biopsy	**Water**		
OTU_1	OTU_1106	0,0207	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_2	OTU_567	0,0230	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__
OTU_3	OTU_1013	0,0115	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__uniformis
OTU_4	OTU_2	0,0138	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_5	OTU_487	0,0091	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus];s__gnavus
OTU_6	OTU_2741	0,0207	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__fragilis
OTU_7	OTU_923	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus];s__
OTU_8	OTU_552	0,0160	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_8	OTU_716	0,0277	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_9	OTU_1545	0,0138	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__
OTU_12	OTU_210	0,0114	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
OTU_12	OTU_597	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia (w) ;s__
OTU_13	OTU_521	0,0139	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_14	OTU_140	0,0023	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_15	OTU_254	0,0185	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__
OTU_16	OTU_601	0,0161	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__ovatus
OTU_17	OTU_247	0,0254	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_17	OTU_1258	0,0114	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_19	OTU_1196	0,0091	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia (w) ;s__
OTU_22	OTU_338	0,0254	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_22	OTU_611	0,0278	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_22	OTU_865	0,0137	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Salmonella (w)
OTU_22	OTU_1503	0,0207	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter (w)
OTU_22	OTU_1751	0,0230	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__

OTU_24	OTU_2278	0,0138	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Odoribacteraceae];g__Odoribacter;s__
OTU_25	OTU_565	0,0069	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_26	OTU_1058	0,0091	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Barnesiellaceae];g__;s__
OTU_27	OTU_713	0,0114	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
OTU_28	OTU_1604	0,0115	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__caccae
OTU_29	OTU_758	0,0138	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__copri
OTU_32	OTU_193	0,0207	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_32	OTU_244	0,0160	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_34	OTU_712	0,0068	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus];s__
OTU_36	OTU_1422	0,0161	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_37	OTU_1452	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_39	OTU_558	0,0069	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea
OTU_40	OTU_361	0,0137	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae
OTU_42	OTU_2141	0,0184	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__distasonis
OTU_45	OTU_87	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_45	OTU_3016	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_47	OTU_753	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__
OTU_48	OTU_536	0,0114	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira;s__
OTU_51	OTU_480	0,0161	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
OTU_53	OTU_716	0,0161	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_54	OTU_1604	0,0278	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__caccae (w)
OTU_56	OTU_271	0,0115	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
OTU_57	OTU_678	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae (b) ;g__Coprococcus (b)
OTU_59	OTU_370	0,0161	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
OTU_62	OTU_185	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus (b)
OTU_66	OTU_1045	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae;g__;s__
OTU_70	OTU_576	0,0114	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_72	OTU_133	0,0138	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila

OTU_72	OTU_2570	0,0231	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila
OTU_73	OTU_1195	0,0160	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_75	OTU_1288	0,0139	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
OTU_76	OTU_760	0,0280	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia (w)
OTU_76	OTU_2189	0,0139	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
OTU_79	OTU_3080	0,0162	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus;s__
OTU_80	OTU_51	0,0091	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_80	OTU_212	0,0207	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
OTU_80	OTU_338	0,0138	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_80	OTU_611	0,0254	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_80	OTU_702	0,0278	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
OTU_84	OTU_231	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
OTU_84	OTU_1073	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae (w) ;g__SMB53 (w)
OTU_86	OTU_1392	0,0068	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae (b)
OTU_91	OTU_1183	0,0161	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Haemophilus; s__parainfluenzae
OTU_93	OTU_1528	0,0160	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__eggerthii
OTU_95	OTU_605	0,0068	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__;s__
OTU_96	OTU_971	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;s__
OTU_98	OTU_173	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_99	OTU_230	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_100	OTU_567	0,0254	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae (w) f__Bacteroidaceae (b) ;g__Bacteroides (b)
OTU_100	OTU_716	0,0254	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_100	OTU_2544	0,0137	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_102	OTU_740	0,0184	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter;s__
OTU_102	OTU_987	0,0207	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter
OTU_103	OTU_183	0,0115	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
OTU_104	OTU_1652	0,0091	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_105	OTU_1307	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__

OTU_106	OTU_868	0,0092	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
OTU_107	OTU_1296	0,0023	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;s__
OTU_110	OTU_2141	0,0184	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__distasonis
OTU_111	OTU_1191	0,0069	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
OTU_114	OTU_2923	0,0068	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Barnesiellaceae];g__;s__
OTU_117	OTU_451	0,0161	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__
OTU_119	OTU_1047	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_122	OTU_2195	0,0114	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_125	OTU_660	0,0023	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_128	OTU_696	0,0208	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea;s__
OTU_130	OTU_711	0,0139	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_131	OTU_1601	0,0278	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__[Eubacterium];s__biforme
OTU_132	OTU_478	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_135	OTU_495	0,0230	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium;s__prausnitzii
OTU_139	OTU_567	0,0207	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__
OTU_140	OTU_1401	0,0092	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_141	OTU_971	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira (w)
OTU_142	OTU_1127	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;s__
OTU_145	OTU_40	0,0278	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
OTU_145	OTU_282	0,0069	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
OTU_146	OTU_1106	0,0184	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_152	OTU_1371	0,0161	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_158	OTU_597	0,0254	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae (w);g__Roseburia (w)
OTU_162	OTU_1357	0,0115	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Holdemania;s__
OTU_164	OTU_184	0,0069	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__longum
OTU_166	OTU_384	0,0138	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
OTU_171	OTU_1144	0,0160	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__

OTU_173	OTU_3140	0,0186	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_176	OTU_342	0,0184	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__;s__
OTU_176	OTU_2795	0,0091	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__;s__
OTU_179	OTU_179	0,0137	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
OTU_179	OTU_217	0,0230	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis (w)
OTU_183	OTU_758	0,0254	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__copri
OTU_186	OTU_2831	0,0069	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar
OTU_190	OTU_2409	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Anaerostipes;s__
OTU_191	OTU_746	0,0184	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
OTU_192	OTU_40	0,0278	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
OTU_197	OTU_299	0,0116	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_198	OTU_331	0,0139	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_199	OTU_626	0,0144	k__Archaea;p__Euryarchaeota;c__Methanobacteria;o__Methanobacteriales;f__Methanobacteriaceae;g__Methanobrevibacter;s__
OTU_202	OTU_1195	0,0277	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_203	OTU_120	0,0161	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas;s__
OTU_203	OTU_425	0,0254	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas;s__
OTU_207	OTU_2831	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar
OTU_209	OTU_179	0,0253	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis (b)
OTU_209	OTU_217	0,0115	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis
OTU_211	OTU_820	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_212	OTU_2432	0,0162	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
OTU_214	OTU_2150	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_218	OTU_2570	0,0278	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila
OTU_219	OTU_1202	0,0207	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_222	OTU_390	0,0093	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_224	OTU_1571	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;s__

