Finding Patterns in Living Networks

Community analysis of the human gut microbiome

Diplomarbeit

zur Erlangung des akademischen Grades eines

Diplom-Ingenieurs

Studienrichtung: Biotechnologie

Verfasser:

Stefan Lackner

Technische Universität Graz

Erzherzog-Johann-Universität

Fakultät für Technische Chemie, Verfahrenstechnik und Biotechnologie

11/2011

Betreuer:

Univ.-Prof. Dipl.-Biol. Dr.rer.nat. Gabriele Berg

Insitute for Environmnetal Biotechnology (TU Graz)

Ass.Prof. Dr.med.univ. Gregor Gorkiewicz

Institute of Pathology, Medical University of Graz

TU Graz
Graz University of Technology
Senat

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am 02.11.2011 .....

.....
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

2nd of November, 2011
.....
date

.....
(signature)

Contents

Abstract 3
1. Introduction 4
2. Methodology 6
2.1 Pyrosequencing 6
2.2 Data Analysis via Qiime 7
2.2.1 Qiime Workflow I - from FASTA to OTU tables 8
2.2.2 Qiime Workflow II – Diversity measurements 9
2.2.3 Qiime Workflow III – Distance Computation: Verification 11
2.3 Development of a Scoring System 12
2.4 Cytoscape and Profile Clustering 14
2.4.1 Limitations of Common Data Representation Methods 14
2.4.2 Development of a Profile Clustering System 15
3. Bacterial community changes in diarrhea 18
3.1 Materials and Methods 18
3.1.1 Study Setup and Sampling 18
3.1.2 DNA Extraction, Amplification and Quality Measurements 19
3.1.3 Pyrosequencing 19
3.1.4 Data Analysis 20
3.2 Results 21
3.2.1 General comparisons 21
3.2.2 Profile Clustering 23
3.3 Discussion 28
4. Bacterial community changes in Enterocolitis 30
4.1 Materials and Methods 30
4.1.1 Study Setup and Sampling 30
4.1.2 DNA Extraction, Amplification and Sequencing 31
4.1.3 Data Analysis 32
4.2 Results 34
4.2.1 Histological assessments 34
4.2.2 Qiime Analysis 35
4.2.3 Network Analysis 37
4.3 Discussion 42
5. Concluding remarks 48
References 50
Acknowledgements 53

Abstract

The human microbiome is a vast source of genetic functionality and contributes to both, health promoting and pathological events in the human body. Trillions of bacteria associated with nearly all organ systems inhabit the human body and outnumber human cells by a factor of ten. The role of this microbial flora in diseases like asthma and chronic inflammation or body states such as obesity, mental health and gut functionality has been a rapidly expanding field in life sciences. To get access to this interconnected system of bacterial species, next generation sequencing technologies have been developed. In this work, we examine the structure and changes of the human gut microbiome via pyrosequencing and subsequent bioinformatics analysis of the generated data. To overcome limitations of statistical methods, a scoring mechanism was created to identify phylotypes of interest that could be associated with physiological states. This was further developed into a network analysis routine, where a bacterial community can be viewed depending on the abundance changes of its members. Two projects have been carried out as a base for further experiments, one being the artificial induction of diarrhea, the other the description of the microbiome in a severe form of enterocolitis. In the first, we found that diarrhea, a commonly occurring misbalance of the gut system, can lead to thorough changes of the human gut microbial community. Washing out events of Bacteroides and other commensals are accompanied by a rise of Proteobacteria and a general reduction of diversity. We could also show that reconstitution occurs after an acute diarrhea phase, when the bacterial system regenerates the pre-diarrhea composition in a patient specific way. We speculate that these changes could serve as starting points for more severe misbalanced states like chronic inflammations of the gut. The second project was the investigation of nongranulomatous chronic idiopathic enterocolitis, a severe inflammation of the small bowel with high mortality rates and elaborate treatment requirements. Bacterial community structure was again determined by pyrosequencing, statistical data analysis and a network approach. We observed washing out events similar to the diarrhea experiment and an accompanying rise of several phylotypes usually associated with the human oral cavity. Therefore, we propose a translocation genesis for NCIE, where species usually associated with oral biofilm formation enter the gut system either by swallowing or blood transfer and form bridgeheads of colonization for a misbalanced microbial community. Subsequent local inflammation events and nutrition changes then lead to further community reshaping and chronic inflammation in typical patches. These changes are commonly known as “dysbiosis”: when the bacterial and the host interactions enter a mode of misbalance, human physiology can change drastically. These findings underscore the importance of examining the human microbiome as part of an intertwined network, where a plethora of parameters and triggers affect the outcome of bacterial-host interactions.

1. Introduction

The human body can be considered a complex ecosystem with a variety of habitats. Each habitat represents a unique niche with special nutrient sources and growth requirements (e.g oxygen tension, pH, etc.) for the bacteria. It has become evident in the last years that this community of microbes is not isolated from human body function but an integral part of our physiology. This is strikingly shown by the fact that the bacterial cells associated with humans outnumber human cells by at least a factor of 10 [1,2] and that there are a hundred times more genes in the microbial metagenome than in the human genome. Therefore it is speculated that the existence of the human microbiome is a reason for the relatively small number of genes found in the human genome [3]. Indeed these bacteria can fulfill a large number of tasks that are not encoded in human DNA.

Of the 100 trillion [4] bacteria associated with humans, most live within the gut [2]. The gut microbiota has been identified to facilitate a large number of tasks important to human physiology. It is capable of degrading otherwise indigestible carbohydrates (glycans), therefore contributing to energy metabolism [1] and host fat storage [5]. Gut bacteria are able to release hormone-like substances and lipids signals conferring a cross talk with the host, stimulate growth of epithelial cells and, very importantly, contribute to the formation of a functional immune system [2]. These traits have a significant impact on host physiology. Their lack can be observed in germ free mice which show dysfunctionality in various organs and tissues and a significant reduced vascularization of the gut tissue. Without these microbe-host interactions, the gastrointestinal system is immature, which hamperes nutritional processing. Interestingly, these mice also show behavioral changes, which indicates that the microbiota could indeed influence our mood and, to some extent, our personalities [6] [7].

Microbial communities of different body parts are not homogenous in their structure and physiological capabilities. When comparing the microbiota of e.g. the oral cavity, the gut or the skin, it became obvious, that a large amount of variation can be found within these habitats [4]. The effect of different body niches is greater than variation due to interpersonal difference, sex or time [4]. Interpersonal variation of the microbiota is far higher than intrapersonal difference in one habitat, even over a long period of time [4]. Large differences in community structure also occur when comparing different micro-habitats, e.g. luminal, mucosal and epithelial areas [5]. A gut core community, a commonly shared microbiota amongst all humans, is not evident on a species or genus level [8]. It was however shown, that a gene level core microbiome exists [9], indicating that a presence of genes and gene-families is more important to fulfill the required functions than membership to a certain taxon.

The structure of the microbiome in healthy and diseased individuals and its participation in beneficial and adverse events has been a highly active area of research in the last few years. It has become clear that patchworks of bacterial communities inhabit the human body and

that these habitats are often interconnected. Disturbances in this intertwined system can lead to severe imbalances of host-bacterial interactions and can result in a plethora of diseases. Diarrhea [10], obesity [9] and chronic inflammatory diseases [3] were identified to be associated with an altered microbiome structure.

In western societies, a very high standard of hygiene has been reached in the last couple of decades. A desirable process per se, it could have long lasting effects on the composition of our microbiota. While many infectious diseases have been virtually wiped out, the rise of chronic and autoimmune diseases is obvious and problematic.

For this work, the human gut microbiome was investigated in both its healthy and misbalanced states. We identified changes and species of interest during diarrhea and applied that knowledge to characterize a severe form of enterocolitis.

2. Methodology

Traditional methods of microbiology, such as isolation, cultivation as well as subsequent morphologic and functional characterization have been used for decades to shed light into the complex system of human associated bacteria. Ultimately though, there is a limit to this approach. Cultivation methods fail to grow a large number of bacteria, preventing a thorough investigation of the whole bacteria-human system. It has become a well-established fact that approximately only 10% of bacteria found in the environment can be cultivated under laboratory conditions. To circumvent these limitations, culure-independent approaches of microbial systems have been developed that enable the amplification and analysis of microbial marker genes (e.g. 16S rDNA) with next-generation sequencing technologies.

2.1 Pyrosequencing

The technique used in the work at hand, known as pyrosequencing, is a member of a new group of genomic analysis methods, commonly known as NGS or next generation sequencing. Compared to traditional Sanger sequencing, a much higher number of sequences can be generated in one run, thus allowing the researcher to investigate a community of bacteria in a cost efficient manner. The methodology was developed in 1996 by Pål Nyrén and Mostafa Ronaghi at the KTH in Stockholm [11].

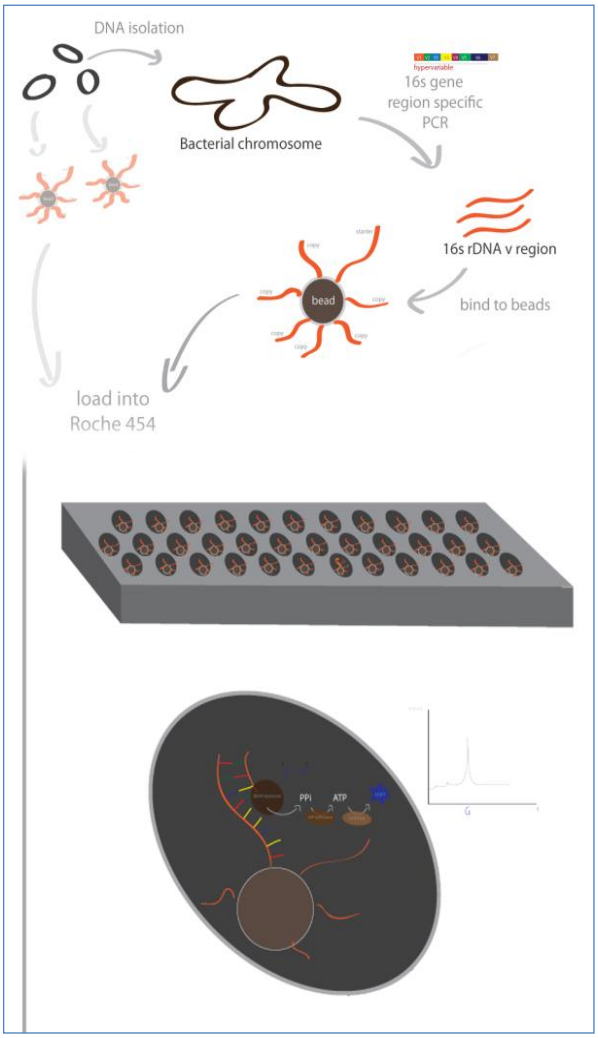

Figure 1: Basic workflow of pyrosequencing

Prior to a pyrosequencing run, the DNA has to be isolated from bacteria. Subsequently, a microbial marker gene suitable for community analysis is amplified via PCR. One such region is the 16S gene for the bacterial ribosomal RNA. It features a variety of hyper variable parts (v1-v9)

specific for microbial taxa, enabling unambiguous identification of bacteria and conserved regions targeted by pan-bacterial primers allowing amplification of the majority of microbes. The 16S rRNA gene has been used as a suitable region for high throughput pyrosequencing in a large number of bacterial community studies so far [12]. The oligonucleotide primers used for PCR fragments harbor a linker (called MID) that allows for the later identification of the sequence as a member of a specific sample. This way, a variety of samples can be analyzed in one sequencing run. At this point, the amplified DNA sequences consist of the 16S gene region, the primer sequence, a linker and the barcode. After DNA amplification, sequences get transferred to beads, where they bind to the surface and get amplified in an enrichment step. Subsequently, the actual pyrosequencing starts, where a deoxynucleotide triphosphate (A, T, G or C) gets supplied to a DNA polymerase, which complements the template sequence. If the base fits to the complement strand, it is incorporated into the strand, a phosphor-diester bond is generated by the DNA polymerase and one pyrophosphate (PPi) is released. An ATP sulfurylase converts this compound to ATP, which in turn activates a luciferase that produces light. The resulting signal is directly proportional to the amount of integrated nucleotides, thus allowing a determination of homopolymers. If the supplied nucleotide is not complementary, it gets degraded by the fourth enzme in the system, an apyrase. Repeating circles of these reactions allow the determination of the actual sequence. The technology is currently limited to sequences of 500-800 bases in length.

2.2 Data Analysis via Qiime

A pyrosequencing device generates three kinds of files: the first is the stored sequence data in FASTA format, which includes all sequence-reads that have been fit the quality parameters. As the distributing process of the PCR fragments amongst the beads is a statistical process rather than a one on one assignment, two or more different sequences will sometimes bind to a bead and, on the other hand, some beads will carry no sequences at all. A respective read (one information unit that is derived from one well and its correspondent bead) can be judged untrustworthy and hence is eliminated from the analysis if it is found ambiguous or empty. This step is the first filtering step in the analysis and is facilitated by using the second resulting file from a pyrosequencing run: the *.qual file. The file indicates a score which is a direct measure of the quality of each sequence in the FASTA file. The third filetype is *.sff, which is a combination archive for both *.fasta and *.qual files.

2.2.1 Qiime Workflow I - from FASTA to OTU tables

The program we mainly used to analyze pyrosequencing data is called Qiime. It is an analysis tool for microbial community sequencing data and was developed by Knight et al. in the last several years [13]. Qiime uses pre – defined scripts to enable users to analyze their data. The above mentioned filtering is part of the first step in this pipeline, a script called "split_libraries". Besides eliminating ambiguous and low quality reads, it is also responsible for splitting the whole sequence data and assigning sequences to their respective samples based on the barcodes. Hence, the output of this step is a quality filtered list, wherein each sequence is assigned explicitly to one sample.

The next step in Qiime analysis is the script "pick_otus_through_otu_table.py". It consists of several subscripts which are shown in Table 1:

Table 1: The Qiime script "pick_otus_through_otu_table"

Subscript	Description
pick_otus.py	clustering sequences of a given similarity into OTUs
pick_rep_set.py	choosing a representative sequence for each cluster
align_seqs.py	aligns the representative sequences to a set of sequences
assign_taxonomy.py	assigns a taxonomy to a sequence cluster
filter_alignment.py	filters the alignments for gaps
make_phylogeny.py	builds a phylogenetic tree based on a chosen method
make_otu_table.py	creates an OTU table

The first sub-step is essential since it determines to what taxonomic level sequences are going to be analyzed. After this step, several clusters of sequences have been created, which are called OTUs (operational taxonomic units) or phylotypes. As a threshold value of any kind cannot be directly translated into classic taxonomic levels like phylum, family or genus, an OTU represents an artificial system in which sequences are clustered by their similarity. We used a sequence similarity of 97% for clustering, which is higher than the genus level requirement of 94% one can find in the Qiime documentation [14]. Our analysis can therefore be considered to work on species level, as methodological studies suggest [15].

Out of these clusters of sequences one is chosen to be the "representative" sequence. The default method for doing this is searching for the most abundant sequence in an OTU.

Aligning the representative sequences to a template alignment is important for following steps as it correlates all OTUs amongst each other. By default, a PyNAST method is chosen here, where a set of pre-aligned sequences (Greengenes - http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) is used as the template. As this alignment creates gaps in the query

sequences, a subsequent filtering step “filter_alignment” is implemented to eliminate gaps which are shared amongst all sequences.

“Assign_taxonomy” searches NCBI (National Center for Biotechnology Information) or RDP (Ribosomal Database Project) databases for matching sequences to the generated query ones. In the default method, RDP classification, small fragments of sequences are compared to pre-aligned sequences of the RDP database and subsequently assigned to the resulting taxonomic classification. As the RDP database allows only for genus level taxonomic classification, the resulting output will be not showing species level, even if the similarity of sequences in an OTU would theoretically allow this.

With data from previous steps, a phylogenetic tree is then created by the script “make_phylogeny”. By default, a FastTree algorithm is used to accomplish this task. This again is an essential step for downstream analysis, as many diversity measurements depend on it.

The last step of “pick_otus_through_otu_table” is the actual generation of an OTU table, which lists all OTUs that were created in the analysis together with their respective reads and taxonomic classification. Together with the OTU table a statistical analysis of the pipeline run is created, where one can find information about sequence length and number of reads per sample. The information content of the OTU table is then used in other scripts (make_otu_heatmap_html.py, make_otu_network.py and summarize_taxa_through_plots.py) to allow a basic graphical representation of the data. While the first scripts create heatmaps and a basic cytoscape network, the later script sums up all reads on a given taxanomic level (e.g. family) and creates bar-charts and pie-charts as graphical representation of the data.

2.2.2 Qiime Workflow II – Diversity measurements

An essential part of any community analysis is the description of the diversity in and amongst samples. The first is called “alpha diversity” or “within – sample” diversity, as it is a direct measure of how large the variety of species in a sample is (or could be). The second group is “beta diversity” or the “diversity between samples”, which measures and illustrates the similarity or dissimilarity of samples when they are compared to each other.

To calculate alpha diversity one must first generate subparts of the main OTU table. This procedure creates smaller OTU tables of a user defined size were a given number of OTUs from the main list is chosen at random for each sub-table. This step is called subsampling or rarefaction. For each of these tables the alpha diversity is calculated by a user defined algorithm and expressed in a number as a fraction of 1. The subsamples are now fused again into one list and their data can be visualized in a “rarefaction plot” (Figure 2). Rarefaction

allows for a rough estimate of inner sample diversity. A typical pyrosequencing run creates between 3000 and 10000 reads per sample. One can imagine that, if only a small fraction of these, e.g. 100, are chosen for subsampling, one will not be able to gain insight into the whole diversity of the community. When raising this number by adding a user defined amount of sequences (which is called iteration step), a larger portion of the community is revealed, thus allowing a more in depth analysis. Essentially, one will reach a plateau, were adding additional sequences will not raise the number of observed OTUs anymore, as new reads are assigned to existing OTUs rather than creating new ones. This limit gives us a reasonable idea about the number of sequences that are needed to fully sample the community. Alpha diversity measurement is therefore an essential part in community analysis. It reveals if the number of sequences generated by a pyrosequencing run (also known as depth of the analysis) is sufficient to describe a given community as a whole. The higher the depth of a pyrosequencing run, the higher the probability of finding rare OTUs, which are only represented by a small fraction of the sequences in one run. Although one could argue, that an OTU is more essential for a community if it is represented by a large number of reads (and therefore represents a dominating taxonomic group of bacteria). Nevertheless, rare OTUs of a bacterial community should not be underestimated.