OTU_231	OTU_2842	0,0208	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus;s__eutactus
OTU_243	OTU_422	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Anaerostipes;s__
OTU_251	OTU_8	0,0184	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
OTU_251	OTU_103	0,0137	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
OTU_253	OTU_462	0,0161	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__Methylobacterium;s__
OTU_255	OTU_269	0,0161	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__[Eubacterium];s__dolichum
OTU_257	OTU_1389	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_260	OTU_982	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_261	OTU_1417	0,0000	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__
OTU_262	OTU_2992	0,0254	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_263	OTU_8	0,0184	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
OTU_263	OTU_103	0,0254	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
OTU_272	OTU_2097	0,0185	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
OTU_274	OTU_1150	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae (w)
OTU_277	OTU_397	0,0254	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella;s__
OTU_277	OTU_2875	0,0208	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella;s__
OTU_285	OTU_1565	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_287	OTU_907	0,0231	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Peptoniphilus;s__
OTU_288	OTU_1545	0,0254	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__
OTU_289	OTU_87	0,0280	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_289	OTU_232	0,0209	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_289	OTU_3016	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_290	OTU_1154	0,0091	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
OTU_291	OTU_2992	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_294	OTU_1601	0,0184	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__[Eubacterium];s__biforme

OTU_304	OTU_1798	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_306	OTU_758	0,0115	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__copri
OTU_309	OTU_231	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae (w) f__Peptostreptococcaceae (b)
OTU_313	OTU_1942	0,0161	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas;s__
OTU_316	OTU_324	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_317	OTU_2257	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_319	OTU_611	0,0231	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter (b)
OTU_319	OTU_865	0,0231	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Salmonella (w) g__Citrobacter (b)
OTU_319	OTU_1503	0,0161	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter
OTU_319	OTU_1751	0,0278	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter (b)
OTU_320	OTU_448	0,0115	k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__;s__
OTU_323	OTU_232	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_323	OTU_3016	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_331	OTU_87	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_331	OTU_3016	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_332	OTU_1240	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_337	OTU_2106	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_341	OTU_353	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae
OTU_341	OTU_1073	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__SMB53 (w)
OTU_346	OTU_597	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia;s__
OTU_349	OTU_331	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_349	OTU_684	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_351	OTU_1172	0,0023	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus];s__
OTU_366	OTU_1305	0,0138	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Eggerthella;s__lenta
OTU_367	OTU_2372	0,0184	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Novosphingobium;s__
OTU_370	OTU_883	0,0092	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Micrococcus;s__

OTU_378	OTU_1739	0,0208	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__;g__;s__
OTU_379	OTU_2992	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_390	OTU_3052	0,0161	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus];s__gnavus
OTU_396	OTU_1150	0,0137	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
OTU_398	OTU_637	0,0069	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];g__;s__
OTU_399	OTU_1183	0,0231	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus;s__parainfluenzae (w)
OTU_401	OTU_2601	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Finegoldia;s__
OTU_408	OTU_247	0,0160	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_408	OTU_1258	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_410	OTU_818	0,0184	k__Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__Chitinophagaceae;g__Sediminibacterium;s__
OTU_415	OTU_695	0,0116	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_420	OTU_2078	0,0023	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__
OTU_427	OTU_2831	0,0255	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar (w) s__parvula (b)
OTU_429	OTU_2918	0,0115	k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__;s__
OTU_435	OTU_1914	0,0278	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;s__
OTU_436	OTU_1267	0,0210	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__;s__
OTU_439	OTU_1047	0,0254	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_441	OTU_1798	0,0207	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_451	OTU_1823	0,0254	k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__;s__
OTU_459	OTU_1213	0,0161	k__Bacteria;p__TM7;c__TM7-3;o__;f__;g__;s__
OTU_461	OTU_13	0,0207	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax (w) g__Delftia (b)
OTU_461	OTU_99	0,0278	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delfia (b)
OTU_461	OTU_112	0,0231	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b)
OTU_461	OTU_276	0,0207	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b)
OTU_461	OTU_386	0,0184	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax (w) g__Delftia (b)
OTU_461	OTU_393	0,0277	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b)

OTU_461	OTU_790	0,0115	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia;s__
OTU_461	OTU_1071	0,0207	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b)
OTU_461	OTU_1126	0,0278	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b)
OTU_461	OTU_1701	0,0254	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Delftia (b)
OTU_461	OTU_2650	0,0254	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Rhodoferax (w) g__Delftia (b)
OTU_462	OTU_2755	0,0161	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__
OTU_468	OTU_567	0,0254	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__
OTU_471	OTU_1819	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
OTU_479	OTU_353	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae (w)
OTU_479	OTU_1073	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae (w);g__SMB53 (w)
OTU_490	OTU_2128	0,0230	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria;s__Subflava (b)
OTU_496	OTU_1011	0,0184	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_497	OTU_51	0,0161	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_497	OTU_212	0,0184	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
OTU_497	OTU_338	0,0231	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_499	OTU_1813	0,0137	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_515	OTU_157	0,0255	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__;s__
OTU_515	OTU_188	0,0207	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__;s__
OTU_517	OTU_1914	0,0091	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;s__
OTU_527	OTU_684	0,0255	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_533	OTU_611	0,0161	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_533	OTU_865	0,0161	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Salmonella (w)
OTU_533	OTU_1503	0,0278	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter (w)
OTU_533	OTU_1751	0,0254	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_537	OTU_2996	0,0091	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae;g__;s__
OTU_549	OTU_1204	0,0160	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
OTU_550	OTU_140	0,0277	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__

OTU_552	OTU_1788	0,0161	k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__Gemella (b)
OTU_553	OTU_217	0,0277	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis (w)
OTU_554	OTU_2037	0,0185	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_556	OTU_1147	0,0161	k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__;s__
OTU_563	OTU_2128	0,0184	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria;s__
OTU_572	OTU_1025	0,0138	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae;g__;s__
OTU_578	OTU_870	0,0115	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__melaninogenica
OTU_581	OTU_154	0,0115	k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__;g__;s__
OTU_582	OTU_271	0,0162	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
OTU_588	OTU_3046	0,0069	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
OTU_589	OTU_51	0,0207	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_589	OTU_212	0,0231	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
OTU_589	OTU_338	0,0160	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
OTU_606	OTU_1053	0,0207	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_623	OTU_22	0,0231	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Oxalobacter (b)
OTU_623	OTU_320	0,0207	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Oxalobacter (b)
OTU_623	OTU_616	0,0254	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Oxalobacter (b)
OTU_632	OTU_1106	0,0207	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
OTU_636	OTU_760	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__
OTU_636	OTU_2189	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia (b)
OTU_642	OTU_2374	0,0185	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae (w) f__[Mogibacteriaceae] (b)
OTU_644	OTU_151	0,0068	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__;s__
OTU_647	OTU_124	0,0184	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__;s__
OTU_647	OTU_194	0,0278	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__;s__
OTU_647	OTU_723	0,0254	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__;s__
OTU_647	OTU_1243	0,0207	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__;s__
OTU_651	OTU_1137	0,0161	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__;g__;s__
OTU_656	OTU_8	0,0278	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas

OTU_656	OTU_103	0,0138	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas
OTU_657	OTU_232	0,0069	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
OTU_666	OTU_2227	0,0184	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__Horikoshii (b)
OTU_667	OTU_516	0,0161	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Paracoccus (w)s__aminovorans (w)
OTU_669	OTU_521	0,0280	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_671	OTU_551	0,0207	k__Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__Chitinophagaceae;g__;s__
OTU_672	OTU_148	0,0186	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Phenylobacterium
OTU_678	OTU_763	0,0278	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Kocuria;s__palustris (w) s__Rhizophila (b)
OTU_678	OTU_2524	0,0208	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Kocuria;s__rhizophila
OTU_687	OTU_2743	0,0256	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Anaerococcus;s__
OTU_698	OTU_331	0,0232	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_699	OTU_631	0,0209	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
OTU_710	OTU_6	0,0278	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Polaromonas (w)
OTU_710	OTU_515	0,0277	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax (w) s__delafieldii (w)
OTU_710	OTU_652	0,0278	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__;s__
OTU_714	OTU_584	0,0092	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
OTU_720	OTU_2880	0,0278	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae;g__;s__
OTU_721	OTU_1819	0,0231	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
OTU_723	OTU_8	0,0138	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__Veronii (b)
OTU_723	OTU_103	0,0161	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__Veronii (b)
OTU_737	OTU_365	0,0138	k__Bacteria;p__Acidobacteria;c__Solibacteres;o__Solibacterales;f__;g__;s__
OTU_758	OTU_1539	0,0115	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
OTU_778	OTU_24	0,0207	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Bradyrhizobium (b)

OTU_778	OTU_477	0,0184	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae g__Bradyrhizobium (b)
OTU_816	OTU_643	0,0234	k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__;s__
OTU_817	OTU_286	0,0138	k__Bacteria;p__Bacteroidetes;c__Cytophagia;o__Cytophagales;f__Cytophagaceae;g__;s__
OTU_827	OTU_2544	0,0230	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae (w) f__Rikenellaceae (b) g__Bacteroides (w)_
OTU_833	OTU_2923	0,0278	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Barnesiellaceae];g__;s__
OTU_835	OTU_270	0,0184	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__
OTU_841	OTU_1793	0,0184	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae
OTU_845	OTU_907	0,0208	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Peptoniphilus;s__
OTU_853	OTU_8	0,0138	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
OTU_853	OTU_103	0,0254	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas
OTU_856	OTU_8	0,0231	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__Viridiflava (b)
OTU_856	OTU_890	0,0278	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__Viridiflava (b)
OTU_863	OTU_190	0,0091	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas;s__
OTU_863	OTU_598	0,0277	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas (b)
OTU_863	OTU_2116	0,0184	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas (b)
OTU_871	OTU_631	0,0280	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium (b)
OTU_878	OTU_1788	0,0161	k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__Gemella (b)
OTU_879	OTU_233	0,0115	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Cryocola;s__
OTU_885	OTU_1200	0,0115	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Lactococcus;s__
OTU_891	OTU_9	0,0045	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas;s__yabuuchiae
OTU_907	OTU_2628	0,0162	k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__;s__
OTU_909	OTU_988	0,0278	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__
OTU_909	OTU_2568	0,0231	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus;s__

OTU_910	OTU_294	0,0184	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__Methylobacterium
OTU_912	OTU_85	0,0254	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__Methylobacterium
OTU_931	OTU_277	0,0091	k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter;s__pylori

† k = kingdom, p = phyla, c = class, f = family, g = genus, s = species

Appendix I - Matches from Jukes-Cantor dataset used in Fisher exact testing

Table 9: The table shows the number of sequences from the respective OTUs identified in the matches implicated to be of importance in the Jukes-Cantor data set and comprise the material used for calculations of Fisher exact value. Sequence number in all samples retrieved from patients presenting both water and biopsies is presented.

	Match 7		Match 6 and 5			Match 4		Match 3		Match 2		Match 1	
Patient number	**OTU 891 (b)**	**OTU 9 (w)**	**OTU 145 (b)**	**OTU 192 (b)**	**OTU 40 (w)**	**OTU 623 (b)**	**OTU 22 (w)**	**OTU 46 (b)**	**OTU 112 (w)**	**OTU 778 (b)**	**OTU 24 (w)**	**OTU 710 (b)**	**OTU 6 (w)**
8	0	13	2	0	8	0	51	0	6	1	22	0	1684
8	0		0	0		0		0		0		0	
32	0	0	0	0	24	0	24	0	1	0	1	0	2
32	0		0	0		0		0		0		0	
36	0	1	3	0	9	0	324	0	0	1	7	0	26
36	0		1	0		2		0		0		0	
37	0	0	0	0	50	0	53	0	0	0	5	0	0
37	0		1	0		0		0		0		0	
39	0	0	0	0	0	0	2	0	1	0	1	0	6
39	0		1	0		0		0		0		0	
39	0		0	0		0		0		1		0	
40	0	0	1	0	2	0	0	0	26	2	448	0	13
41	0	0	0	0		0	9	0	62	0	2	0	29
41	0		0	0		0		0		0		0	
42	0	0	5	0	7	0	233	0		0	0	0	329
42	0		0	0		0		0	1	0		0	
43	0	9	0	0	0	0	23	0		0	48	0	0
43	0		0	0		0		0	0	0		0	
44	0	0	0	0	0	0	0	1	1	0	1441	0	137
44	0		0	0		0		0		0		0	
44	0		0	0		0		0		0		0	
47	0	0	0	0	13	0	0	0	0	0	169	0	3
47	0		0	0		0		0		0		0	
48	0	0	0	0	46	0	0	0	0	0	7	0	2

48	0		0	0		0		0		2		0	
48	0		0	0		0		0		0		0	
49	0	0	0	0	0	0	1	0	0	0	23	0	0
49	0		0	0		0		0		0		0	
55	0	0	1	0	1	0		0	0	0	11	0	162
55	0		0	1		0		0		0		1	
55	0		1	0		0	3	0		0		0	
58	0	0	0	0	4	0	263	0	338	0	7	0	6
58	0		0	0		0		0		1		0	
59	0		0	0	0	0	0	0	0	0	4	0	3
59	0		0	0		0		0		0		0	
59	0		0	0		0		0		0		0	
59	0		0	0		0		0		0		0	
60	0	0	3	0	0	0	0	0	4	0	104	0	0
60	0		0	0		1		1		0		0	
60	0		0	0		0		0		1		0	
61	0	0	0	0	0	0	1	0	0	0	0	0	0
62	0	2	0	0	0	0	0	0	0	0	15	0	1
62	0		0	0		0		0		0		0	
63	0	16	0	0	3	0	0	0	0	0	5	0	0
63	0		2	0		0		0		0		0	
64	0	5	0	0	0	0	0	0	114	0	5	0	21
64	0		1	0		0		0		0		0	
64	0		1	0		0		0		0		0	
65	0	0	16	0	0	0	0	0	0	2	43	4	10
66	0	0	0	0	0	0	172	0	0	0	42	0	1
66	0		0	0		0		0		1		0	
67	0	0	1	0	0	0	0	0	0	0	3	0	4
67	0		0	0		0		0		0		0	
69	0	0	0	0	35	0	0	0	0	0	585	0	0
70	0		0	0	6	0	0	0	0	1	36	0	2