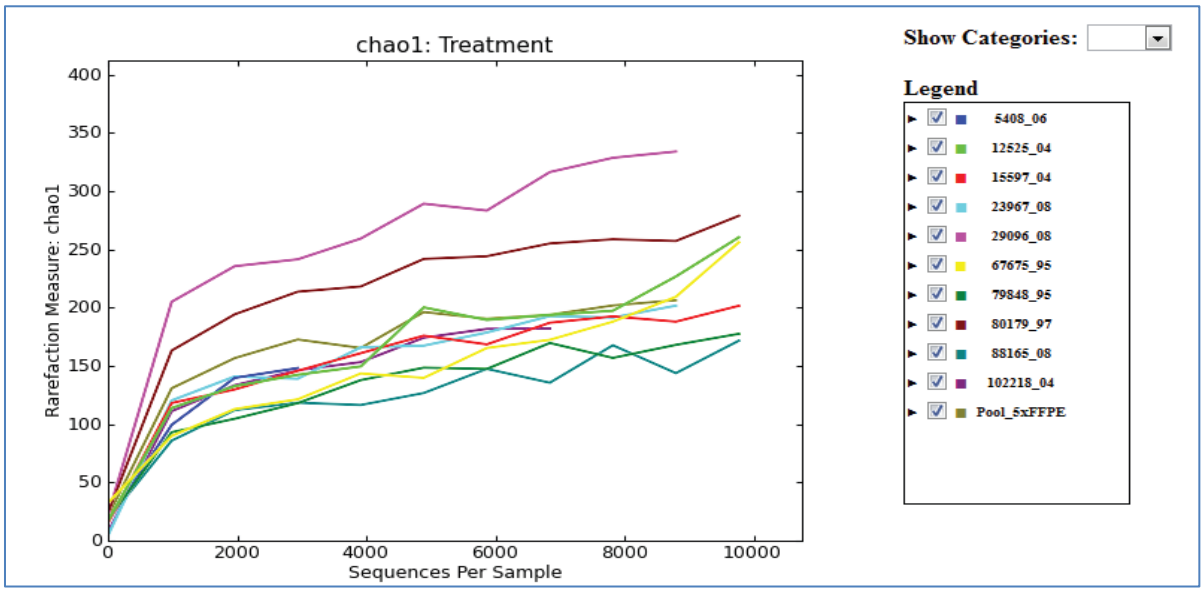

Figure 2: A typical alpha rarefaction plot. As the depth of the subsamples increases, a significant increase in OTU numbers is noticeable. Once a certain treshhold is reached (around 3000 sequences in this example) all curves reach a plateau indicating that additional reads do not lead to a significant increase in new OTUs.

As in alpha diversity, measurement of beta diversity starts with subsampling, in which single file rarefactions are produced and their beta diversity calculated separately. Subsequently, principal coordinates are generated and all samples are mapped to an orthogonal plot. Principal coordinates (PC) are distance calculation results, which give insight into the amount

of variation between samples. They are usually depicted in a 2-D or 3-D plot, where the PC with the highest variation is assigned to one axis, the PC representing a lower amount of variation to the second and the one associated to the least variation to the third axis of the diagram. In the resulting plot one can estimate the similarity of samples based on their proximity, where samples with high similarity in their communities form clusters in the three dimensional system.

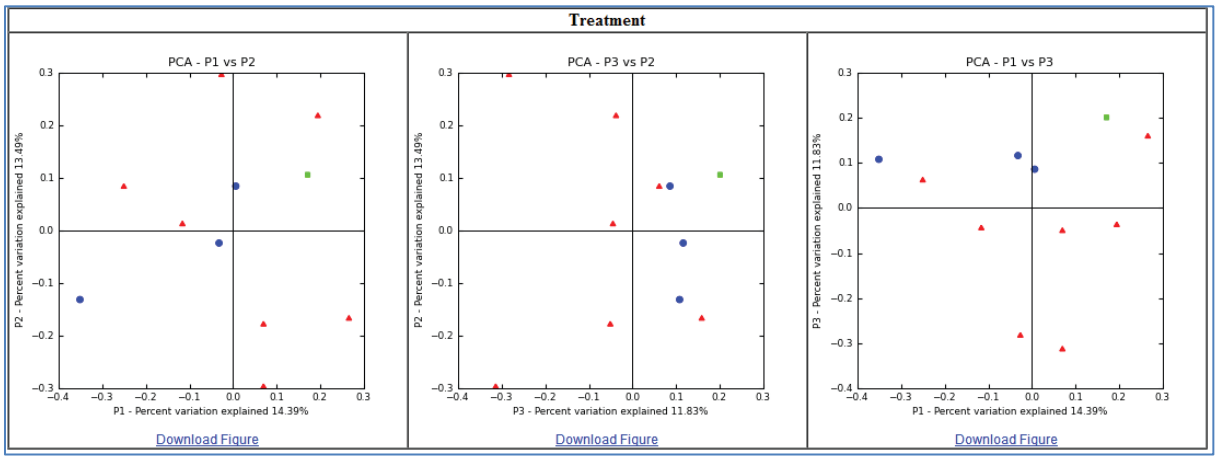

Figure 3: A typical PCA plot depicting the dissimilarity of gut samples. Principle components 1 to three account for up to 13.49% variability. The percentage number on each axis represents the amount of variation that is associated with the axis. In this figure a three dimensional PCA plot is shown in three independent 2D plots.

2.2.3 Qiime Workflow III – Distance Computation: Verification

Verification of distance computing can be carried out via jackknifing, one example of this being the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) tree. Jackknifing starts by creating a master (main) tree based on information from previous beta diversity distance values and the OTU table. Subsequently, as in beta diversity, subsampling and beta diversity distance computation leads to a number of rarefied lists. These are used to generate subsampled trees, which in turn are compared to the master tree. Branching points (or nodes) that are consistent over a high number of comparisons are identified

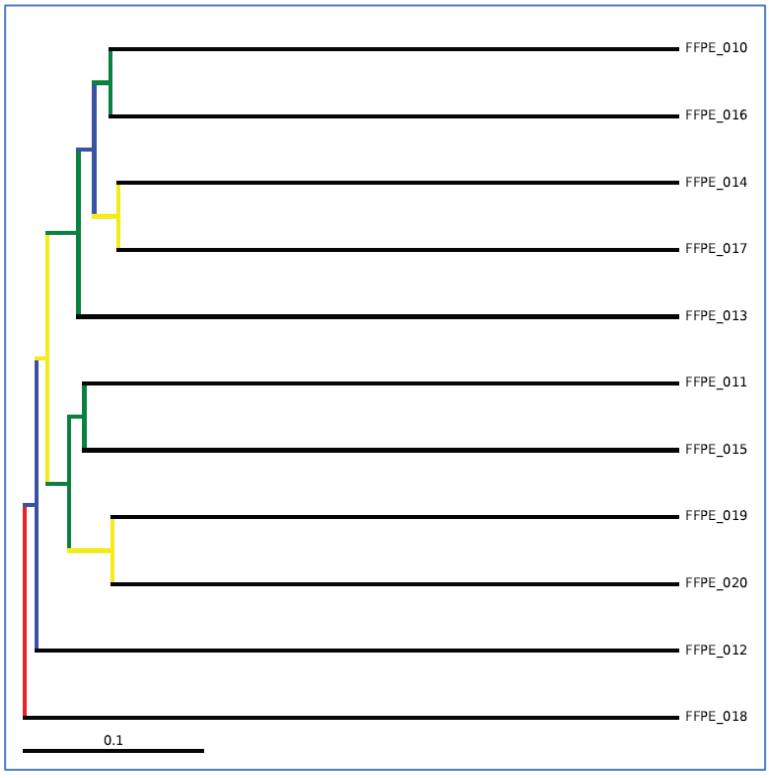

Figure 4: A typical UPGMA tree. Nodes are colored by their "trustworthiness" / consistency: highly supported nodes are shown in red (75-100% support), yellow (50-75% support), green (25-50% support) and blue (0-25% support).

as robust, while nodes that vary significantly between comparisons are characterized as non-robust.

2.3 Development of a Scoring System

The main project of this thesis was to develop a method for analysing the microbiota of a patient with a severe form of Enterocolitis. As stated in chapter 0, the bacterial communities which can be found in humans differ significantly, even between members of a family. This leads to a number of problems when analysing pyrosequencing data:

a. As the interpersonal variation between patients is high, it is difficult to find OTUs, where progressions during the illness / the experiment are similar in all patients.
b. Hence, when observing only the mean changes amongst patients, OTUs of interest often get excluded due to normalizing effects.
c. Analysing the basic sample differences (as in beta diversity measurement) gives an idea of the similarity between samples but will not produce results on e.g. species level.
d. Trying to use the advantage of bar charts which require a summarize_taxa approach, where groups are formed on a certain taxonomic level (like genus), exhibits similar problems, as it is not sure if these groups behave similar internally. A high rise in one particular *Bacteroides* OTU could thus be covered up by a high number of small declines in other *Bacteroides*.
e. Statistical analysis demand sample sizes which are high enough to function correctly. There are several cases (rare diseases, stored material with no access of gaining new samples, highly invasive sampling procedures in humans, ethical questions or financial limitations) when only a small number of samples can be sequenced and analysed.

To overcome these problems, one must circumvent the drawbacks of classical statistical approaches but still be able to find OTUs of interest. To achieve this, I designed a global scoring system, which weighs OTUs on their abundance in patients rather than producing a mean value amongst all patients.

The scoring systems input file is the main OTU table, which is generated in Qiime analysis. For explanation purposes, we will only focus on the standard case where a disease is examined first (acute state) and a late examination, when the disease is cured or latent (remission state). As by default the OTU table features absolute reads rather than portions, so a relative value is calculated for each OTU in each sample. This allows for comparing of

samples with different absolute read values. In the next step, a ratio is calculated between different time points / health states. The formula for this step is %acute / (%remission + 0.1). The addition of 0.1 serves two purposes: First, it avoids division by 0, which would yield a false value. Secondly, it prohibits small value changes (as from acute: 0.1% to remission: 0.2%) from entering the scoring boundaries. This approximately relates to 10 reads per sample (in a sample of 10 000 reads) and represents the error of PCR / pyrosequencing, which could arise and falsify the results. After this step, every OTU in each sample now has a ratio value assigned to it. Depending on these ratios, a local score is formed: it counts the number of times, an OTU reaches a certain threshold ratio (e.g. above 1.5) when looking at all samples. In the easiest system of two patients with two states only three values can be the result of this step: If a ratio > 1.5 between acute and remission is found in both, the local score is 2, if it is found in one, the score is 1 and 0 if an OTU does not reach the given ratio in any patient. The last step in this system is the introduction of a global score. It is calculated with the formula:

$$Global\ Score = (\sum local\ scores) \times (Number\ of\ Patients\ in\ whom\ an\ OTU\ occurs\ with\ a\ ratio > 1{,}5)$$

The sum of local scores by itself is a direct measure of how many times an OTU showed a significant increase. As it is quite unlikely, that the bacterial communities of two patients show a similar behaviour, this event should be emphasised when it actually occurs. This is facilitated by the introduction of a multiplication by the number of times a significant OTU ratio change is detected in more than one patient. Hence, two OTUs that show the same local score sum over two patients can have a different global score, where the OTU that shows the increase in more patients scores higher.

The introduction of this global scoring system simplifies the identification of genera that feature a user defined ratio change. The OTU of interest does not have to behave the same way in all patients or all conditions but could still score high if certain criteria (several local scores > threshold, found in more than one patient) are met. This approach also circumvents the limitations of statistics when it comes to small sample sizes. This is essential for cases like our experiments, where due to the rare nature of the disease or invasive nature of the sampling method (gut biopsies in diarrhea) only a small number of samples are available.

After a global score has been determined for each OTU, the correspondent sequence can be blasted via NCBI blast tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) or similar services, most notably RDP SeqMatch (http://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp) [16].

According to the nature of the setup, one can also choose different ratios or ratios < 1 to find OTUs which decrease over the course of the experiment.

An alternate way to use the scoring system is to not use fold changes (e.g. between acute and remission states) but “changing patterns”, like increasing or decreasing to create local

scores. The subsequent steps are then the same as described above. The diarrhea study discussed in this work is an example for this approach, which we called “profile clustering”.

2.4 Cytoscape and Profile Clustering

The next methodological task was the development of an OTU based network analysis which allows the identification of bacterial groups which behave in certain patterns over the course of the experiment. We developed two network types, which could handle data from different experimental setups. The first one is applicable for two point setups, where these two points can either represent time, disease state, treatment type or any other groups defined by the researcher. The second network is similar to the first but the fact that it allows OTU progression identification in experiments with more than two sampling points.

2.4.1 Limitations of Common Data Representation Methods

The basic purpose of profile clustering is to find bacteria of interest on OTU level, without losing biologically important information about the interpersonal variety of the system and representing their progression in a graphical manner. It became clear early in the development process, that pie charts or bar charts are not suitable for this task due to the high number of OTUs which arise from this kind of analysis. One can plot a small number of OTUs on a classical diagram with the time points on x-axis and the relative abundance in a sample on the y-axis, but as with bar charts, the number of OTUs one can follow in one picture is low, especially if the progressions differ amongst patients. The first step was to take lists generated with our global scoring system and map their respective relative abundance on diagrams as described above. After comparing a number of high scoring bacterial OTUs it was obvious, that reaction patterns to an impetus differed greatly amongst patients. This is attributed to the high inter-personal variation within the individual microbiomes and leads to the assumption that disease patterns or shared symptoms are not necessarily attributed to the same changes in the bacterial community of each patient. Examples of this can be found in chapter 4.2. There was, however, an interesting finding in all bar charts, namely, that progressions showed similar reaction patterns when comparing different OTUs. Our first approach of developing an OTU based identification system therefore led to statistical graphical analysis, where curves can be fitted onto standard curves with a certain degree of variation. One could therefor design a system, in which OTUs are mapped to certain model progressions. Unfortunately, the scarcity of suitable material for our experiments prohibited that approach due to the low number of samples we obtained.

2.4.2 Development of a Profile Clustering System

These limitations led to a more straightforward approach for our first network experiment, the bacterial community changes in diarrhea. Here, three time points had to be compared in four patients.

In our model study, three time points related to the stressor (diarrhea) can be defined, pre-diarrhea, diarrhea and post-diarrhea. Therefore, an OTU can only feature three progressions when comparing two points. It can either increase, decrease or stay within certain limits of abundance changes which define the state "non-changing". The slope of a corresponding imagined straight line is therefore > 0, < 0 or 0. The same applies to the next relation between time points two and three. To allow for this comparison, the absolute read values (correspondent to sequences) of each OTU had to be transferred to relative values based on the percentage they represent in a sample. Subsequently we calculated the abundance changes for each OTU between the first two (pre-diarrhea) samples. Together with the corresponding abundance values for diarrhea and post-diarrhea status, a three point profile (pre-diarrhea – diarrhea – post-diarrhea) of each OTU could be drawn. Subsequently, a scoring system was introduced that assigned values of -1 (decreasing abundance value between two states), +1 (increasing abundance value) or 0 (relative abundance change < +/- 0.05%) to the (two) slopes of this profile. Each OTU therefore generated a specific overall score that related to one of the nine possible progressions. When comparing the relative values of OTUs in the OTU table, most (60-70%) followed a "non-changing" pattern and only 30-40% showed significant increase or decrease. Omitting taxonomic units with a total score of 0 (no abundance changes > +/- 0.05%) we created a network in Cytoscape. The latter is a program especially designed for network analysis and was developed at the Institute of Systems Biology in Seattle [17].

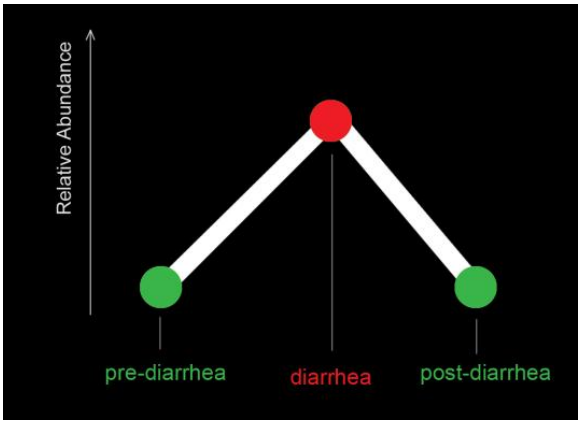

Figure 5: One of nine possible progressions (profiles) of OTU abundance during the course of a three time point experiment (pyramid profile).

Our Network Profiling features a number of distinct features: Each OTU is connected to its corresponding profiles in a spring embedded manner, placing units that relate to more than one progression nearer to the center of the network, while those corresponding to one distinct profile are clustered around the periphery. Therefore, the behavior of a specific OTU during the course of the experiment can easily be identified. The width and opacity of the lines indicate the number of patients in which an OTU clustered to a specific profile (The

more often an OTU is associated with a certain profile, the more dominant an edge will show in the network). The color of the lines represents connections to OTUs either considered associated with an abundance increase (red) or decrease (green) in the diarrhea state. Node sizes correspond to the mean relative abundance change between the pre-diarrhea and the diarrhea state, allowing an estimation of the impact each OTU receives when diarrhea is induced. Larger nodes indicate high abundance changes of an OTU, both negative and positive in relation to the pre-diarrhea state. Profiles with connections to a large number of OTUs can be considered dominant for a given investigated condition, which is further emphasized by the size of the nodes clustered near a profile. Therefore, certain conditions in the body could be identified by different distributions of OTUs amongst the 9 profiles.