70	0	3	0	0		0		0		0		0	
71	0	9	0	0	0	0	0	0	0	0	102	0	5
71	0		0	0		0		0		1		0	
73	0	22	0	0	5	0	0	0	0	0	63	0	0
73	0		0	0		0		0		0		0	
75	0	0	1	0	1	0	0	0	1	2	6	0	0
75	0		0	0		0		0		0		0	
76	0	0	11	0	0	0	0	0	0	0	3	0	0
76	0		16	0		0		0		0		0	
76	0		29	0		0		0		0		0	
77	0	0	0	0	117	0	0	0	0	0	7	0	0
77	0		0	0		0		0		0		0	
78	0	0	3	0	0	0	0	0	0	0	3	0	1058
78	0		6	0		0		0		0		0	
78	0		8	0		0		0		0		0	
79	0	0	0	0	0	0	0	0	0	0	7	0	3
79	0		0	0		0		0		0		0	
81	0	2	1	0	26	0	8	0	1	0	12	0	165
81	0		0	0		0		0		1		0	
81	0		0	0		0		0		0		0	
83	0	0	0	0	36	0	37	0	1	0	4	0	1178
83	0		0	0		0		0		0		0	
83	0		0	0		0		0		0		0	
85	0	17	16	0	38	0	8	0	0	0	6	0	14
86	0	1	2	0	8	0	4	0	14	0	36	0	11
86	0		0	0		0		0		0		0	
86	0		0	0		0		0		0		0	
87	0	0	0	0	56	0	1	0	1	0	3	0	0
87	0		2	0		0		0		0		0	
88	0	0	0	0	40	2	0	0	1	0	18	0	6

90	0	0	1	0	9	0	10	0	1	0	8	0	716
90	0		2	0		0		0		0		0	
90	0		5	0		0		0		0		0	
93	0	1	2	0	11	0	0	0	222	0	19	0	1
93	0		0	0		0		0		0		0	
95	0	0	0	0	21	0	18	0	0	0	10	0	7
95	0		0	0		0		0		0		0	
97	0	1	0	0	19	0	156	0	0	0	2	0	0
97	0		8	0		0		0		0		0	
98	0	0	1	0	4	0	0	0	0	0	11	0	1724
98	0		0	0		0		0		0		0	
101	0	0	0	0	14	0	3	0	418	0	62	0	7
101	0		0	0		0		0		0		0	
102	0	1	0	0	21	0	0	0	1	0	53	0	164
102	0		0	0		0		0		0		0	
102	0		0	0		0		0		0		0	
103	0	0	0	0	2	0	0	0	2	0	63	0	0
103	0		0	0		0		0		0		0	
104	0	2	11	0	3	0	0	0	0	0	14	0	25
104	0		18	0		0		0		1		0	
104	0		14	0		0		0		0		0	
105	0	0	0	0	17	0	0	0	0	0	65	0	1100
105	0		1	0		0		0		0		0	
105	0		1	0		0		0		0		0	
106	0	1	54	0	66	0	0	0	1	0	567	0	67
106	0		49	0		0		0		0		0	
106	0		34	0		0		0		0		0	
107	0	16	0	0	23	0	1	0	0	1	211	0	1155
107	0		1	0		0		0		0		0	
107	0		0	0		0		0		0		0	
107	0		2	0		0		0		0		0	

108	0	0	0	0	2	0	0	0	173	0	195	0	1
108	0		0	0		0		0		0		0	
109	0	0	0	0	51	0	0	0	0	0	25	0	4
110	0	2	1	0	78	0	19	0	0	3	25	0	373
110	0		1	0		0		0		4		2	
110	0		0	0		0		1		3		1	
112	0	3	1	0	77	0	483	0	17	0	12	0	0
112	0		0	1		0		0		0		0	
112	0		0	0		0		0		0		0	
113	0	15	0	0	23	0	1	0	0	1	1	0	1
113	0		0	0		0		0		1		0	
113	0		0	0		0		0		0		0	
116	0	0	0	0	4	0	0	0	1	0	16	0	1740
116	0		1	0		0		0		0		0	
116	0		2	0		0		0		0		0	
116	0		0	0		0		0		0		0	
117	0	0	0	0	60	0	0	0	1	1	335	0	16
117	0		0	0		0		0		0		0	
117	0		0	0		0		0		1		0	
120	0	0	0	0	15	0	0	0	1	0	57	0	89
120	0		1	0		0		0		0		0	
122	0	0	0	0	42	0	0	0	2	0	1	0	19
122	0		0	0		0		0		0		0	
125	0	0	0	0	20	0	0	0	2	0	42	0	29
125	0		0	0		0		0		0		0	
132	0	24	0	0	54	0	4	0	1	0	12	0	9
132	0		0	0		0		0		0		0	
132	0		0	0		0		0		0		0	
1009	0	0	0	0	0	1	0	0	0	0	59	0	1
1009	0		0	0		0		0		0		0	

1009	0		2	0		1		0		0		0	
1010	0	0	1	0	5	0	0	0	1	0	330	0	15
1010	0		0	0		0		0		0		0	
1010	0		2	0		0		0		0		0	
1011	0	11	0	0	25	0	0	0	1	0	463	0	0
1011	0		0	0		0		0		0		0	
1011	0		0	0		0		0		0		0	
1012	0	54	0	0	14	0	261	0	1	0	3	0	3
1012	0		2	0		0		0		0		0	
1012	0		0	0		0		0		0		0	
1013	0	0	0	0	0	0	0	0	10	0	2	0	30
1013	0		0	0		0		0		0		0	
1013	0		0	0		0		0		0		0	
1014	0	1	1	0	0	0	0	0	77	0	63	0	340
1014	0		0	0		0		0		0		0	
1014	0		0	0		0		0		0		0	
1015	0	0	0	0	9	0	0	0	1	0	2	0	22
1015	0		0	0		0		0		0		0	
1015	0		1	0		0		0		0		0	
1017	0	0	1	0	33	0	6	0	1	0	56	0	85
1017	0		0	0		0		0		0		1	
1018	0	1	0	0	0	0	1	0	1	0	3	0	13
1018	0		0	0		0		0		0		0	
1018	0		0	0		0		0		0		0	
1021	0	0	9	0	37	0	0	0	61	17	45	0	6
2005	0	7	4	0	0	0	0	0	41	0	100	0	0
2005	0		5	0		0		0		0		1	
2005	0		8	0		0		0		0		0	
2006	0	0	0	0	1	0	2	0	43	0	1	0	823
2006	0		0	0		0		0		0		0	
2006	0		0	1		0		0		0		0	

2007	0	1	0	0	0	0	0	0	138	1	92	0	735
2007	0		1	0		0		0		0		0	
2007	0		1	0		0		0		0		0	
2009	0	1	1	0	0	0	0	0	12	0	3	0	0
2009	0		0	0		0		0		0		0	
2009	0		0	0		0		0		0		0	
2011	0	1	0	0	18	0	1	0	0	0	4	1	1
2011	0		2	0		0		0		0		0	
2011	0		0	0		0		0		0		0	
2012	0	1	0	0	19	0	1	0	0	0	1	0	1
2012	0		0	0		0		0		0		0	
2012	0		0	0		0		0		0		0	
2012	0		0	0		0		0		0		0	
2013	0	0	0	0	38	0	0	0	0	0	12	0	1
2013	0		0	0		0		0		0		0	
2013	0		0	0		0		0		0		0	
2014	0	0	32	0	0	0	0	0	0	0	5	0	0
2014	0		1	3		0		0		0		0	
2015	0	0	0	0	0	0	0	0	0	0	143	0	0
2015	0		0	0		0		0		0		0	
2016	0	0	0	0	0	0	0	0	1	0	1	0	1
2016	0		0	0		0		0		0		0	
2017	0	0	1	0	9	0	0	0		0	2	0	118
2017	0		0	0		0		0	112	0		0	
2020	0	0	0	0	42	1	0	0	0	0	8	0	761
2020	0		0	0		0		0		0		0	
2020	0		1	0		0		0		0		0	
2021	0	0	0	0	89	0	0	0	0	1	74	0	92
2021	0		1	0		0		0		0		0	
2021	0		0	0		0		0		0		0	

2021	0		1	0		0		0		0		0	
2022	0	10	0	0	138	0	0	0	0	0	0	0	0
2022	0		0	0		0		0		0		0	
2022	0		0	0		0		0		0		0	
2023	0	0	1	0	15	0	3	0	10	0	54	0	0
2023	0		4	0		0		0		0		0	
2024	0	7	1	0	12	0	1	0	437	0	14	0	1
2024	0		0	0		0		0		0		0	
2024	0		0	0		0		0		0		0	
5001	0	1	0	0	139	0	0	0	2	0	40	0	24
5001	0		0	0		0		0		0		0	
5001	0		0	0		0		0		0		0	
5004	0	0	1	0	0	0	0	0	0	1	14	0	1
5004	0		1	0		0		0		0		0	
5004	0		0	0		0		0		0		0	
5005	0	1	0	0	0	0	25	0	1	0	1	0	1070
5005	0		0	0		0		0		0		0	
5005	0		0	0		0		0		0		0	
5008	0	0	1	0	0	0	2	0	0	0	25	0	0
5008	0		0	0		0		0		0		0	
5008	0		2	0		0		0		0		0	
5009	0	0	0	0	24	0	0	0	0	0	14	0	0
5009	0		1	0		0		1		0		1	
5010	0	48	1	0	14	0	0	0	0	0	40	0	1
5010	0		1	0		0		0		0		0	
5011	0	0	0	0	3	1	0	0	175	0	14	0	4
5011	0		0	0		0		0		0		0	
5013	0	0	0	0	39	0	32	0	0	0	111	0	4
5013	0		0	0		0		0		0		0	
5013	0		0	0		0		0		0		0	
6001	0	80	17	0	13	0	1	0	2	0	95	0	8

6001	0		0	0		0		0		0		0	
6001	0		0	0		0		0		1		0	
6002	0	0	0	0	19	0	0	0	0	0	484	0	1
6002	0		14	0		0		0		0		0	
6003	0	0	0	0	105	0	0	0	0	1	5	0	9
6003	0		0	0		0		0		0		0	
6005	0	0	0	0	9	0	47	0	1	0	1	0	211
6005	0		1	0		0		0		0		0	
6005	0		0	0		0		0		0		0	
6005	0		0	0		0		0		0		0	
6006	0	0	0	1	0	0	0	0	1	0	13	0	0
6006	0		0	0		0		0		0		0	
6006	0		0	0		0		0		0		4	
6008	0	0	3	0	1	0	0	0	0	1	50	0	0
6008	0		12	0		0		0		0		0	
6009	0	3	0	0	1	0	1	0	1	0	38	0	6
6009	0		0	0		0		0		0		0	
6011	0	0	2	0	0	0	1	0	13	0	2	0	472
6011	0		0	0		0		0		0		0	
6014	0	0	1	0	22	0	0	0	41	0	18	0	0
6014	0		2	0		0		0		0		0	
6014	0		1	0		0		0		0		0	
6015	0	0	0	0	0	0	0	0	211	0	12	0	0
6015	0		1	0		0		0		0		0	
6016	0	10	1	0	39	0	0	0	0	0	12	0	2
6017	0	0	4	0	18	1	4	0	1	0	8	0	638
6017	0		3	0		0		0		0		1	
6017	0		1	0		0		0		0		0	
6017	0		6	0		0		0		0		0	
6018	0	0	0	0	19	0	0	0	2	0	3	0	754

6019	0	1	0	0	1	0	0	0	0	0	143	0	1
6019	0		0	0		0		0		0		0	
6020	0	0	0	0	7	0	1	0	0	1	177	0	0
6022	0	0	0	0	27	0	0	0		0	133	0	0
6022	0		0	1		0		0	1	0		0	
6023	0	0	0	0	18	0	0	0	2	0	19	0	0
6023	0		0	0		0		0		0		0	
6023	0		0	0		0		0		0		0	

Appendix J – Matches from biopsy data set used in Fisher exact testing

Table 10: The table shows the number of sequences from the respective OTUs identified in the matches implicated to be of importance in the biopsy data set and comprise the material used for calculations of Fisher exact value. Sequence number in all samples retrieved from patients presenting both water and biopsies is presented.

Patient number	Match 7		Match 6		Match 5 and 4			Match 3		Match 2		Match 1	
	OTU 4 (b)	OTU 2 (w)	OTU 582 (b)	OTU 271 (w)	OTU 179 (b)	OTU 217 (w)	OTU 179 (w)	OTU 288 (b)	OTU 1545 (w)	OTU 56 (b)	OTU 271 (w)	OTU 19 (b)	OTU 1196 (w)
8	36	0	0	0	0	0	0	0	0	9	0	139	0
8	37		0		0			0		7		108	
32	31	13	0	21	1	18	76	0	0	6	21	195	6
32	151		0		12			0		11		90	
36	10	60	0	26	0	53	49	0	1	3	26	5	4
36	4		0		0			0		2		8	
37	2	20	0	11	0	32	46	0	0	1	11	70	11
37	0		0		0			0		1		48	
39	299	19	0	193	0	19	230	0	1	84	193	0	12
39	65		0		0			0		17		0	
39	91		0		0			0		29		0	
40	0	0	0	3	0	8	2	0	0	0	3	64	0
41	373	1	0	1	0	5	1	0	1	22	1	375	0
41	120		0		0			0		29		284	
42	8	2	0	0	0	0	1	0	0	0	0	181	0
42	6		0		0			0		3		83	
43	2	11	0	2	0	2	0	0	0	0	2	66	0
43	3		0		0			0		1		63	
44	1110	0	0	0	0	1	0	0	0	2	0	147	0
44	101		0		0			0		6		154	
44	102		0		2			0		2		178	
47	488	3	0	17	0	30	26	0	3	0	17	47	21
47	203		0		0			0		0		18	
48	650	5	12	0	0	0	0	0	0	0	0	120	1