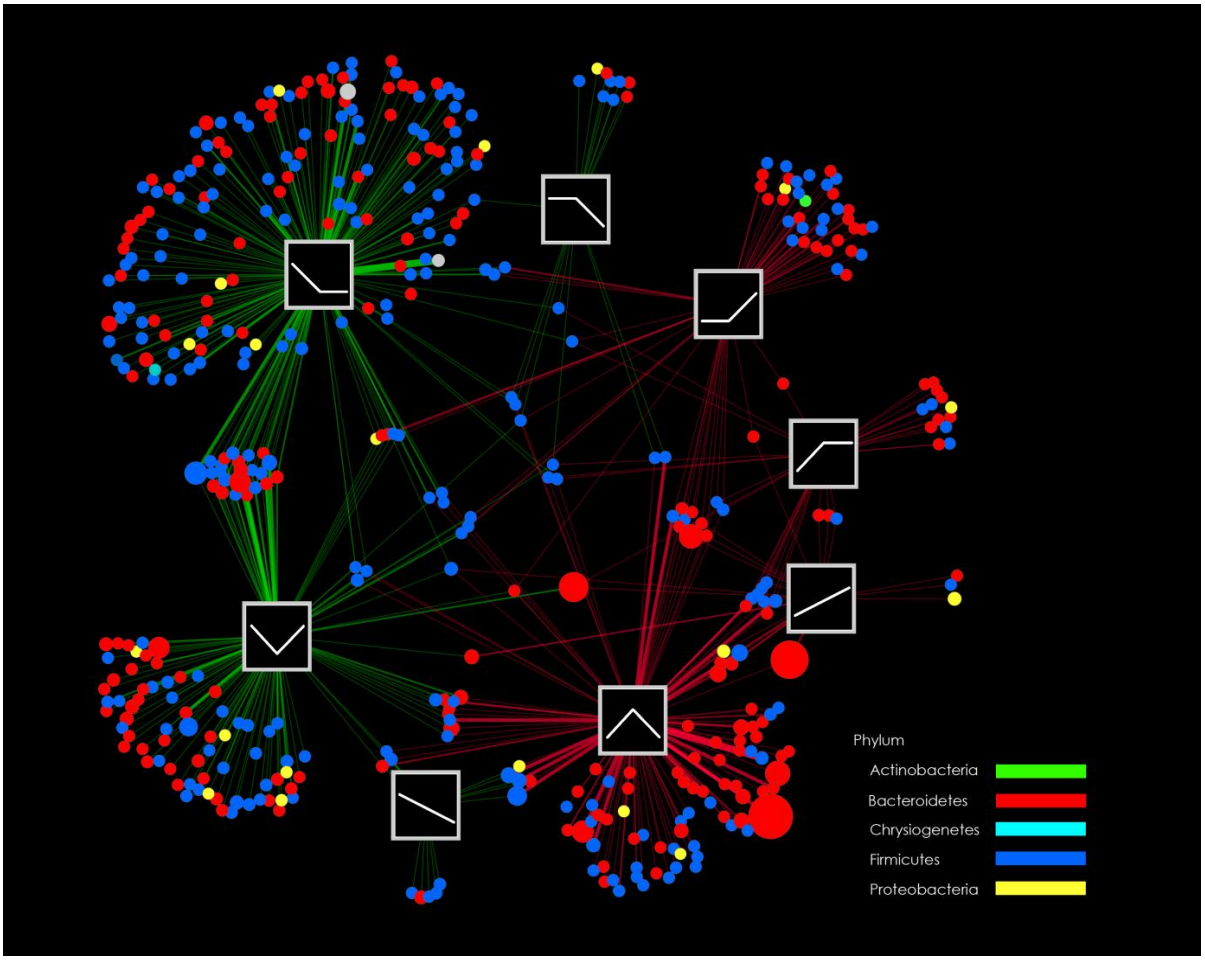

Figure 6: Cytoscape profile cluster analysis from stool samples showing the relative abundance change of phylotypes in response to diarrhea. Three point patterns including pre-diarrhea, diarrhea and post-diarrhea abundance levels are shown. Displayed are only phylotypes which were assigned to a respective reaction pattern in at least two individuals (corresponding to thin lines). The width of lines correlates to the number of individuals a respective phylotype was assigned to a specific reaction pattern. Phylotypes showing an abundance change of at least 50% are represented as full colored nodes; below that level, nodes appear transparent. Size of nodes correlates to the mean relative abundance change comparing pre-diarrhea to diarrhea samples. Phylotypes are colored according to their phylum membership and named according to the phylogenetic level conferred by the RDP classifier with 80% identity.

We propose that our profile clustering approach facilitates the identification of relevant taxa and could empower bacterial community analysis in the course of microbiome experiments with low sample counts.

Although the implementation of this network system in our experimental setup proved successful, there occurred a number of problems that should be considered when using this approach. Firstly, due to the calculation of the difference (a subtraction) between time points in certain cases a minimal change (as 20.1 to 20.3%) can lead for the OTU to be considered “increasing”. This has been circumvented by the introduction of a fold change regulation, where only OTUs get represented as full circles that reach a fold change of > 1.5. The higher the numbers of patients this rule applies, the higher the opacity of the corresponding node. Hence, OTUs that show small differences > 0.05% but do not reach a fold change > +- 50% are more transparent and therefore less prominent when viewing the network. The second limitation of the current network version is the inability to show the relative changes (node sizes) for more than two time points. This will be a concern for the ongoing development of the system.

For experimental setups with only 2 conditions like diseased/healthy, treated/untreated etc., the same principles are applicable as in three time point networks. But due to the fact that a similar number of OTUs is now clustered to only two possible profiles, one must find ways to enhance the accessibility of the network graphic. A good way to facilitate this proofed to be the combination of the network profiling approach with the global scoring system described in chapter 2.3. This allows a better filtering of OTUs of interest based on the global scoring system.

It is also important to note that both approaches, global scoring as well as profile clustering, are currently developed for the identification of OTUs which significantly change in the highest number of patients. There might be circumstances though, as when one is searching for species that are uniquely found in one sample, when both systems have to be adjusted to meet the new requirements.

3. Bacterial community changes in diarrhea

3.1 Materials and Methods

3.1.1 Study Setup and Sampling

The principal idea behind the first project discussed in this thesis was to investigate the behavior of the bacterial gut community when confronted with the effects of induced diarrhea. Because diarrhea is an accompanying symptom in a large number of gastrointestinal diseases, the intrinsic effects of this condition are of great interest.

The study was conducted with four healthy male Caucasians, which were subjected to a diarrhea inducing therapy with polyethylene-glycol 4000 (PEG). Neither had a history of diarrhea or was treated with antibiotics for a time period of at least 1 year prior to the start of the study. Stool sampling was carried out on four time points, t1 – t4, where t1 represents the starting point of the study, 7 days before a defined diet was administered to the participants. This diet consisted of a defined mixture of protein, fat and sugars for a time period of five days. On day two of this defined diet, another stool sample was gathered (t2). Subsequently, on the third day of the diet, diarrhea was induced by oral intake of 150 g PEG. This treatment was sustained for three days, after which another stool sample (t3) was taken. The final sample was acquired seven days after the withdrawal of PEG and return to a free diet (t4). Figure 7 shows the study timeframe:

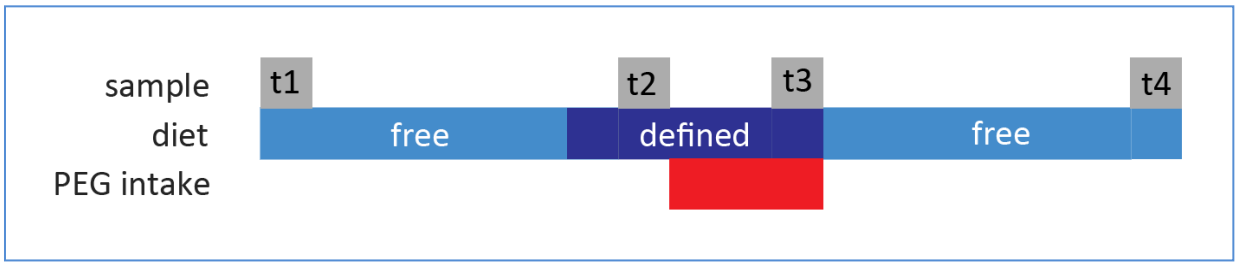

Figure 7: Study timeframe for induction of diarrhea with definition of diet periods and PEG intake duration. One box space represents one day.

In addition to stool samples, biopsies of the colon were carried out for three of the four participators on time points t2 and t3. The biopsy samples were taken from the sigmoid mucosa after washing the area with a physiological saline solution to avoid stool contamination.

Both stool and mucosa samples were immediately frozen at -20°C after sampling.

3.1.2 DNA Extraction, Amplification and Quality Measurements

DNA Extraction was carried out with QIAamp DNA Stool Mini Kit (Qiagen) for stool samples and with QIAamp DNA Mini Kit (Qiagen) for biopsies according to standard protocols. The only deviation from standard protocol was an additional incubation step for stool samples in a boiling water bath for five minutes.
The target gene for the subsequent pyrosequencing analysis was the 16S rRNA gene, which has been proven to be a suitable target according to prior studies [12]. A variable region (v1-2) of the 16S gene was amplified by using oligonucleotide primers BSF8 and BSR357. Each primer additionally harbored a barcode sequence (MID) to allow for sample allocation after sequencing The MID part was linked to the primer by a linker region. PCR conditions can be found in Table 2 and Table 3:

Table 2: PCR Mix for 16S V1-2 region amplification in stool and tissue

Material	Stool	Tissue
Sample	100 ng	10 ng
1x HotStar Master Mix (Qiagen)	50 µL	50µL
Primer BSF8	20 µM	20 µM
Primer BSR357	20 µM	20 µM

PCR conditions were chosen as follows:

Table 3: PCR Setup for 16S V1-2 region amplification in stool

PCR step	Temperature	Duration	Stool	Tissue
Initial Denaturation	95 °C	12 min		
Denaturation	95 °C	30 sec		
Annealing	56°C	30 sec		
Elongation	72 °C	60 sec	22 cycles	35 cycles
Final Step	72 °C	7 min		

For each sample, three separate amplifications were carried out under the same conditions. Resulting PCR products were separated via 1% agarose-electrophoresis in 1xTAE buffer and subsequently extracted and purified via the Qiagen Gel Extraction Kit according to the standard protocol. Extracted PCR products were then pooled and analyzed for fragment length and DNA quality by the BioAnalyzer 2100 DNA 1000 system (Agilent Technologies). DNA concentrations were measured with the QuanitDect reagent (Invitrogen).

3.1.3 Pyrosequencing

Pyrosequencing preparations included the mixture of targeted gene PCR products with a bead solution (Roche - FLX System). Targeted gene fragments and beads were mixed in an

equi-molar fashion and the mix applied to PicoTiter Plates (Roche) according to the manufactures recommendations.

3.1.4 Data Analysis

Pyrosequencing reads derived from the diarrhea project were analyzed with SnoWMAn (http://SnoWMAn.genome.tugraz.at), an analysis pipeline for pyrosequencing data developed at the Graz University of Technology. After an initial quality filtering step we discarded all sequences that featured a length <150 bp, any ambiguous characters or were not matching the forward primer (distance > 2)). Subsequently, the remaining sequences were clustered by complete linkage with Infernal V1.0., allowing for 0 to 5 % sequence similarity. Representative sequences from each cluster were picked and classified via the RDP Bayesian classifier 2.0.1. to assess taxonomy. Subsequent statistical Analysis was carried out within the software package “R” (V.2.12.1). A profile clustering approach was developed to identify OTUs of interest with methods described in chapter 2.4. Healthy gut status at time points t1 and t2 were evaluated together and their mean values were used for comparison against the diarrhea state t3.

The identification of phylotypes of interest was carried out with a global scoring system and a subsequent profile clustering network analysis. We compared the pre-diarrhea to the diarrhea state and the latter to the resulting post-diarrhea state. To calculate pre-diarrhea abundance a mean value of tp1 and tp2 was generated. A significant change was indicated by a difference between time points of at least +- 0.05%. If progressions showed changes below this threshold, they were considered “non-changing” and were not included in the network analysis.

3.2 Results

3.2.1 General Comparisons

Detailed Snowman and R analyses that were carried out over the course of the experiment can be found in the manuscript this master thesis supplements [18]. For all 16 stool samples that were sequenced, an average read depth of 18288 +/- 3198 was achieved. The 6 biopsy samples achieved an average of 25300 +/- 1075 reads. Phylum level evaluation of stool and mucosal samples showed a clear dominance of Firmicutes in the latter and Bacteroidetes in these stools (Figure 8).

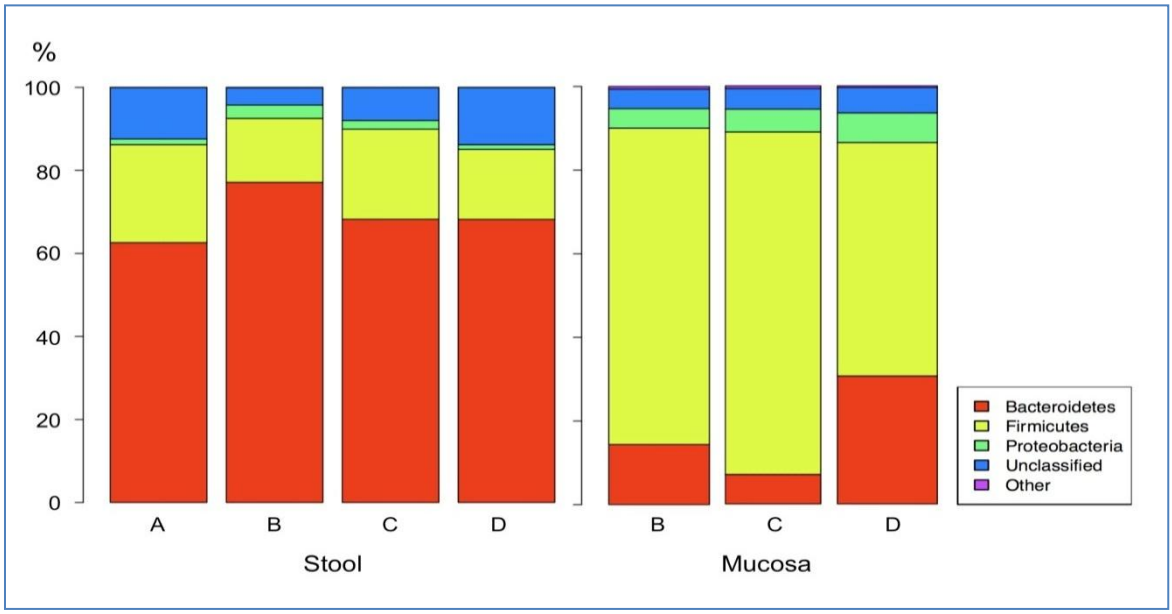

Figure 8: A bar chart representation of pooled stool and mucosa samples. Bacteroidetes are clearly dominating the stool community, while Firmicutes make up for the majority of reads in mucosa samples. Even on the low taxonomic phylum level variation between patients can be observed (reproduced from [17]).

Proteobacteria make up a comparatively group in both the stool (2%) and mucosa (5.7%) environments.

Beta diversity measurements by PCA analysis revealed a high level inter-individual variation of stool and mucosal microbiotas with a tendency of mucosal microbiotas to convergate in the event of diarrhea (Figure 9). This effect cannot be observed in stool samples, where inter-personal variation seems to be the

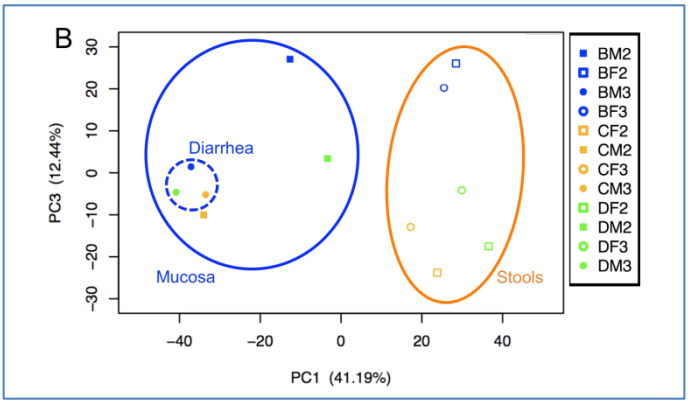

Figure 9: Representation of PC1 and PC3 with mucosa and stool sampling (reproduced from [17]). Legend: first position: participant, second position: mucosa (M) or stool (F) content. Numbers indicate timepoints

stronger clustering force. In general, a clustering of mucosa and stool samples can be observed, an effect that is due to the differences in these two different habitats and their resulting colonization patterns.

Alpha diversity measurements and rarefaction analysis suggested a diversity loss in acute diarrhea samples indicated by a lower slope plateau value (see Figure 10).