48	921		9		0			0		2		36	
48	768		8		0			0		0		117	
49	0	0	0	0	1	0	0	0	0	42	0	1	0
49	2		0		0			0		47		0	
55	2	3	0	16	0	19	75	0	0	10	16	69	8
55	1		0		1			0		6		71	
55	1		0		2			0		14		67	
58	176	1	0	8	1	18	1	0	0	6	8	7	1
58	197		0		0			0		0		19	
59	387	3	0	25	0	6	226	0	1	9	25	1	4
59	408		1		0			0		2		2	
59	449		0		0			0		11		0	
59	709		0		0			0		3		2	
60	11	10	0	0	0	2	6	0	0	3	0	64	0
60	5		1		0			0		10		45	
60	8		0		0			0		3		72	
61	31	100	0	12	0	20	12	0	4	9	12	9	3
62	0	0	0	27	0	59	43	0	7	0	27	26	5
62	0		0		0			2		0		14	
63	16	15	0	54	0	21	44	0	1	14	54	4	2
63	8		0		0			0		15		3	
64	21	0	0	0	0	0	0	0	0	21	0	2	0
64	27		0		2			0		23		4	
64	33		0		4			0		30		4	
65	1	0	4	0	0	0	0	0	0	6	0	1	0
66	2198	0	0	0	0	0	0	0	0	0	0	0	0
66	1997		0		0			0		0		5	
67	38	5	0		0	61	47	0	0	0	17	6	5
67	103		0	17	0			0		0		0	
69	0	0	0	3	0	0	8	0	1	0	3	42	0
70	27	7	0	3	0	3	20	0	0	0	3	20	0

70	20		0		0			0		0		38	
71	262	0	0	0	4	0	0	0	0	5	0	3	0
71	400		0		5			0		2		1	
73	0	1	0	0	0	0	0	0	0	0	0	66	0
73	1		0		0			0		0		62	
75	12	15	0	20	0	168	31	0	0	6	20	25	37
75	3		0		0			0		0		22	
76	354	9	0	23	0	20	70	0	0	59	23	126	5
76	614		1		0			0		71		165	
76	1288		0		0			0		297		42	
77	1	0	0	0	0	0	0	0	0	6	0	90	0
77	2		0		0			0		10		47	
78	2	0	1	0	0	0	0	0	0	40	0	9	0
78	2		6		0			0		34		12	
78	5		7		0			0		51		16	
79	13	64	0	15	0	101	32	0	3	57	15	22	5
79	12		0		0			0		63		14	
81	5	6	1	12	1	24	16	0	1	31	12	4	13
81	5		2		1			0		31		4	
81	4		4		1			0		28		4	
83	0	0	0	0	0	0	0	0	0	0	0	26	0
83	0		0		1			0		1		18	
83	2		0		3			0		2		35	
85	0	0	0	0	0	0	0	0	0	4	0	10	0
86	440	0	0	0	0	0	1	0	0	0	0	0	0
86	238		0		0			0		0		0	
86	157		0		0			0		0		0	
87	0	0	0	0	0	0	0	0	0	0	0	0	0
87	0		0		0			0		0		0	
88	0	1	0	0	1	0	0	0	0	1	0	16	0

90	1	0	5	0	0	0	0	0	0	5	0	3	0
90	0		0		0			0		3		2	
90	0		0		0			0		2		6	
93	463	0	0	0	0	0	0	0	0	9	0	14	0
93	284		0		0			0		14		4	
95	63	1	0	0	0	0	0	0	0	19	0	18	0
95	83		0		0			0		8		26	
97	0	0	0	0	0	0	0	0	0	45	0	1	0
97	1		0		1			0		78		6	
98	20	0	0	0	0	0	0	0	0	17	0	1	0
98	35		0		0			0		6		2	
101	0	0	0	0	0	0	0	0	0	0	0	0	0
101	0		0		0			0		0		1	
102	1	1	0	0	0	0	0	0	0	6	0	3	0
102	3		0		0			0		4		2	
102	5		0		0			0		5		12	
103	0	0	39	0	9	0	0	0	0	0	0	164	0
103	1		15		11			0		1		99	
104	0	0	0	0	0	0	0	0	0	10	0	22	0
104	0		0		0			0		7		19	
104	1		0		1			0		7		23	
105	0	0	0	0	0	0	0	0	0	35	0	3	0
105	0		0		0			0		12		8	
105	1		0		0			0		21		11	
106	0	1	0	0	0	0	0	0	0	6	0	14	0
106	0		0		0			0		3		18	
106	0		0		1			0		1		14	
107	207	0	0	2	0	5	0	0	0	0	2	156	0
107	251		0		0			0		0		134	
107	207		0		0			0		0		155	
107	226		0		0			2		0		100	

108	1	0	1	0	0	0	0	1	0	2	0	53	0
108	0		0		0			0		2		42	
109	146	0	0	0	0	0	0	0	0	3	0	55	0
110	1	0	0	1	0	0	0	0	0	19	1	70	0
110	1		3		0			4		9		40	
110	3		0		0			0		14		37	
112	133	0	0	0	0	0	0	0	0	0	0	0	0
112	230		0		2			0		0		0	
112	255		0		1			0		0		1	
113	0	0	0	0	0	0	0	0	0	7	0	32	0
113	1		1		0			0		6		19	
113	1		0		0			0		5		45	
116	4	0	0	0	13	0	0	0	0	1	0	9	0
116	12		0		24			0		2		10	
116	42		0		12			0		0		5	
116	5		0		12			0		0		20	
117	3	0	24	0	0	0	0	0	0	0	0	0	0
117	13		18		0			0		0		3	
117	5		7		0			0		0		0	
120	0	0	0	0	0	0	0	0	0	0	0	7	0
120	0		0		0			0		0		8	
122	21	2	0	0	0	0	0	0	0	11	0	114	0
122	45		0		0			0		12		163	
125	677	0	28	0	4	0	0	0	0	0	0	22	0
125	712		4		0			0		1		38	
132	98	3	0	0	0	0	0	8	0	16	0	15	0
132	101		0		0			11		17		8	
132	136		0		0			10		21		16	
1009	0	38	1	33	0	115	20	0	0	5	33	23	7
1009	1		0		0			0		2		37	

1009	1		0		0			0		2		32	
1010	22	3	0	0	8	0	0	0	0	9	0	19	0
1010	19		0		8			0		10		13	
1010	15		0		11			0		3		19	
1011	0	26	0	19	0	32	18	0	6	0	19	1	3
1011	0		0		0			0		0		0	
1011	3		0		0			0		0		0	
1012	141	0	0	0	0	0	0	0	0	0	0	73	0
1012	210		0		0			0		0		31	
1012	397		0		0			0		0		14	
1013	0	0	0	0	1	0	0	0	0	0	0	4	0
1013	0		0		1			0		0		8	
1013	0		0		2			0		0		7	
1014	2	5	0	5	3	77	1	0	0	14	5	11	2
1014	0		1		7			0		27		22	
1014	0		0		3			0		16		19	
1015	24	8	0	97	0	0	13	0	0	0	97	402	0
1015	4		0		0			0		0		364	
1015	18		0		0			0		0		366	
1017	1	0	0	0	0	0	0	0	0	7	0	35	0
1017	0		2		0			0		3		30	
1018	88	1	0	9	0	8	30	0	0	6	9	14	0
1018	108		0		0			0		1		10	
1018	205		0		0			0		10		19	
1021	1921	0	0	0	0	0	0	0	0	0	0	4	0
2005	91	5	0	3	0	16	13	0	0	0	3	0	1
2005	77		0		0			0		0		0	
2005	77		0		0			0		0		0	
2006	1	4	0	4	0	9	0	0	0	0	4	582	1
2006	0		0		0			0		0		249	
2006	0		0		0			0		0		360	

2007	1	4	0	3	0	22	4	0	0	9	3	112	0
2007	1		0		0			0		12		61	
2007	0		0		0			0		20		69	
2009	64	0	24	0	0	0	0	0	0	43	0	1	0
2009	92		3		0			0		15		0	
2009	48		18		0			0		37		2	
2011	22	10	0	13	1	6	51	0	0	0	13	242	4
2011	30		0		0			0		0		174	
2011	21		0		0			0		0		135	
2012	43	31	11	10	0	78	10	0	1	7	10	117	3
2012	95		6		0			0		5		179	
2012	48		5		0			0		17		26	
2012	72		0		0			0		8		10	
2013	6	18	0	8	0	98	17	0	0	12	8	118	2
2013	7		0		0			0		11		107	
2013	4		0		0			1		8		29	
2014	2	0	13	0	0	0	0	0	0	0	0	10	0
2014	2		16		0			1		0		9	
2015	0	0	0	0	0	0	0	0	0	3	0	79	0
2015	0		0		0			5		2		56	
2016	0	0	0	0	0	0	0	0	0	21	0	52	0
2016	0		0		1			0		16		42	
2017	227	1	0	1	0	1	1	0	0	15	1	0	0
2017	8		0		0			0		59		0	
2020	0	0	0	0	0	0	0	0	0	2	0	69	0
2020	0		0		0			1		1		68	
2020	0		0		0			0		0		44	
2021	97	1	0	0	0	0	0	0	0	5	0	16	0
2021	105		0		0			0		8		23	
2021	162		0		0			0		7		27	

2021	137		0		0			0		3		18	
2022	8	0	0	0	1	0	0	0	0	7	0	1	0
2022	12		0		1			0		6		8	
2022	0		0		1			0		35		2	
2023	0	0	0	0	1	0	0	0	0	20	0	69	0
2023	0		0		2			0		29		39	
2024	2	0	2	0	7	1	0	0	0	27	0	1	0
2024	12		0		2			0		21		1	
2024	7		2		3			0		33		1	
5001	0	7	0	7	0	0	0	0	0	12	7	272	0
5001	0		0		0			0		8		303	
5001	1		1		0			0		13		323	
5004	1	0	0	0	0	0	0	1	0	0	0	1	0
5004	0		0		0			0		0		1	
5004	0		0		0			0		0		0	
5005	0	1	1	6	0	40	6	0	0	3	6	0	4
5005	0		2		0			0		0		3	
5005	0		5		0			0		7		5	
5008	11	1	0	0	4	0	0	44	0	34	0	64	0
5008	19		0		3			8		18		25	
5008	13		0		0			12		19		62	
5009	1	1	0	0	0	0	0	1	0	67	0	1	0
5009	590		0		0			0		0		84	
5010	480	0	0	0	0	0	0	0	0	0	0	33	0
5010	802		4		1			0		0		19	
5011	0	0	0	0	0	0	0	0	0	2	0	13	0
5011	0		0		0			0		0		26	
5013	212	0	0	0	0	0	0	0	0	5	0	0	0
5013	246		0		0			0		9		0	
5013	254		0		0			0		6		0	
6001	718	0	0	0	1	2	0	0	0	35	0	5	0

6001	209		0		1			0		9		19	
6001	265		0		1			0		26		18	
6002	2	6	0	0	3	0	0	0	0	40	0	1	0
6002	3		0		4			0		27		0	
6003	2	0	0	0	4	0	0	0	0	10	0	7	0
6003	0		0		9			0		4		11	
6005	238	79	0	5	0	39	12	0	3	0	5	3	1
6005	463		0		0			0		6		4	
6005	168		0		0			0		12		0	
6005	38		0		0			0		5		2	
6006	10	2	0	7	0	4	7	0	0	15	7	43	1
6006	18		1		0			0		9		12	
6006	8		0		0			0		5		18	
6008	62	0	0	0	0	0	0	0	0	0	0	0	0
6008	104		0		0			0		0		0	
6009	0	2	5	4	0	4	0	0	0	12	4	0	0
6009	2		14		0			0		32		0	
6011	50	55	0	5	0	25	7	0	0	0	5	11	1
6011	1296		0		0			0		0		3	
6014	2	0.0	0	0	0	0	0	0	0	13	0	7	0
6014	0		0		0			0		6		9	
6014	2		0		0			0		36		10	
6015	23	0.0	1	0	0	0	0	0	0	53	0	126	0
6015	46		1		0			0		18		76	
6016	0	0.0	0	0	0	0	0	0	0	7	0	42	0
6017	835	0.0	0	0	18	0	0	0	0	28	0	3	0
6017	1069		0		29			0		19		2	
6017	1439		0		8			0		21		1	
6017	742		0		34			0		41		1	
6018	2	0.0	0	0	0	0	0	0	0	30	0	0	0

6019	86	0.0	0	0	0	0	0	0	0	8	0	0	0
6019	382		0		0			0		3		0	
6020	1689	0.0	0	0	0	0	0	0	0	0	0	0	0
6022	188	1.0	0	0	0	0	0	0	0	42	0	0	0
6022	225		0		0			0		122		0	
6023	95	0.0	0	0	0	0	0	0	0	89	0	7	0
6023	95		0		0			0		79		11	
6023	174		1		0			0		46		0	

Appendix K – Top 50 water OTUs identified as match by Jukes-Cantor

Table 11: shows the fifty most prevalent water OTUs identified as match by Jukes-Cantor, its biopsy match, and their relative prevalence in percent.

Prevalence‡ %	**Biopsy OTU**	**Water OTU**	**Prevalence† %**	**Taxonomy***
0,0040	OTU_710	OTU_6	3,8856	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Polaromonas;s__
0,0061	OTU_778	OTU_24	2,1040	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__;s__
0,0013	OTU_461	OTU_112	1,0614	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__;s__
0,0011	OTU_623	OTU_22	0,9817	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__;s__
0,0490	OTU_145	OTU_40	0,6753	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
0,0006	OTU_891	OTU_9	0,6245	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas;s__yabuuchiae
0,2944	OTU_45	OTU_87	0,5766	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
1,5470	OTU_12	OTU_210	0,5626	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
0,1402	OTU_72	OTU_133	0,5457	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila
0,0013	OTU_461	OTU_13	0,5456	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax
0,2046	OTU_32	OTU_244	0,5248	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__