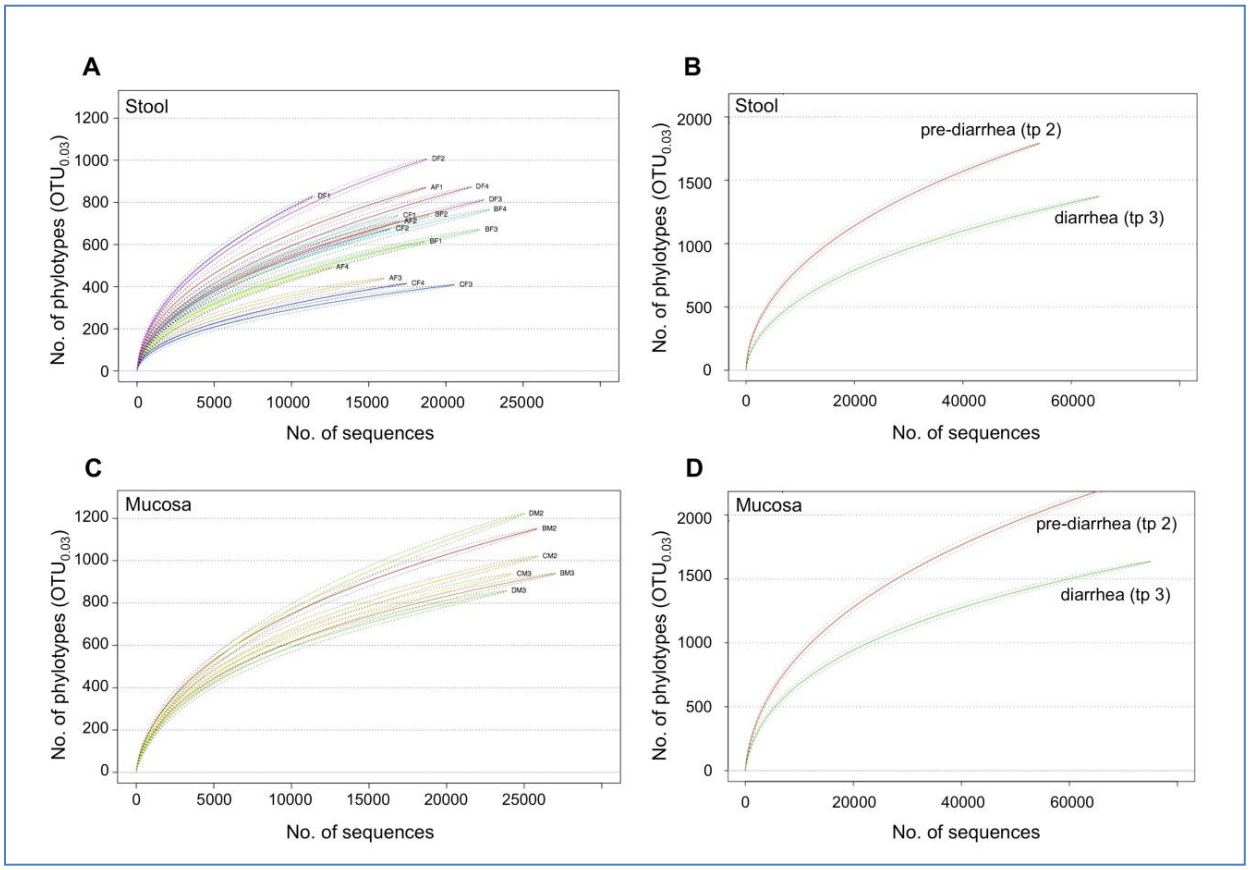

Figure 10: Alpha rarefaction analysis of stool and mucosal communities. Curves on the left side show alpha diversity measurements of all samples, whereas plots B and D approximate the mean diversities of time points tp2 (pre-diarrhea) and tp3 (diarrhea) - reproduced from [17].

The plots on the right hand side (B, D) of Figure 10 indicate a decreased species richness, or, in other words, the loss of a multitude of species (OTUs) formerly building up the microbial community. This can be observed in both, stool and mucosa samples.

After a period of diarrhea, the gut microbiota is reconstituting, where samples regain their pre-diarrhea species richness and taxa distribution. Graphically this was again assessed via PCA plots to investigate cluster forming of samples. Figure 11 shows two PC comparisons of the three dimensional PCA result:

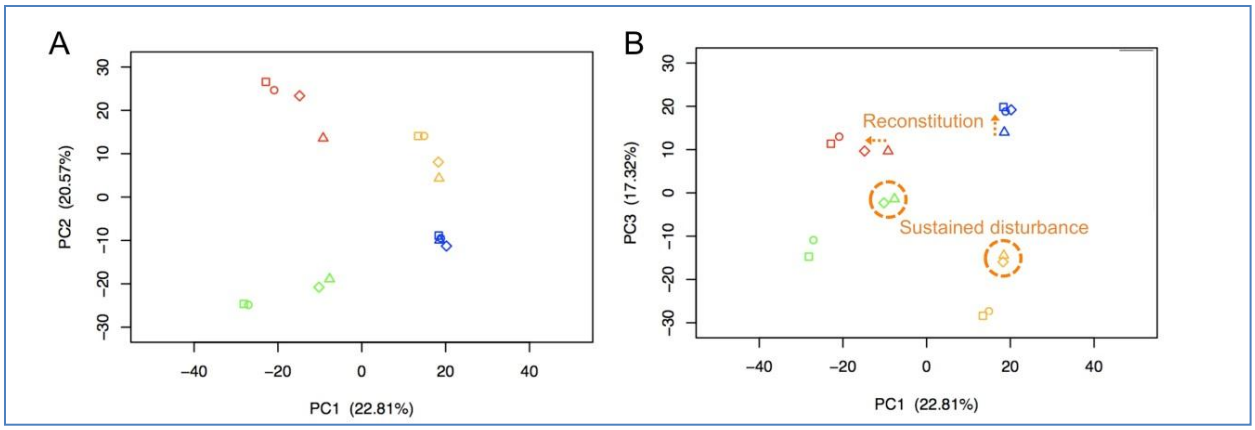

Figure 11: Reconstitution events in stool samples. Colors represent different patients. Notably, this reconstitution could only be observed in two patients (50%) (reproduced from [17]).

Figure 11 shows that reconstitution appeared only in two individuals (50%), while the other two showed a sustained alteration of the microbiotas one week after cessation of PEG.

3.2.2 Profile Clustering

Profile clustering analysis was carried out with 5885 OTUs of the combined stool and tissue sample data set. Of these, 5427 or 92.2% in stool and 5567 or 94.9% in mucosa samples were considered “non-changing”, as their respective relative abundance did not change within our threshold value of +- 0.05% over the course of the experiment. Phylotypes for stool and mucosa samples were then clustered to their progressions as described in Chapter 3.1.4. The resulting profile clustering network for stool is shown in Figure 13.

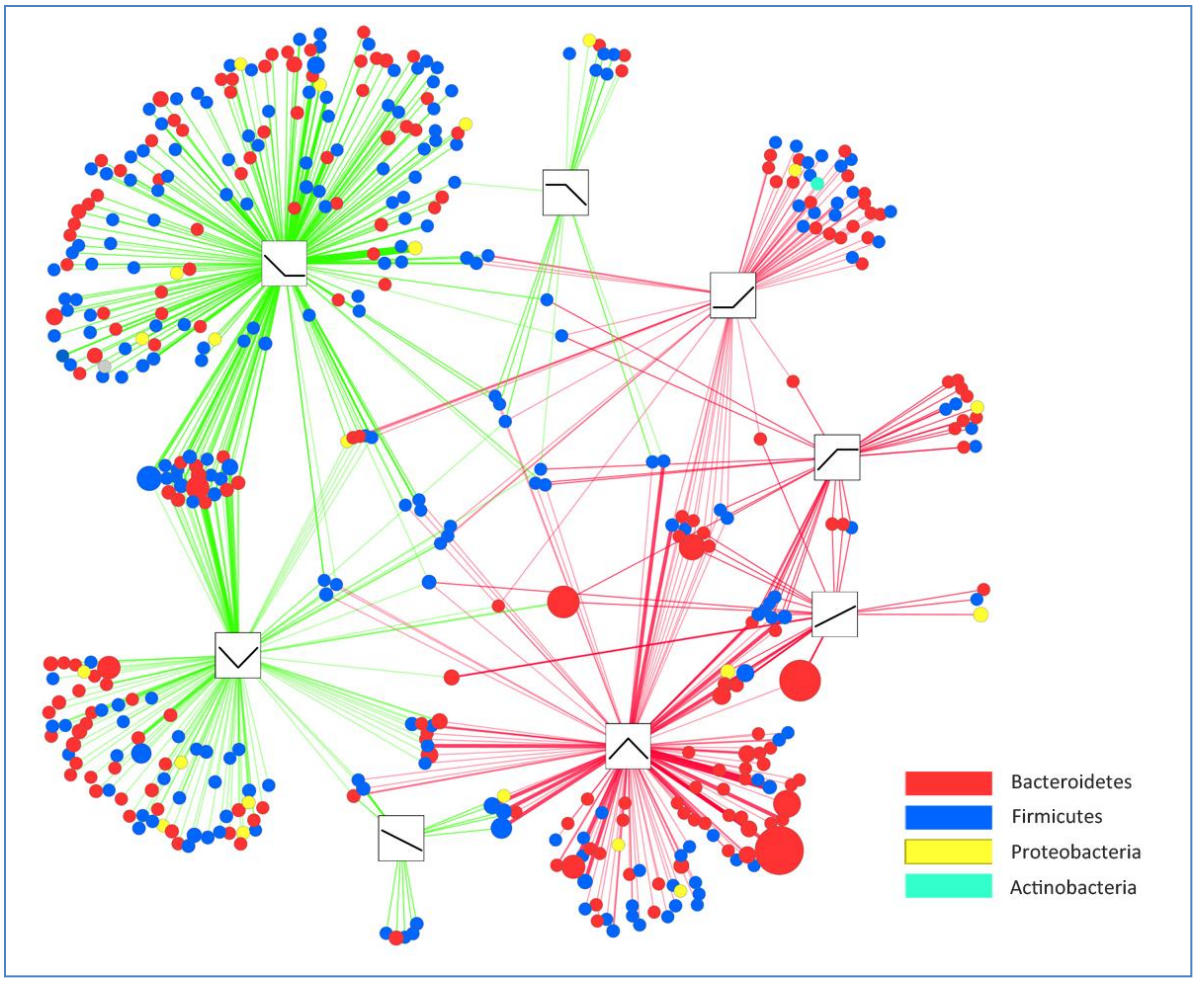

Figure 12: The diarrhea profile clustering Cytoscape network generated from comparing OTU progressions over the course of the experiment in stool samples: Each OTU is connected to its corresponding profiles in a spring embedded manner, placing units that relate to more than one progression nearer to the center of the network, while those corresponding to one distinct profile are clustered around the periphery. Therefore, the behavior of a specific OTU during the course of the experiment can easily be identified. The width and opacity of the lines indicate the number of patients in which an OTU clustered to a specific profile. The more often a phylotypes is associated with a certain profile, the more dominant an edge will show in the network. The color of the lines represents connections to OTUs either associated with a high abundance in diarrhea (red edges) or a low one (green edges). Node sizes correspond to the mean relative abundance change between the pre-diarrhea and the diarrhea state, allowing an estimation of the impact each phylotype receives when diarrhea is induced. Larger nodes indicate high abundance changes of an OTU, both negative and positive in relation to the pre-diarrhea state. Profiles with connections to a large number of OTUs can be considered dominant for a given investigated condition, which is further emphasized by the size of the nodes clustered near a profile.

A clear dominance of Bacteroidetes and Firmicutes is obvious in the gut microbiota. Numbers of Protebacteria and Actinobacteria, whose numbers in stools are naturally low [5] represent only a minor fraction of the microbiota. It should be noted that certain progressions are more likely to occur then others as revealed by the analysis. A dominant reaction of phylotypes in case of diarrhea is decreased richness, in other words a wash out effect during diarrhea (low-high-low (pyramid) and high-low-high (well) abundance profiles) and a loss of microbial phylotypes in stool without recovery after diarrhea (top left hand profile). Other progressions, like increasing abundance over all three time points or its opposite, a steady decreasing pattern, are less likely to occur; only few OTUs are clustered with these profiles. It is not uncommon for OTUs to behave in a different pattern in different

patients, hence they are clustered not around a specific profile but on the intersections between two or, when featuring a wide variety of reactions, in the middle of the network. Identifying these can be of great interest when searching for phylotypes with a more flexible behavior, e.g. when looking further into displacement effects. If, for example, one is comparing a pyramid and a well profile, OTUs clustering to both profiles could indicate bacteria that can react two ways, depending on the overall state of the bacterial community. Further insights into the general nature of the system can therefore be generated by examining subparts of the network and their respective clustered phylotypes.

To enhance the viewing characteristics of the network and identify OTUs of interest within it, a sub-network was created from Figure 12 only showing phylotypes whose progressions were found in at least two patients (Figure 13).

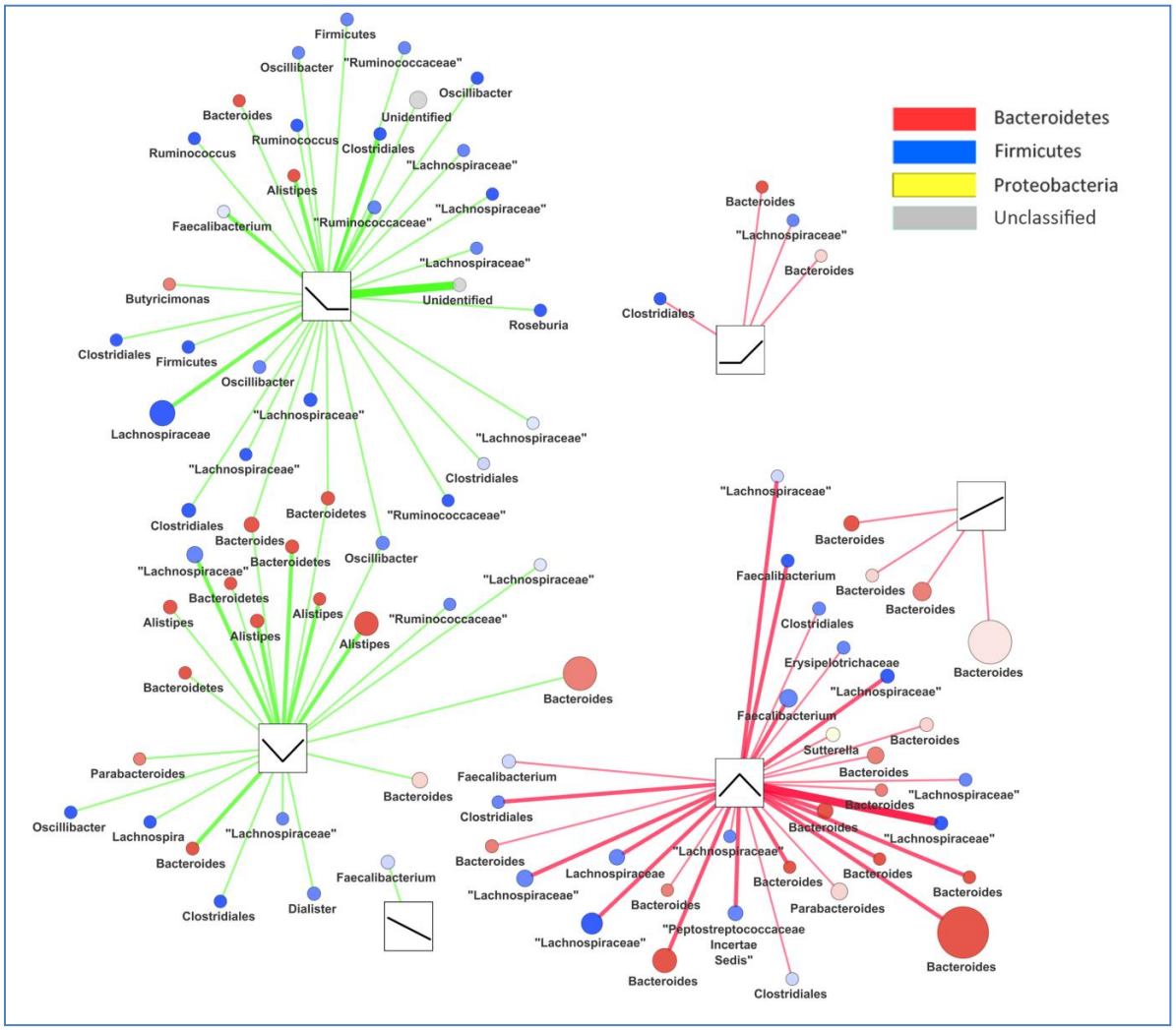

Figure 13: The diarrhea profile clustering Cytoscape network for stool samples simplified by only showing phylotypes that show a common progression in at least 2 patients. High opacity of circles indicates major fold change differences whilst more transparent circles indicate low fold change values for a given OTU progression. Nomenclature of nodes is based on the highest taxonomic classification with at least 80% confidence.

The network analysis revealed that Firmicutes and Bacteroidetes changed dominantly in stool samples of at least two patients. At genus level *Bacteroides* are associated with a pyramid pattern, *Oscillibacter* and *Ruminicoccus* show a decrease in diarrhea and do not rise in numbers in the post-diarrhea state. Proteobacteria were not obviously affected by diarrhea in stool samples, the exception was *Sutterella*, which shows a low fold change difference. Lachnospiraceae family members seem to interchange amongst each other - they can be found in 4 of 6 profiles indicating a wide range of reactions to the diarrhea impetus. A more in depth classification of this group could reveal if displacement amongst Lachnospiraceae family members occurs and which genera are taking part in this process. Data from mucosa samples was also visualized in a network (Figure 14).

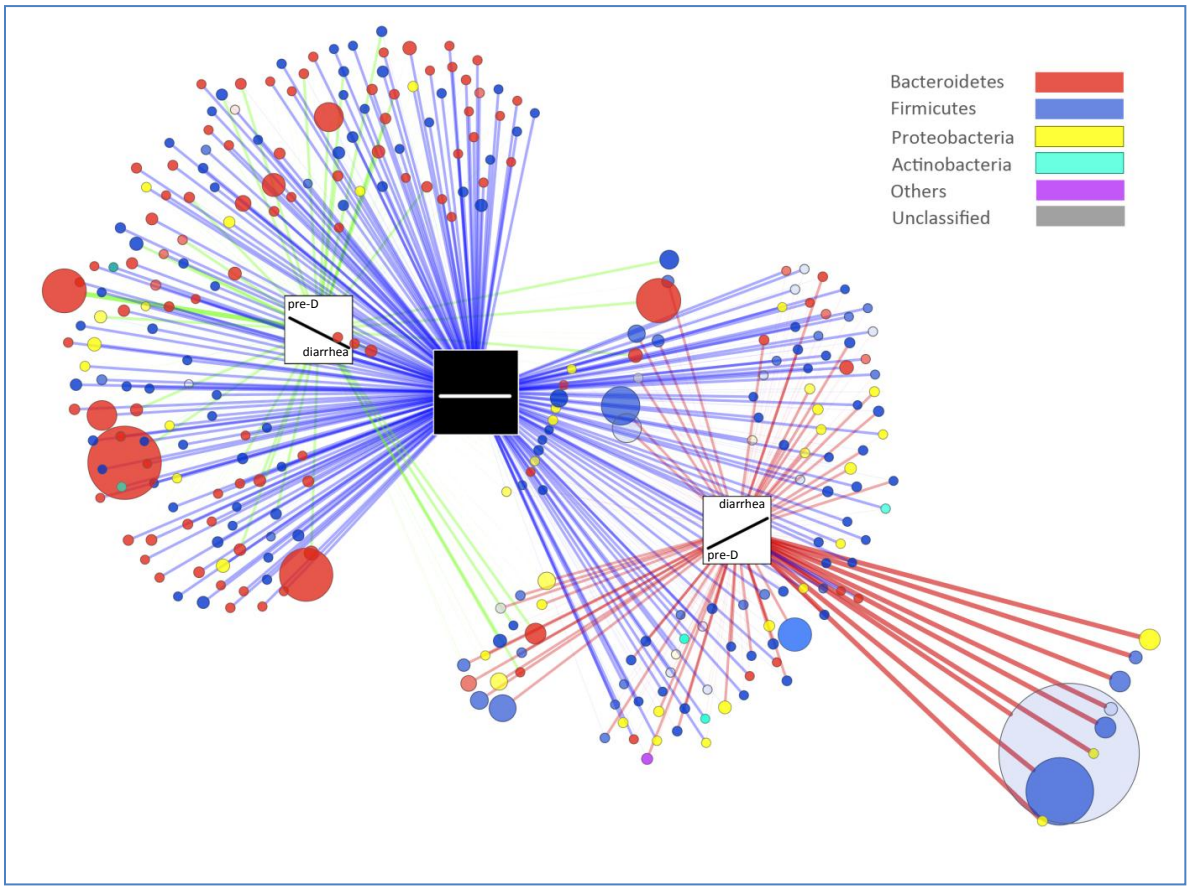

Figure 14: The diarrhea profile clustering Cytoscape network generated from comparing OTU progressions over the course of the experiment in tissue samples. The progression of pre-diarrhea state (pre-D) to diarrhea state is shown, increasing phylotype edges are colored in red, decreasing in green. Size of nodes indicates relative abundance difference. Connections to a non-changing behavior are shown. Notably, the majority of OTUs do not show a general pattern of either increase or decrease amongst all three patients. It is common for phylotypes which are found changing in one patient to maintain their relative abundance in others.

Figure 14 clearly illustrates that stable progressions with no significant abundance change account for the majority of OTU behaviors. Phylotypes showing significant increase or decrease in one patient often exhibit a non-changing trend in others.

Omitting non-changing phylotypes and only using connections found in at least two individuals allowed for a simplified representation of the mucosal microbial communities (Figure 15).

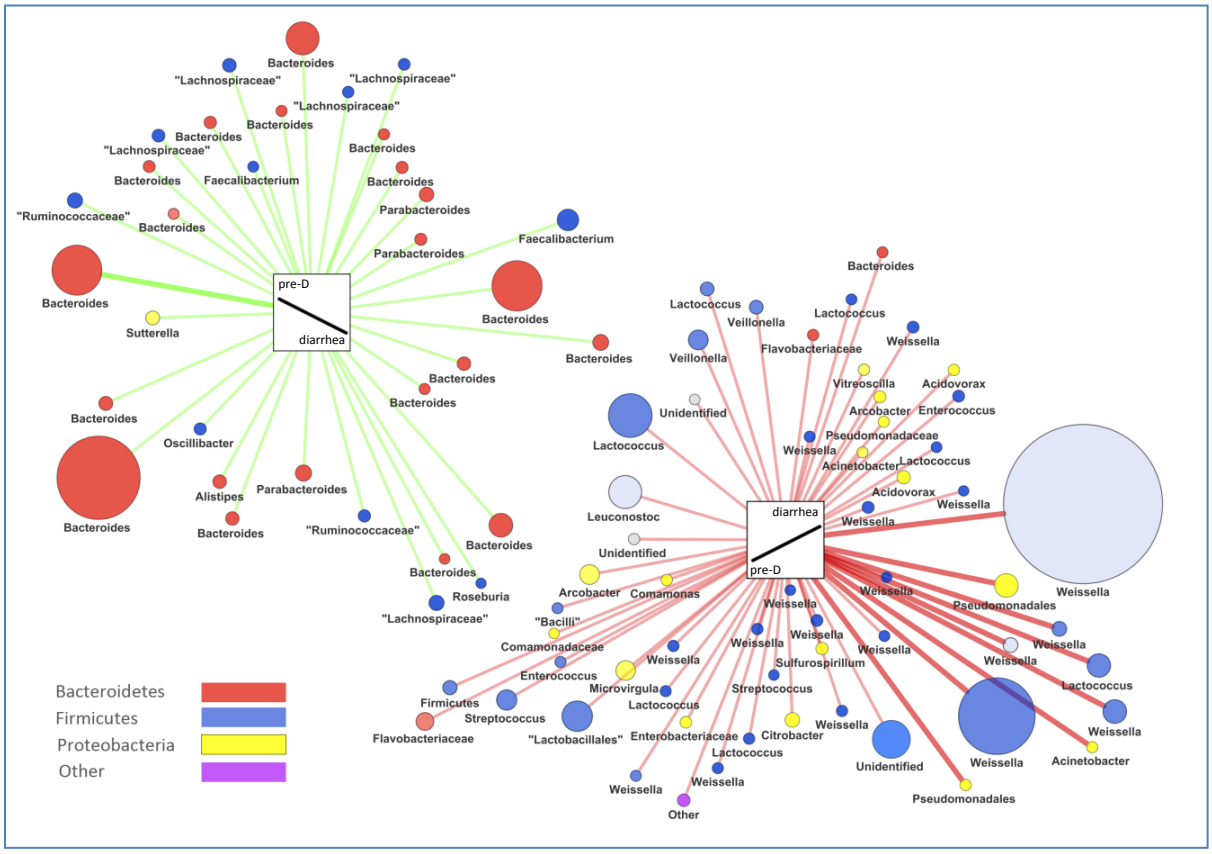

Figure 15: The diarrhea profile clustering Cytoscape network for tissue samples simplified by only showing phylotypes that show a common progression in at least 2 patients. The progression of pre-diarrhea state (pre-D) to diarrhea state is shown, increasing phylotype edges are colored in red, decreasing in green. Size of nodes indicates relative abundance difference; high opacity of circles indicates major fold change differences. Nomenclature of nodes is based on the highest taxonomic classification with at least 80% confidence.

Bacteroidetes are clearly associated with a decreased abundance on the mucosa while Firmicutes and Proteobacteria make up the majority of OTUs associated with an increased abundance during diarrhea. Especially *Weissella* and *Lactococcus* genera, members of the Firmicutes phylum, dominated the changing phylotypes. Also several Proteobacteria dominated the increased abundance pattern, with more than 15 different genera (amongst others: *Citrobacter, Acinetobacter, Acrobacter, Acidovorax, Microvirgula*). Genera that showed a significant decrease at the mucosa during diarrhea are represented by *Parabacteroides* and, to a bigger extend, *Bacteroides*.

3.3 Discussion

The majority of bacteria found in the human gut belong to either the Firmicutes or Bacteroidetes phylum. Data from tissue sampling though clearly shows that it is especially the Firmicutes, Bacteroidetes and Proteobacteria that increase significantly after induction of diarrhea. The major decreasing group on the mucosa is the Bacteroidetes phylum, which accounts for more than 75% of decreasing OTUs. This finding suggests that the community occurring in a healthy gut system is indeed ruptured by diarrhea and is displaced to some extend with phylotypes of the Proteobacteria phylum when comparing relative abundance values.

The most abundant genera found in human stool samples by Eckburg *et al*. were *Bacteroides, Alistipes, Parabacteroides, Dorea, Ruminococcus, Clostridium, Eubacterium, Faecalibacterium, Collinsella, Roseburia* with *Bacteroides being the most abundand* [3] but as in most studies in the gut microbiome field, no data was generated on tissue / mucosal samples.

The separate analysis and comparison between mucosa and stool habitats is vital for a successful description of the gut microbiome, as community structure differs greatly between the two habitats [2]. Implementation of both environments allowed for a comparative approach. In our study, a clear indication for the displacement effect is the high amount of Bacteroidetes found in stool samples and their decline on the mucosa. It seems that with changing environmental factors, Bacteroidetes get washed out of the community system. It could be shown in previous studies that Bacteroidetes numbers can significantly change with antibiotic therapy [19] or human weight loss [8]. In the latter it could also be shown that the level of functional diversity in a gut system is directly linked to the relative abundance of Bacteroidetes. Lachnospiraceae and Ruminoccocaceae are the other two groups found in mucosa samples that significantly decrease in abundance during diarrhea. As for Bacteroidetes, some of them might get washed out as indicated by an association with the pyramid profile. Additionally, they are also associated with the other two main changing patterns in stools (decrease-increase and decrease-nonchanging), which indicates that their elimination from the community is not associated with a specific time point but a more long lasting effect, which takes place in a wider timeframe. One could say that the reaction of Bacteroidetes to diarrhea is a more acute event, with subsequent fast normalization in its washing out behavior, while Lachnospiraceae and Ruminococcaceae show a more diverse and enduring abundance change. Both families are common in gut communities [8]. Bacteroidetes can be seen as (pan-)bacterial genomic libraries, as they harbor a plethora of genes that can be non-active for a long time but can be switched on if environmental conditions require it [20].This could be the reason for the more “acute” response of Bacteroides to diarrhea.

Members of the Proteobacteria phylum seem to increase after diarrhea. A possible explanation for this finding is that members of this phylum quickly occupy niches in the gut with a successful local colonization. *Lactobacillus* and *Lactococcus* on the other hand are usually considered beneficial human gut bacteria (i.e. probiotic bacteria) [6]. Their numbers increase on the mucosal layer. This effect could be attributed to the loss of genera like Bacteroidetes and the subsequent rise of phylotypes that flourish not due to pathological conditions but simply due to loss of competition for available nutrients. This finding is backed up by the results of our snowman / R analysis [18], especially the PCA analysis and rarefaction curves, where a clear decrease in diversity is observable (Figure 10). By promoting certain beneficial bacteria or a decrease in immunologic actions against them, reconstitution of a normal bacterial gut community could be enhanced and accelerated. When comparing the similarities between samples, such a relationship can also be seen in two patients and their respective bio diversities (Figure 9). While it seems to take longer for some individuals to recover from an event of diarrhea, others show a strong reconstitution with a fast return to the pre-diarrhea diversity.

As science investigates the complex interactions of the bacterial – human relationship it becomes more and more obvious that collective reactions like these are not an exception but a general principle of biological systems. Data acquired from this analysis suggests that events like diarrhea, which are usually considered harmless, can indeed have pronounced effects on the human gut microbiome. Due to significant changes in the bacterial community, a short lasting bacterial infection could lead to a misbalance in the bacterial-host relationship and subsequently to the development of an altered (disordered) community.

4. Bacterial community changes in Enterocolitis

4.1 Materials and Methods

4.1.1 Study Setup and Sampling

Nongranulomatous chronic idiopathic enterocolitis (NCIE) is a chronic inflammation of the human small intestine, which leads to severe diarrhea and subsequent dehydration. High mortality rates and an unknown but suspected infectious etiology make it a very interesting target for investigation. Our principal hypothesis was that a dysbiosis (misbalance) of the human small intestinal microbial community led to an over-stimulation of the immune system, hence causing chronic inflammation.

Human biopsies of three female patients (age 20-49) diagnosed with NCIE were available in FFPE (formalin fixed paraffine embedded) material and investigated using a pyrosequencing approach. All samples were obtained from the Biobank of the Medical University of Graz (http://www.medunigraz.at/biobank). The bacterial community was analyzed in two ways: the first was a classic diversity comparison, which should give insights into the general community dynamics in NIEC. The second was to develop a method to analyze the samples in more detail on the OTU level. Thus, OTUs of interest should be found that changed significantly between the two main states (acute/remission). As the interpersonal variety in the gut microbiome is very high, an approach overcoming the limitations of statistics when dealing with small sample size experiments was chosen. That led to the development of the scoring system described in chapter 2.3, with further development into the profile clustering system.

Table 4 gives an overview of samples acquired from FFPE biopsies.

Table 4: Sample nomenclature of NCIE patients

Patient	Sample Name Run1	Disease State	Location
A	FFPE10	Remission	Small bowel (Duodenum)
	FFPE11	Acute	Small bowel (Duodenum)
	FFPE12	Acute	Small bowel (Duodenum)
B	FFPE13	Remission	Small bowel (Duodenum)
	FFPE14	Acute	Small bowel (Duodenum)
	FFPE15	Acute	Small bowel (Duodenum)
	FFPE16	Acute	Small bowel (Duodenum)
C	FFPE17	Remission	Small bowel (Duodenum)
	FFPE18	Acute	Small bowel (Duodenum)
	FFPE19	Acute	Small bowel (Duodenum)

As NCIE is a very rare disease, only a small number of patient samples were available. Histological assessment available from all samples allowed for stratification of samples into remission / acute disease states (see Table 4). All samples in this project were derived from FFPE (formalin fixed paraffin embedded) material, a common agent to conserve and fix tissue samples for histopathological assessment.

4.1.2 DNA Extraction, Amplification and Sequencing

Tissue from FFPE material was cut via a microtome and extracted with the Qiagen DNA mini Kit (Qiagen) according to the standard protocol. Subsequent 16S V1-2 amplification was carried out with Roche standard primers (primer_fw and primer_rev) and 6base MID codes (Table 7). Separation of the amplified sequences and purification was then achieved by gel electrophoresis. Sequencing was accomplished on a Roche FLX Sequencing System in PicoTiter Plates.

Table 5: PCR mix for 16S amplification

Material		Concentration
Sample	2 µL	Range: 0.5 – 12.2 ng/µL
1x HotStar Master Mix (Qiagen)	10 µL	
Primer_fw	0.6 µL	10 pmol/µL
Primer rev	0.6 µL	10 pmol/µL
H_2O	6.8 µL	

PCR conditions were chosen as follows:

Table 6: PCR parameters for 16S amplification

PCR step	Temperature	Duration	
Initial Denaturation	95 °C	12 min	
Denaturation	95 °C	30 sec	Run1
Annealing	56 °C	30 sec	
Elongation	72 °C	60 sec	35 cycles
Final Step	72 °C	7 min	

A compilation of sequenced samples with their respective sampling dates and primer design can be found in Table 7.

Table 7: Samples sequenced and analyzed in both runs with sequences of primers and data of biopsy sampling

Patient	Sample ID	Run	Mid	Primer	Sampling Date
A	FFPE_010	1	ATGCAG	AGAGTTTGATCCTGGCTCAG	08-08-06
	FFPE_011		ATGCTC	AGAGTTTGATCCTGGCTCAG	08-02-27
	FFPE_012		CAGAGC	AGAGTTTGATCCTGGCTCAG	08-03-10
B	FFPE_013		CAGATG	AGAGTTTGATCCTGGCTCAG	06-02-27
	FFPE_014		CAGCAG	AGAGTTTGATCCTGGCTCAG	04-09-24
	FFPE_015		CAGCTC	AGAGTTTGATCCTGGCTCAG	04-02-10
	FFPE_016		CATCTG	AGAGTTTGATCCTGGCTCAG	04-02-03
C	FFPE_017		CATGAG	AGAGTTTGATCCTGGCTCAG	97-09-23
	FFPE_018		CTCATG	AGAGTTTGATCCTGGCTCAG	95-09-18
	FFPE_019		CTGATC	AGAGTTTGATCCTGGCTCAG	95-11-06

4.1.3 Data Analysis

The study was based on a retrospective assessment of specimens available in the tissue bank. Therefore, time between sampling and number of samples varied amongst the patients. Figure 16 shows sampling points and disease states of all three patients on a coherent time scale.

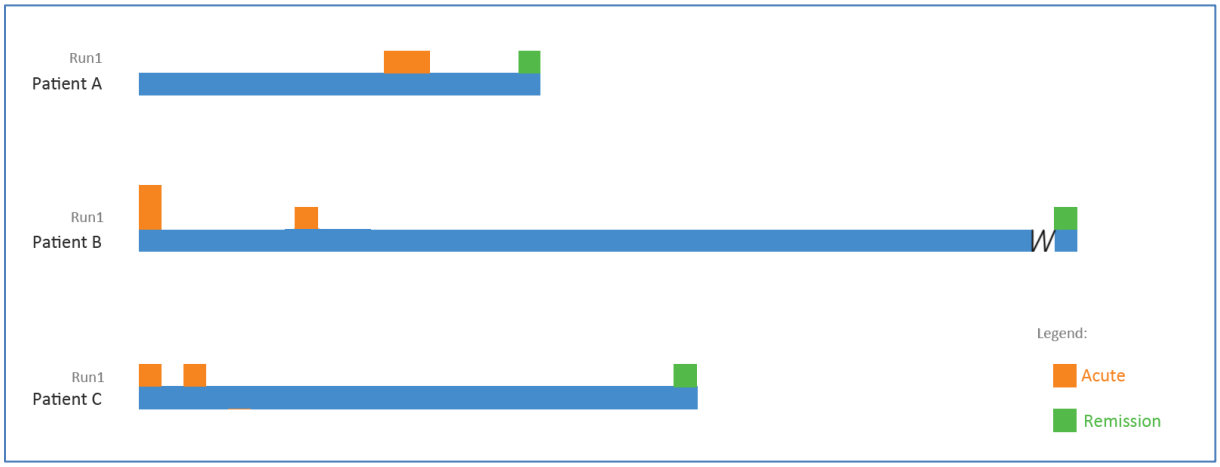

Figure 16: Sampling timepoints (colored boxes) for all three patients. Timeframes are drawn to scale, with one box space representing one month. Two boxes stacked onto each other indicate two sampling points in this particular month.

The variability amongst sampling times led to a two state comparison approach, where mean values for acute vs. remission states were compared rather than longitudinal progressions (as in the diarrhea project).

The data analysis pipeline package Qiime was used to create OTUs, diversity measurements and phylogenetic trees. Custom parameters can be found in Table 8. For filtering out

ambiguous and low quality reads, the quality file of the GS FLX system was used in the split_libraries script.

Table 8: Qiime parameters for both enterocolitis runs

Qiime parameter	Value
OTU picking algorithm	UClust
Cluster similarity	0.97
Picking representative OTUs	most_abundand
Alignment method	PyNast
Minimum sequence length	150 Bp
Taxonomy assignment method	RDP
Taxonomy assignment Confidence	0.8
Tree building method	fasttree
Rarefaction depth	3727
Number of reps	100

After sequencing and basic data analysis with Qiime, a global scoring system was developed to find OTUs of interest. Subsequent clustering of progressions was achieved with a profile network approach (see chapter 2.4). A cluster similarity of 0.97 (97%) was chosen to enable OTUs presenting species. The RDP classifier was used to determine taxonomic classification; output was restricted to an 80% confidence threshold allowing for taxonomic assignment. Comparing the relative abundance values for the remission state versus acute states in any patient, two point profiles (remission versus various acute states) could be drawn. After omitting taxonomic units with a total score of 0 (relative abundances in two samples differ by less than 0.05%) we created a Cytoscape network, where each OTU was connected to its corresponding profiles. To enhance visualization properties, only OTUs with high global scores were used to build the network in Figure 23. For the global scoring system, acute versus remission states have been compared and their respective difference calculated. For all comparisons, a local score has been calculated that indicates, how many times an OTU is changing in one patient and into what direction (increasing or decreasing manner). From these local scores a global score could be calculated as explained in chapter 2.3.

OTUs of interest that reached a high global score but could only be classified to the genus level with the RDP classifier approach were later blasted with RDP SeqMatch to reach species level classification (http://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp) and the blast option on HOMD (Ver. 10.1 – standard parameters) (http://www.homd.org/modules.php?op=modload&name=RNAblast&file=index&link=upload). HOMD is a bacterial community analysis platform for the human mouth microbiome [21].

For validation of global scoring, a second approach was used, where acute versus remission was calculated as division rather than difference.

4.2 Results

4.2.1 Histological assessments

Nongranulomatous chronic idiopathic enterocolitis leads to severe alteration in the structure of the human small intestinal surface by chronic inflammation. These changes in the human mucosal layer of the small gut (Duodenum) are shown in Figure 17.

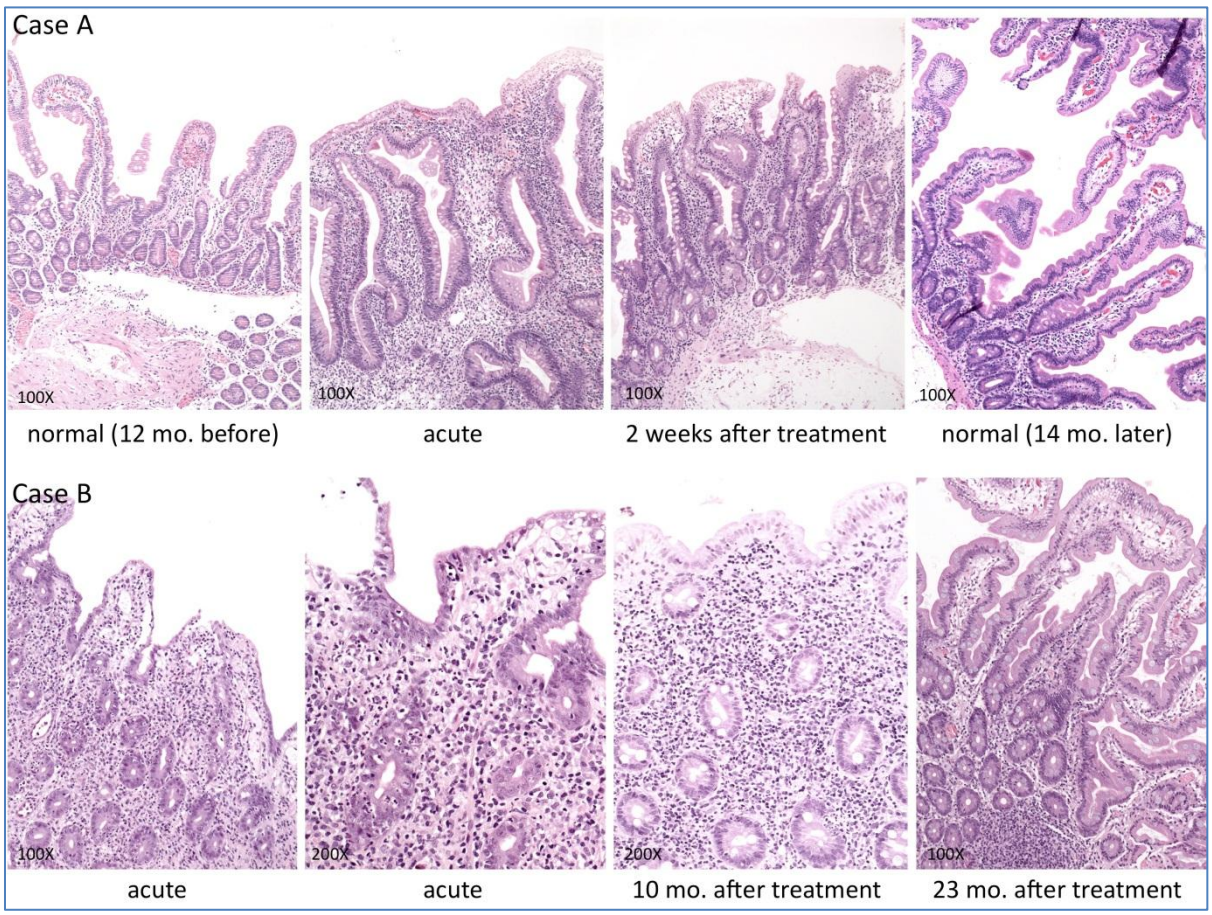

Figure 17: Histopathological representation of tissue samples in NCIE cases. A healthy duodenal mucosa is composed of a multitude of villi and crypts to increase the surface in the gut. During inflammation one can observe a loss of this surface structures are shown. Histological assessment is shown for patients A and B. Healthy tissue before acute onset of the disease is shown in the top left panel for case A. Direct comparison shows that after antibiotic and anti-inflammatory treatment a regeneration of the layer to pre-disease status could be accomplished (right side panels).

4.2.2 Qiime Analysis

On average, a sample achieved 10082 reads in the Qiime analysis (+/- 2908). 1695 OTUs have been identified at 97% similarity clustering. Of these, 1076 have been identified as non-changing, indicating that they did not change their relative abundance in all acute versus remission comparisons or did not reach the threshold level of +- 0.05% abundance change. This equals to 63% of all OTUs.

Alpha rarefaction analysis showed that a high level of coverage was reached throughout all samples. Additional sequencing effort would therefore have to be substantial to identify new phylotypes (Figure 18).

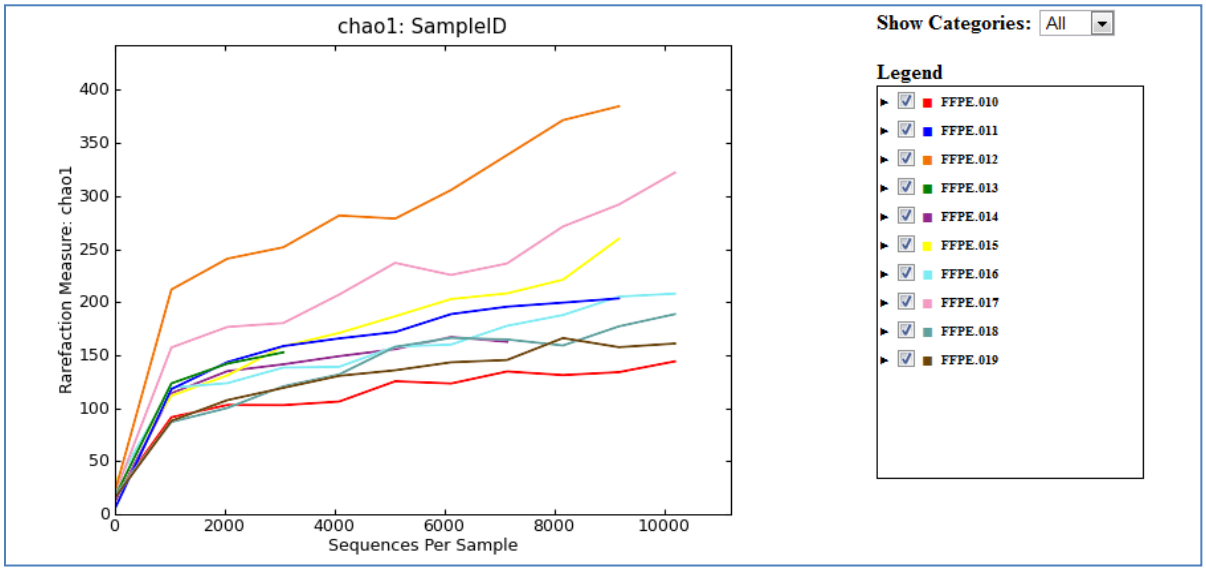

Figure 18: Alpha Rarefaction analysis of Enterocoliits FFPE samples. (Chao1 diversity computation is shown)

Phylum level analysis revealed a decrease in Firmicutes and Bacteroidetes and a parallel rise of Actinobacteria as well as Proteobacteria in the acute state (Figure 19). Comparing distinct samples (Figure 20) on phylum level confirmed that trend in all three patients, although in different forms. A high phylum-level inter sample variation could be observed between patients which indicated a thorough change in the bacterial community.

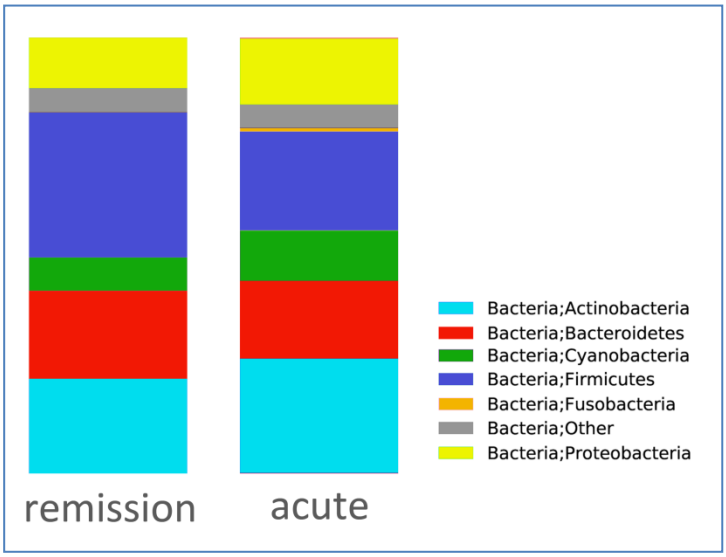

Figure 19: Phylum level abundance in remission and acute states (pooled data)

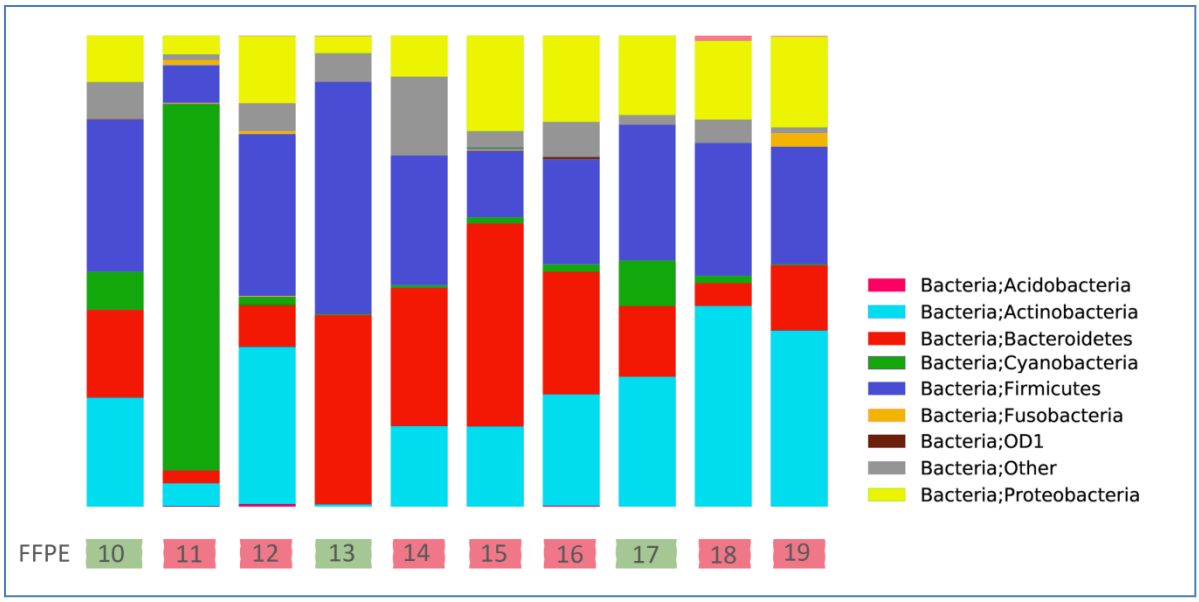

Figure 20: Phylum level abundance in various samples. FFPE sample numbers with green background indicate remission states, red background displays acute states.

The high number of Cyanobacteria present in FFPE11 was possibly caused by contamination of the sample.

Inter sample diversities represented via PCA plots (Figure 21) revealed a significant disturbance in the gut microbial community. Patients formed loose clusters, where acute samples showed large dissimilarities to the remission states. When coloured by disease state, clustering of remission states occurred to some extent, although no general principle of reconstitution could be identified.

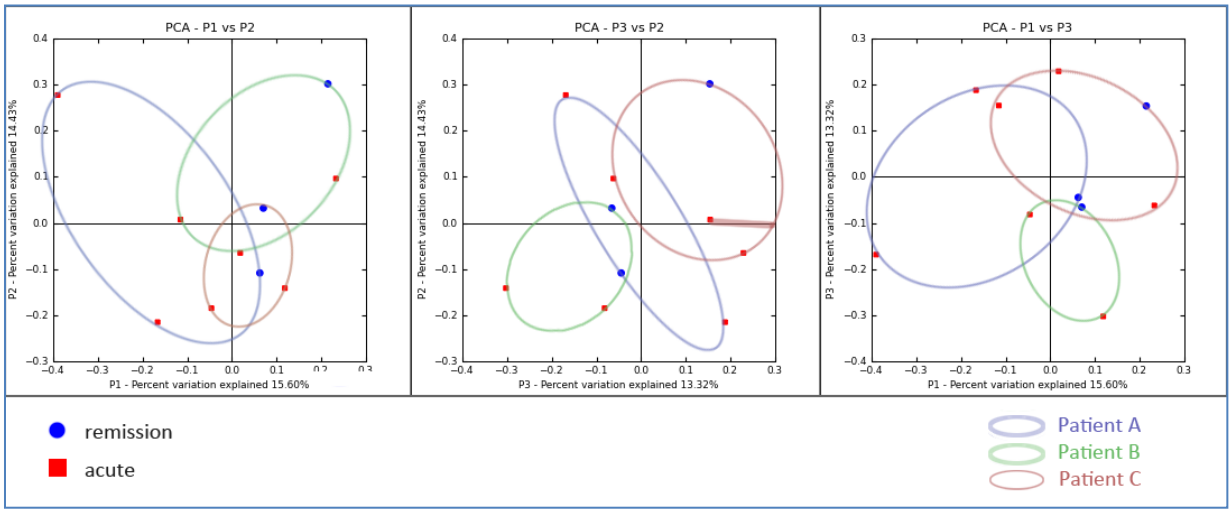

Figure 21: Principle Component Analysis of FFPE samples 10-19. Remission (blue dots) and acute states (red squares) are connected in an ellipse to show patient affiliation. Three PCA dimensions (1-3) are shown.

Figure 20 and Figure 21 clearly demonstrate that community compositions are not only dissimilar between patients but also when comparing the microbial community of one patient over a longer period of time. This initial whole community approach was therefore

limited by a high inter and intra personal variation. An OTU based approach had been developed to overcome these limitations.

4.2.3 Network Analysis

The network profiling (Figure 22) using a global scoring system led to a number of findings not accessible via a basic sample comparison. While Firmicutes and Bacteroidetes make up for the major part of a healthy human gut [2,5], the duodenum features high numbers of Proteobacteria as well as Actinobacteria that change significantly amongst all three patients (Figure 22).

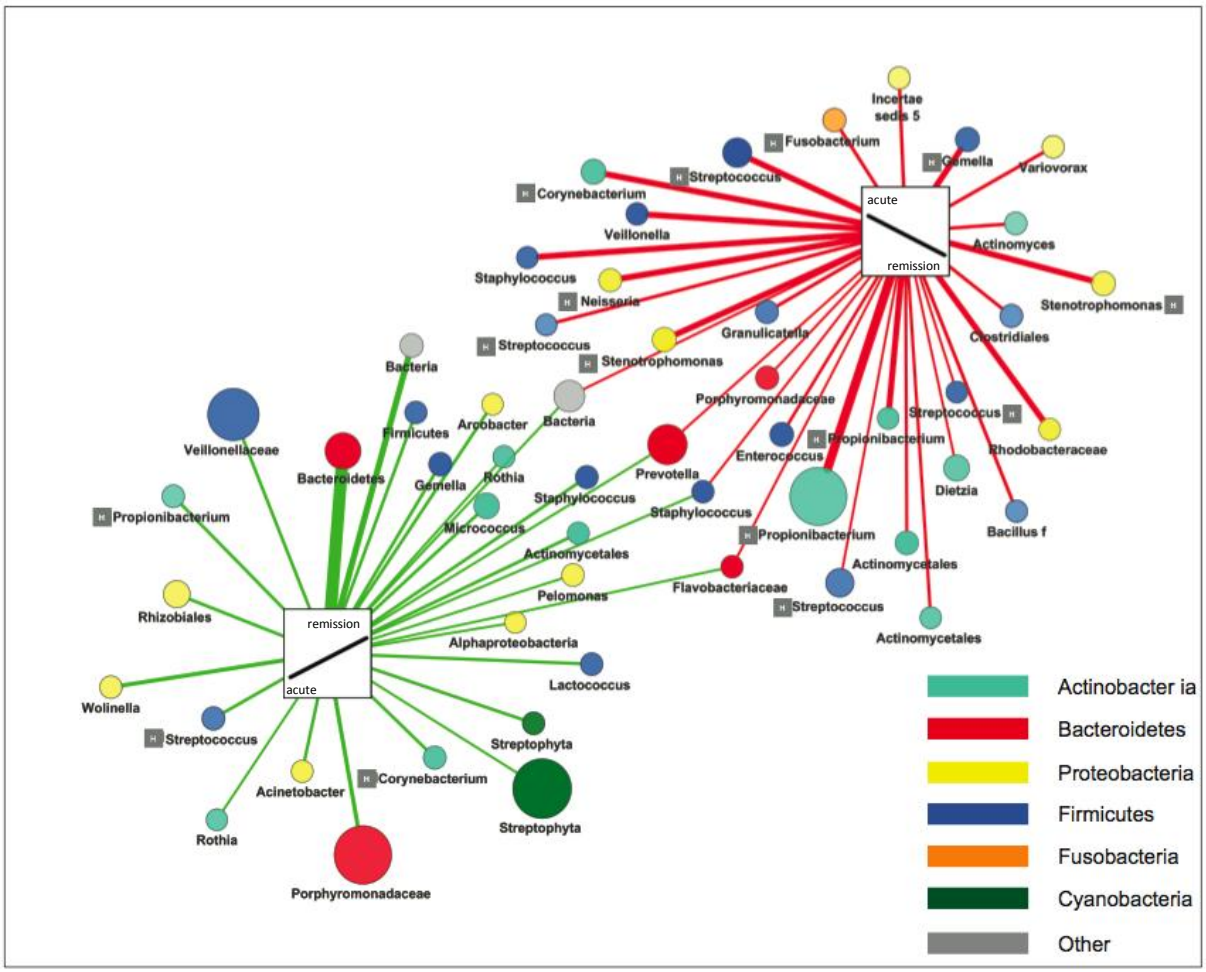

Figure 22: Phylotypes showing an increasing (red) or decreasing (green) relative abundance comparing acute phase of disease with the remission state. OTUs showing the respective abundance change are connected to their respective profile. The width of the lines indicates the number of patients in which an OTU has clustered to a specific profile. Node sizes correspond to the mean relative abundance change between the acute and the remisson state. The transparency of nodes is dependend on the number of times a given fold change value (+-50%) was reached. The node labels show the highest possible taxonomic classification with at least 80% confidence (Ribosomal Database). OTUs that are prominent members of the human oral microbiota are indicated with an H.

The majority of OTUs do not cluster with one specific profile in more than 1 patient. This is a reasonable finding, given the large interpersonal variety of the human gut microbiome and the resulting possibilities for the gut community to react to an impetus. However, the small number of OTUs that showed a similar behavior in many patients could represent common microbial agents that are crucial for the development of the investigated condition. In our study, we identified a number of agents that could be viewed as a base disturbance in the case of NCIE acute state (e.g. *Fusobacterium, Propionibacterium, Stenotrophomonas, Streptococcus and Staphylococcus*). Some of them are already known for acting detrimental in other forms of chronic GI inflammation.

Figure 23 depicts the highest global scorers (minimum global score = 9) in NIEC.

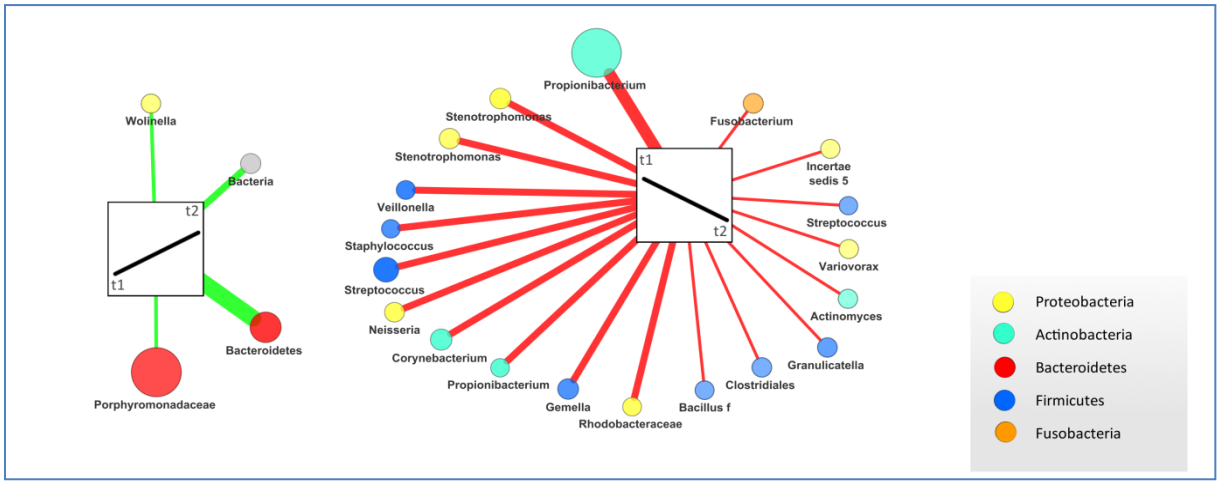

Figure 23: A profile clustering approach with a global scoring system incorporated. The global score is made up of difference local scores and therefore a measure of how many times an OTU shares a common progression in all patients. OTUs are clustered to progressions between states t1 (acute state) and t2 (remission). Of the several thousand OTUs resulting from the pyrosequencing analysis, only those with high global scores are shown. Edge thickness is based on the global score while the size of nodes is related to the mean abundance change a phylotype features between the acute and remission measurements. Nomenclature of nodes is based on the highest taxonomic classification with at least 80% confidence.

Even though the global score is a measure to identify OTUs that behave similarly in all patients, a longitudinal representation showed the large interpersonal variation that even these high global scorers featured (Figure 24). Whilst genera like the top right hand *Propionibacterium* or *Stenotrophomonas* somewhat showed a common trend, others like *Corynebacterium* behaved significantly different in at least one of the patients.

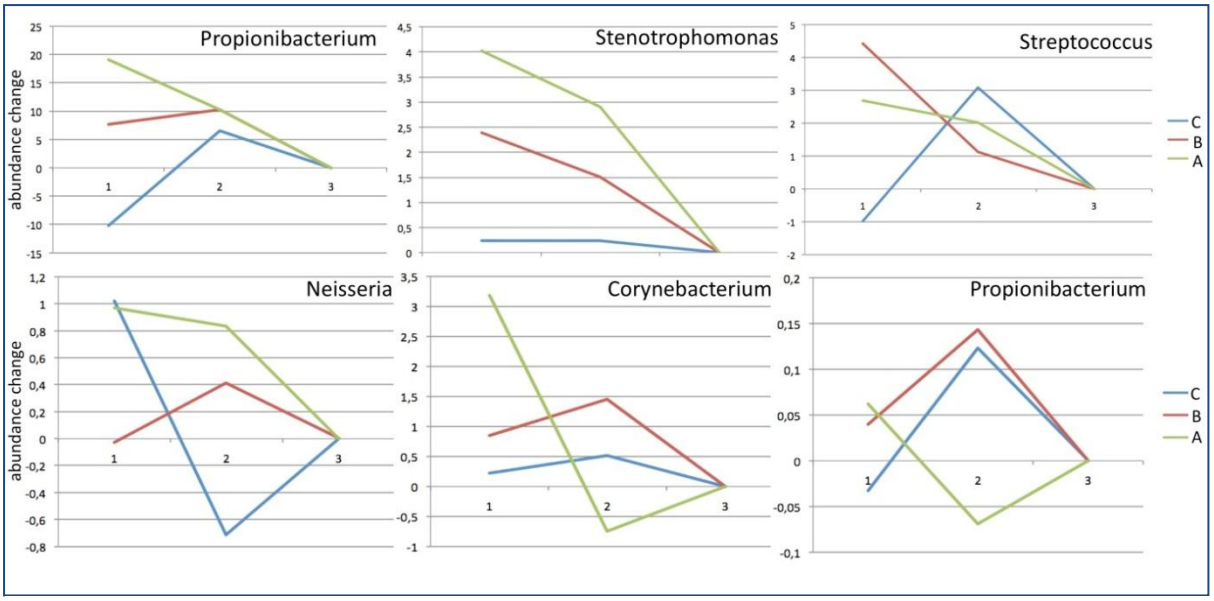

Figure 24: Abundance change (y-axis) of selected OTUs during the course of disease (x-axis: time-point 1 & 2, acute phase; time-point 3, remission). Samples from all three individuals (A, B, C) are shown. Although these high scoring phylotypes were found to be related to the disease status not every OTU shows a similar behavior (abundance change) in every patient. Data is normalized to the abundance value of t3.

For the top scoring OTUs identified via the global scoring system, their respective representative sequences were blasted with HOMD. When comparing the results to RDP classifier (whole database) hits, a far higher identity match could be observed in HOMD. Top scoring strains that are associated with a high relative abundance in the acute disease state are shown in Table 9.

Table 9: HOMD database output for OTUs of interest (high global scorers). Only the top ranking classification results are shown. Phylotypes are arranged to reflect counter-clockwise decreasing global scores as seen in Figure 23. Identities/ Mismatches are calculated with gaps and AGCTU.

Network Name	HOMD Identification (10.1)	Identities	Mismatch
Propionibacterium	Propionibacterium acnes | Oral Taxon 530 | Strain 63597 | AF145256 | 70 | N	100	0/247
Stenotrophomonas	Stenotrophomonas maltophilia | Oral Taxon 663 | Strain LMG 958 | X95923 | 28 | N	97.4	6/228
Stenotrophomonas	Stenotrophomonas maltophilia | Oral Taxon 663 | Strain LMG 958 | X95923 | 28 | N	99.2	2/237
Veillonella	Veillonella atypica | Oral Taxon 524 | Strain DSM 20739 | X84007 | 208 | N	99.1	2/219
Staphylococcus	Staphylococcus epidermidis | Oral Taxon 601 | Strain ATCC 14990 | D83363 | 3 | N	100	0/227
Streptococcus	Streptococcus mitis bv 2 | Oral Taxon 398 | Strain SK34 | AY005045 | 410 | N	98.2	4/220
Neisseria	Neisseria flavescens | Oral Taxon 610 | Strain LNP444 | AJ239280 | 115 | N	99.5	1/220
Corynebacterium	Corynebacterium mucifaciens | Oral Taxon 835 | Strain DMMZ 2278 | Y11200 | 0 | N	93.8	15/243
Propionibacterium	Propionibacterium acnes | Oral Taxon 530 | Strain 63597 | AF145256 | 70 | N	99.7	1/324

Gemella	Gemella haemolysans | Oral Taxon 626 | Strain ATCC 10379 | L14326 | 537 | N	100	0/229
Rhodobacteriaceae	Rhizobium loti | Oral Taxon 659 | Strain USDA 3455 | U50166 | 3 | N	86.9	32/244
Bacillus	Bacillus subtilis | Oral Taxon 468 | Strain ATCC 6633 | AB018486 | 46 | N	86	36/257
Clostridiales	Oribacterium sp. | Oral Taxon 108 | Strain FTB41 | AF287770 | 22 | U	89.3	24/224
Granulicatella	Granulicatella adiacens [para-adiacens] | Oral Taxon 534 | Strain TKT1 | AB022027 | 515 | N	97.8	5/229
Actinomyces	Actinomyces sp. | Oral Taxon 170 | Clone AP064 | AF287749 | 36 | U	99.1	2/234
Variovorax	Variovorax paradoxus | Oral Taxon 717 | Strain IAM 12373 | D30793 | 7 | N	97.6	5/206
Streptococcus	Streptococcus australis | Oral Taxon 073 | Strain T1-E5 | AF385525 | 179 | N	99	2/210
Incertae sedis 5	Leptothrix sp. | Oral Taxon 025 | Clone AV011a | AF385528 | 11 | P	95.1	12/243
Fusobacterium	Fusobacterium periodonticum | Oral Taxon 201 | Strain ATCC 33693 | X55405 | 53 | N	100	0/224

OTUs with high global scores associated with the remission state are shown in Table 10:

Table 10: HOMD database output for OTUs of interest (high global scorers). Only the top ranking classification results are shown. Phylotypes represent those associated with an increasing pattern in Figure 23. Identities/ Mismatches are calculated with gaps and AGCTU.

Network Name	HOMD Identification (10.1)	Identities	Mismatch
Porphyromonadaceae	Helicobacter pylori | Oral Taxon 812 | Strain ATCC 43504 | M88157 | 0 | N	87.3	28/220
Bacteroidetes	Catonella morbi | Oral Taxon 165 | Clone MB5_P12 | DQ003629 | 92 | N	79.7	43/212
Bacteria	Prevotella sp. | Oral Taxon 820 | Strain P2A_FAAD4 | AF537212 | 0 | U	84.8	33/217
Wolinella	Tannerella sp. | Oral Taxon 286 | Clone BU063 | AY008308 | 16 | P	88.5	26/227

Using a slightly different scoring approach (ratio comparison) and starting a new run, the whole process of finding OTUs of interest was repeated and yielded the same results for most high scoring phylotypes. The results of this second analysis, which should be seen as validation, can be seen in Table 11.

Table 11: HOMD database output for OTUs of interest (high global scorers). Repetition of analysis that led to Table 9 with altered scoring system and analyzed separately in a new Qiime run.

HOMD Identification (10.1)	Identities	Mismatch
Gemella haemolysans | Oral Taxon 626 | Strain ATCC 10379 | L14326 | 537 | N	100	0/229
Streptococcus mitis bv 2 | Oral Taxon 398 | Strain SK34 | AY005045 | 410 | N	98.2	4/220
Corynebacterium mucifaciens | Oral Taxon 835 | Strain DMMZ 2278 | Y11200 | 0 | N	93.8	15/243
Stenotrophomonas maltophilia | Oral Taxon 663 | Strain LMG 958 | X95923 | 28 | N	97.4	6/228
Neisseria flavescens | Oral Taxon 610 | Strain LNP444 | AJ239280 | 115 | N	99.5	1/220

Propionibacterium acnes | Oral Taxon 530 | Strain 63597 | AF145256 | 70 | N	100	0/247
Microbacterium sp. | Oral Taxon 185 | Clone AV005b | AF385527 | 2 | P	88.1	26/218
Streptococcus australis | Oral Taxon 073 | Strain T1-E5 | AF385525 | 179 | N	99	2/210
Fusobacterium periodonticum | Oral Taxon 201 | Strain ATCC 33693 | X55405 | 53 | N	100	0/224
Bergeyella sp. | Oral Taxon 422 | Clone C4MKM119 | AY278614 | 14 | P	90.3	21/216

Except for two phylotypes (*Microbacterium* and *Bergeyella*) the lists for difference and ratio global scoring were identical. This is due to the relatively stable global scoring system, which emphasizes changing patterns rather than absolute values.

4.3 Discussion

Our investigation of the impact of NCIE on the bacterial community confirmed a number of hypotheses that we initially generated. Firstly, the impact of NCIE on the human gut is evident not only in histological differences of the small bowel tissue. It can also be assessed by a sequence driven community analysis, which shows a major disturbance in the microbiota of the duodenal mucosa. Changes that can be seen during acute enterocolitis are thorough up to phylum level. Notably, certain percentile changes in OTU abundance, as it is the case for *Propionibacterium acne*, number in the 20s to 30s. This implies that the bacterial community gains and loses certain highly abundant phylotypes, which in turn are the apex of a once balanced system. Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria make out a major part of OTUs that change significantly amongst all three patients.

It is still unclear, if the disturbed bacterial gut community in NIEC is caused by a certain agent or a disturbed immune response and a subsequent loss of beneficial bacteria or another mechanism that is unknown up to now. It is possible that a member of the discovered high scoring taxa is actually causing the chronic inflammation which subsequently leds to the prominent phylum level disturbances due to an altered microbial niche. This hypothesis is, however, opposed by the clear differences of high scoring bacteria (as seen in Figure 24). If only one agent was the cause of NCIE, it should be found in a somewhat similar abundance in the acute state of the disease in all three patients, while in remission its numbers should dwindle. *Stenotrophomonas maltophilia* is the only phylotype that meets this general requirement, but shows different abundance changes in all patients.

We therefore propose that NCIE is caused by a misbalance in the bacterial community of the human gut. This effect, which is known as dysbiosis, is characterized by the comprehensive community changes we can also find in enterocolitis patients. Hence, it is not one pathogenic agent but a multitude of bacteria that are responsible for the chronic inflammation, which devastates the gut tissue and leads to an initial misbalance. Driven by these changes, a newly composed gut flora could arise that harbors detrimental features. Hereafter, further displacement of commensal bacteria, like Bacteroidetes or *Faecalibacterium* and subsequent loss of healthy immunologic triggering effects take place. The inflammatory potential is directly influenced by modulation of the ubiquitin-proteasome system in epithelial cells by commensal gut bacteria [22]. The ability of certain beneficial gut microbes to induce anti-inflammatory regulatory T-cells was identified to be vital for the prevention of several chronic disorders [6]. For instance, *Bacteroides fragilis* has been shown to produce a polysaccharide (PSA) which leads to specific proliferation of CD4+ regulatory T cells and subsequent anti-inflammatory IL-10 expression [20]. Without these regulating cycles of activation and deactivation, the immune system enters a state of permanent over-induction without ways to balance itself as commensal bacteria leave the gut-bacterial

system. Once the new microbial community is in place, only extensive antibiotic therapy can levitate the pressure on the host.

The question is how a formally healthy system can change so drastically.

The oral translocation hypothesis

The discovery that high scoring genera could be identified better in the human oral microbiome database than in the more general annotated RDP database led us to another mechanism that we propose as hypothesis in this work. As described above, every part of the human body has a distinct bacterial community, where most members are considered mutualists. This system facilitates a number of reactions, from simple neutral inhabitation to complex host-bacteria interactions like production of hormones and neuro-active substances or degradation of chemical compounds [2,5,19]. Invasive species of bacteria which are usually not associated with a healthy human-bacterial system usually find it hard to out-compete commensal species. An invading pathogenic bacterium that comes from outside the system has to overcome the native microbial flora, which can be considered one of the first lines of defense. In a classic infection, only bacteria that find a growth advantage in some way can maintain a certain threat for the human body. But what would happen if it was not one pathogenic agent but a whole community?

The displacement of parts of the human oral bacterial community, as indicated by the types of species found in acute enterocolitis samples, could play a major role in development of NCIE. Such a translocation event could be triggered by a multitude of factors, may it be diarrhea, a former illness or something as common as teeth brushing with aggressive tooth paste [23]. Once these oral bacteria find a suitable niche within a habitat they usually do not inhabit, community dynamics could lead to local inflammatory reactions. A healthy gut system would balance out bacteria with immune-regulation via IgA, proteolytic enzymes or protection by bile acid [24]. But in contrast to bacteria which can be found outside the human system, these oral bacteria, as they were tolerated to some extend in the oral cavity, could harm the mucosa at this atypical niche. Moreover, immune response would lead to an increasingly suitable environment for the newcomers, as cellular parts and body fluids could provide excellent nutrients.

A schematic proposal of events is shown in Figure 25:

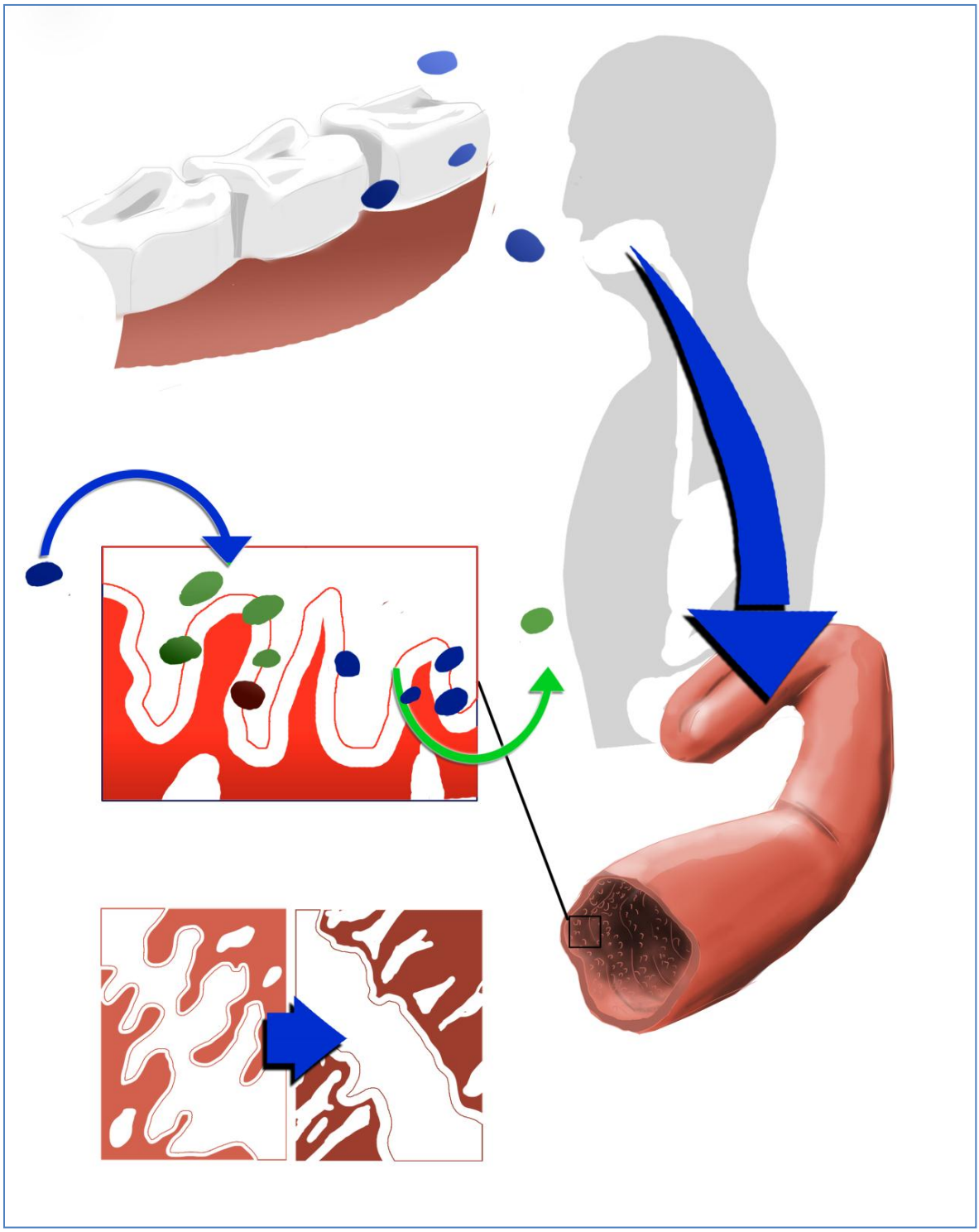

Figure 25: A proposed effective scheme of human oral microbiome translocation: Oral bacteria get swallowed and pass the stomach's acidic environment to settle in locations throughout the small intestine. If events like diarrhea or local inflammation lead to a growth benefit or if they are translocated in a high numbers, commensal gut bacteria (green) get displaced by oral bacteria (blue). This leads to further immunologic responses and increased inflammation, further promoting the dysbiosis in the system. Once a chronic inflammation status is reached, the affected tissue has very thoroughly altered its appearance- Cavities and intestinal villi close up which leads to the classic "flat" gut surface that can be observed in coloscopy. A reason for this change could very well be the minimization of surface that is available to the now detrimental bacterial gut community.

Adjusting to these new conditions, a formerly commensal community could switch to "disease enhancing" as anatomic constraints are weakened. A similar translocation event can be observed during "bacterial overgrowth syndrome", a severe complication in a number of gut related diseases and immune deficiency disorders, where the bacterial flora of the small gut is "overgrown" or displaced with communities usually associated with the colon [24].

Oral communities feature less variation amongst different humans than other habitats [4]. The similar behavior of certain OTUs amongst patients could therefore indeed be an indicator for a less variable bacterial system getting access to a system of high variability and functional redundancy. Thirty percent of human oral microbial OTUs are classified as Proteobacteria, with *Neisseria* found most often but is virtually depleted of Bacteroidetes [25]. Thus, an oral community "used to" a high number of Proteobacteria and low numbers of Bacteroidetes could find itself benefiting from washing out events during inflammation.

Nutrients and species of interest.

While certain bacteria indeed receive a boost in growth, it is Bacteroidetes that decline in NIEC. Bacteroides are responsible for the majority in the colon [26] and a rise of a competing community could lead to their repression during acute enterocolitis. Additionally, Bacteroides have been observed to actively induce gene expression in human host cells, indicating a tight communication link between bacterial and human system [20]. As in case of diarrhea, Bacteroides could play a pivotal role in community restructuring as their large set of genes could makes them more flexible under adverse (or opportunistic) conditions.

Antibiotic treatment leads to a decrease of butyrate producing genera like *Faecalibacterium* or *Clostridium* in the colon [19], further amplifying nutrient competition in the gut. Importantly, butyrate is also a major nutrient source for GI epithelial cells [3].

It is known that certain genera of Proteobacteria are highly effective in exploiting the iron rich living conditions provided by blood [10]. Furthermore, it could be shown, that it is mostly Enterobacteria which require a readily available iron source for proliferation and virulence, while many commensals, as *Lactobacillus*, do not. The supplementation of iron led to competition advantages for pathogenic strains on cost of "barrier" genera [10]. Therefore, tissue damage, which is accompanied by loss of iron into the gut lumen, could in fact cause major proliferation of these species rather than their removal.

Propionibacterium acnes was found to be one of the possible causative agents for granulomatous colitis [27]. The family Fusobacteriaceae is known to efficiently ferment amino acids, a source of nutrients that are also increased in inflamed guts [26]. *Fusobacterium nucleatum* has been associated with inflammatory diseases in the oral cavity and is common in mucosa isolates from the gut in IBD cases [28]. It features a wide range of abilities that would make it very effective as co-inductor of adverse NCIE effects: readily attachment to a wide range of cell types would enhance its capacity to form bridgeheads of infection in the

gut, enhanced compatibility of aggregation with other bacterial species and production of tissue irritating substances [28] would boost changes in the bacterial community structure. As for *Propionibacterium*, it could be shown that *Fusobacterium* is associated with IBD [29]. In addition to that, Fusobacteria are also producing butyric acid, which, in excess is a toxin for mucosal cells which induces apoptosis and ulceration [29]. *Stenotrophomonas maltophilia* has been associated with neonatal infections and bacteraemia [30]. *Streptococcus mitis* is known for a highly effective epithelial attachment system, translocation via the bloodstream and the causative agent for a form of endocarditis [31]. It is highly antibiotic resistant and expresses an Immunoglobulin A protease, which would make it a perfect "first settler" for infectious sites in the gut.

As stated before, nearly all of the global high scoring bacteria in NCIE (see Table 9) are common members of the oral microflora and/or associated with disease states in the oral cavity. Their appearance mimics the appearance of oral plaques, which led to the biofilm hypothesis in inflammatory bowel diseases that we generated over the course of the study.

Biofilms and antibiotic treatment

The following paragraph is a short summary of the excellent review on oral biofilms written by Filoche et al. [32] with additional remarks regarding its connection to chronic enterocolitis. According to the ecological plaque hypothesis, the oral bacterial community can switch its behavior – from more mutualism driven to pathogenic states. The development of oral biofilms usually starts with a triggered adhesion to the tooth or surrounding tissue. It could be shown by Ritz and colleagues that streptococci belong to the first to colonize epithelial surfaces, followed by *Actinomyces* (high scoring in NCIE) and gram negative filamentous bacteria that attach to the "bridgehead" colonies. Over time, a biofilm will evolve, whose appearance and composition is tightly linked to the host health state and the biofilms specific location in the oral cavity. If human nutrition is heavily based on sugars, acidogenic bacterial species flourish and the whole community enters a pathogenic state – which is commonly known as dental caries. These pathogenic biofilms are constructed amongst others by *Streptococcus mutans*, *Streptococcus mitis* as well as *Rothia* and *Actinomyces* species. Similar species alterations can be found in paradontosis, where *Fusobacterium nucleatum* and *Porphyromonas* sp. contribute to the disease state. In recent years, a theory has evolved that describes the formation of biofilms not only as a one by one occupation event by different species but actual attachment of free moving ("planktonic") bacterial aggregates to the oral surfaces with subsequent colonization of a local spot. The concerted action of a bacterial "multicellular" structure could indeed protect the colonizing bacteria from host immune reaction that single cells would not be able to overcome. Apart from attachment and biofilm evolving, the detachment of cells from the aggregate has become a major area of research. Active detachment could very well be facilitated by a bacterial community in order to avoid nutrition shortage or adverse conditions (aggressive tooth paste, frequent use of mouth fresheners, smoking).

The moving aggregate could find its way into the human gastrointestinal system by swallowing. Upon its arrival in the stomach, its matrix and layered architecture would prevent the annihilation of viable microbes by the acidic gastric environment and protect it until arrival in the duodenum or subsequent parts of the bowel. With an attachment event in the gut, various displacements lead to the formation of a viable translocated community that has been described in this chapter. Due to repeated attachment and detachment events, the patchy inflammatory appearance could occur, which is so typical for chronic inflammatory diseases of the gut. The large variety of group genome content in a biofilm could very well be one of the reasons for the ability of the oral microbial community to outcompete the functionally potent Bacteroidetes or Firmicutes group in the gut. Alternatively,"biofilm priming" could also be facilitated by bacteria like *Streptococcus mitis* and its migration to the gut via the bloodstream to be later enriched with oral bacteria via plaques or blood vessel translocation.

The biofilm hypothesis is also backed up by the resilience of chronic bowel diseases to antibiotic treatment.

When treated with antibiotics, patients show decreased gut concentrations of propionate, butyrate and isobutyrate [26]. Woodmansey *et al*. could also observe that numbers of *Staphylococcus, Streptococcus* and *Enterococcus* in feces increase under antibiotic treatment. This rise was accompanied by higher abundance of *Propionibacterium acnes*. In the oral microbiome, it could be observed that upon antimicrobial treatment, the inner layers of a biofilm community were sheltered and would be the source for reformation when the antibiotic treatment was ceased [32]. These aggregates could function as a source for new inflammation and could be the reason for the immediate onset of acute inflammation NCIE patients suffer after antibiotic deduction. The effects of the antibiotic intake, necessary to levitate symptoms of enterocolitis, could therefore lead to further dysbiosis themselves. Antibiotic treatment has already been linked to a number of chronic diseases like asthma and atopic disease [19], underlining the necessity for a responsible use.

5. Concluding remarks

We showed that diarrhea, a very common and usually harmless condition of the human body, could indeed lead to major disturbance in the gut microbial community with phylum level changes and washing out of important commensal taxa. Such an event could be one of the priming factors for the development of IBD, as the free competitive space if quickly filled up by opportunistic pathogens, one of them being the large group of Proteobacteria. This rampaged system is susceptible to a number of adverse environmental conditions and is more sensitive to community reshaping by bacteria normally rare in the Gi microbial community.

We believe that the source for the onset and severe progression of NCIE are microbes of the oral bacterial community. Bridgeheads of translocations are formed by either blood vessel transport or swallowing of biofilm particles and subsequent attachment (by pioneering species like *Streptococcus mitis* or members of the *Actinomyces* genus) and colonization of the gut. After this initial priming phase and first inflammatory reactions of the host, the bacteria receive a growth advantage with further enhancement of biofilm diversity and pathologic potential. The role of butyrate and other nutrients in this event remains to be examined, as it can either act as toxin or essential cell nutrient. A shortage of butyrate in the early stages of IBD development due to displacement of the native flora could be followed by the subsequent rise of functional redundant (butyrate producing) bacteria like *Fusobacterium*, which could lead to misbalance, overproduction and cellular damage in the long run. Displacement of commensal bacteria, like members of the Bacteroidetes phylum takes place and the functional diversity of the gut microbial system gets impeded. Eventually, an altered bacterial community has established itself in the gut, accompanied by the IBD characteristic patchy inflammation where translocation has occurred. Antibiotic therapy in combination with anti-inflammatory treatment attacks biofilms and helps reconstitute a native epithelial structure. With it, formerly outcompeted native bacteria gain in numbers and the community enters a state of remission again. As therapy will not be able to get rid of all biofilm particles, there will always exist sources for repeated colonization, as indicated by the severe relapses upon antibiotic treatment stop.

These findings underscore the importance of a multi-level approach when examining host-bacterial systems. Data that can be acquired from formerly poorly described systems will have to be supplemented by functional analysis and causative modes of metabolic co-existence in the future. Next and next – next generation sequencing methods will allow for a more in depth analysis of the microbial agents associated with humans and enhancing spatial distinguishability between communities will help resolving the dependencies in these systems. With the progression of the Human Microbiome Project, we can hope for the generation of baseline parameters in bacterial-host systems, which would serve as guidelines for further studies. Despite these advances, we are only at the verge of describing

the system of biological agents and their human hosts, e.g. the large area of viral-human interactions is virtually not explored on a community level.

Nonetheless, there has been a paradigm change in medical life sciences in the last couple of years. Focusing on a limited number of parameters (a small number of species, genomes or an effector molecule) is important to understand basic biological rules and direct relationships. There are, however, limitations to this approach, as the complex network of microbes with its hosts often acts indirectly and system wide. Bacterial communities could indeed be linked over a large number of effector pathways and body parts. The events started in one of the habitats found in and on human bodies are not isolated but interchange information in various ways. The path that participants of this living space take and the results on human health and wellbeing are dependent on a large multitude of factors, beginning from host immune state to commensal bacterial community structure, location, nutrition, diseases and their treatments, environmental and geographical parameters or psychological conditions. Rather than forming direct causative chains of effectors, they all belong to a network of interdependency and concerted reactions forming the human-bacterial system.

Therefore, understanding the network of human-bacterial interactions would bring us closer to not only battling adverse conditions, but getting new insights into one of the fundamental questions of mankind: what actually defines a human being.

References

[1] Tschöp H.M., Hugenholtz P., Karp C.L. : « **Getting to the core of the gut microbiome** « *Nature Biotechnology*, Vol. 27/4, 344-346, (2009).

[2] Turnbaugh P.J., Ley R.E., Hamady M., Fraser-Liggett C.M., Knight R., Gordon J.I. : « **The Human Microbiome Project** ». *Nature*, Vol.449, 804-810, (2007).

[3] Eckburg P.B., Bik E.M., Bernstein C.N., Purdom E., Dethlefsen L., Sargent M., Gill S.R., Nelson K.E., Relman D.A. : "**Diversity of the Human Intestinal Microbial Flora**". *Science* Vol.308, 1635-1638, (2005).

[4] Costello E.K., Lauber C.L., Hamady M., Fierer N., Gordon J.I., Knight R.: "**Bacterial Community Variation in Human Body Habitats Across Space and Time**". *Science*, Vol.326, 1694-1697, (2009).

[5] Eckburg P.B., Bik E.M., Bernstein C.N., Purdom E., Dethlefsen L., Sargent M., Gill S.R., Nelson K.E., Relman D.A. : "**Diversity of the Human Intestinal Microbial Flora**". *Science* Vol.308, 1635-1638, (2005).

[6] Ehlers S., Kaufmann S.H.E.: "**Infection, inflammation, and chronic diseases: consequences of a modern lifestyle**". *Trends in Immunology*, Vol.31/5, 184-190, (2010).

[7] Bercik P, Denou E, Collins J, Jackson W, Lu J, Jury J, Deng Y, Blennerhassett P, Macri J, McCoy KD, Verdu EF, Collins SM: "**The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice**". *Gastroenterology*, Vol.141/2, 599-609, (2011).