0,0800	OTU_164	OTU_184	0,4972	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__longum
0,1208	OTU_203	OTU_120	0,4937	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas;s__
0,0381	OTU_80	OTU_51	0,4527	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
0,0717	OTU_251	OTU_8	0,4089	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
0,2732	OTU_62	OTU_185	0,3869	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
0,5938	OTU_99	OTU_230	0,3755	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
0,0013	OTU_461	OTU_99	0,3729	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__;s__
0,2046	OTU_32	OTU_193	0,3650	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
0,0717	OTU_251	OTU_103	0,3241	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__
0,0381	OTU_80	OTU_212	0,3105	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
0,0390	OTU_179	OTU_179	0,2802	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__
0,0013	OTU_461	OTU_1701	0,2774	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__;s__
0,0037	OTU_553	OTU_217	0,2753	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis
1,7485	OTU_15	OTU_254	0,2487	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__
0,0667	OTU_98	OTU_173	0,2464	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
1,2215	OTU_14	OTU_140	0,2201	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
1,0743	OTU_51	OTU_480	0,2119	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
0,0465	OTU_197	OTU_299	0,2059	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
0,0126	OTU_515	OTU_157	0,2056	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__;s__
5,3543	OTU_135	OTU_495	0,1968	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium;s__prausnitzii
0,0016	OTU_647	OTU_124	0,1960	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__;s__
0,0395	OTU_198	OTU_331	0,1816	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
0,4590	OTU_56	OTU_271	0,1787	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__

0,2334	OTU_84	OTU_231	0,1777	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
0,0643	OTU_59	OTU_370	0,1752	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
0,0490	OTU_145	OTU_282	0,1656	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
0,0154	OTU_316	OTU_324	0,1579	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
1,4141	OTU_16	OTU_601	0,1554	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__ovatus
0,0037	OTU_289	OTU_232	0,1551	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__
0,0061	OTU_778	OTU_477	0,1523	k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae
5,5465	OTU_4	OTU_2	0,1513	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__
0,2240	OTU_13	OTU_521	0,1386	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
0,7706	OTU_40	OTU_361	0,1261	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae
2,2424	OTU_5	OTU_487	0,1170	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus];s__gnavus
0,0013	OTU_461	OTU_386	0,1163	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax
0,2541	OTU_17	OTU_247	0,1163	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
1,5470	OTU_12	OTU_597	0,1139	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia;s__
1,1711	OTU_8	OTU_716	0,1032	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
0,0006	OTU_863	OTU_190	0,1000	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas;s__

* k = kingdom, p = phyla, c = class, f = family, g = genus, s = species

† = prevalence water OTU

‡ = prevalence biopsy OUT

Appendix L: Results of ASCA-ANOVA analysis

Table 12: Shows the output of the ASCA-ANOVA analysis performed on different groups from the biopsy data set.

Terms	Principal components	P-value
Variable 1 (Gut part)	2	1
Variable 2 (GI localization)	6	1
Variable 3 (Inflammation category)	2	0,2518
Variable 4 (Diagnosis)	5	0,0002
Variable 5 (Age)	2	0,0381
Var 1 x Var 2	5	1
Var 1 x Var 3	8	1
Var 1 x Var 4	15	1
Var 1 x Var 5	6	1
Var 2 x Var 3	19	1
Var 2 x Var 4	20	1
Var 2 x Var 5	14	1
Var 3 x Var 4	17	0,6186
Var 3 x Var 5	6	0,5698
Var 4 x Var 5	10	0,0001

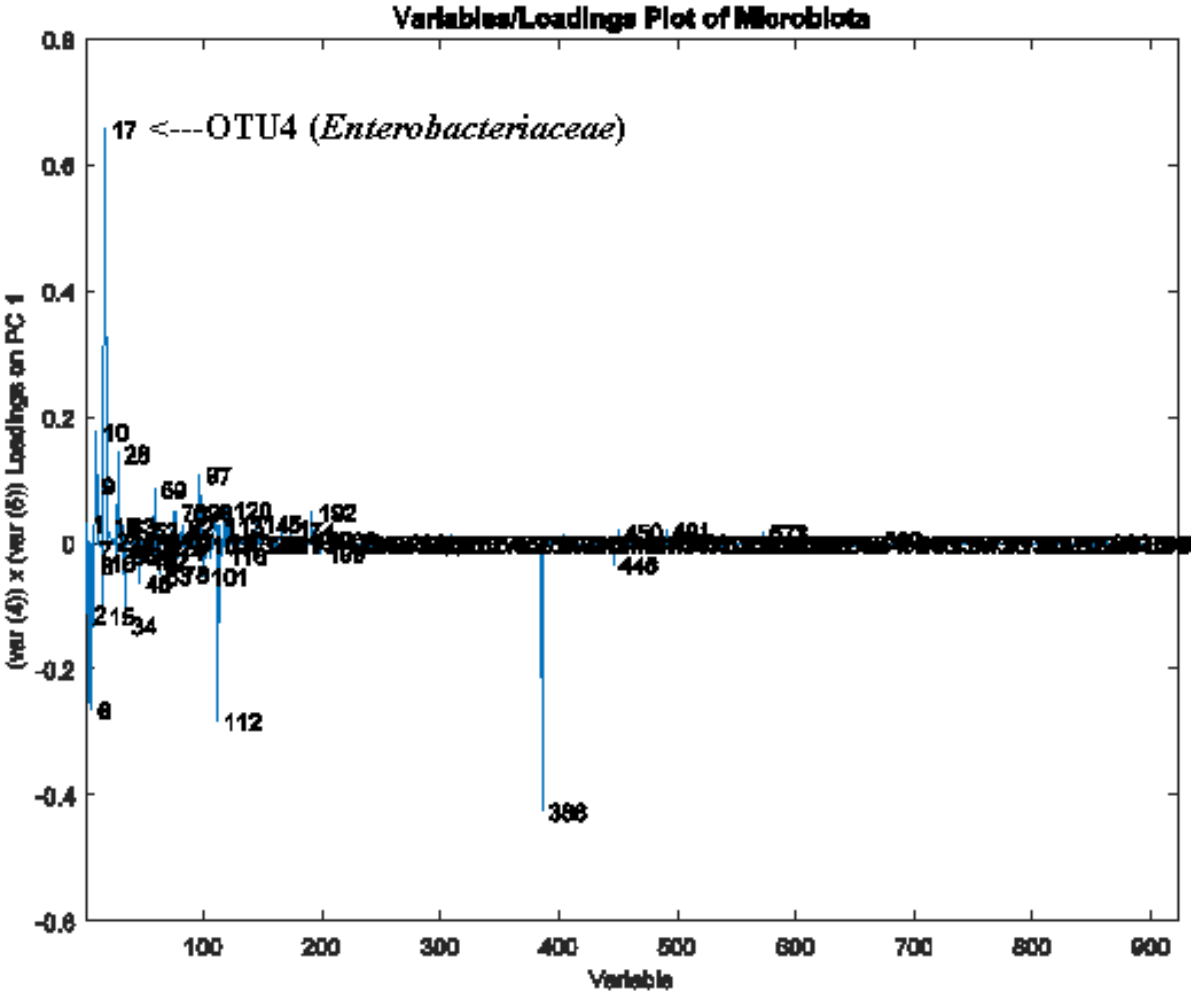
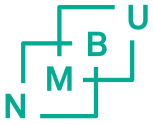


Figure 1: The figure shows loading plot and loading scores of variables potentially explaining variations between age and diagnosis, as tested in ASCA ANOVA.



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway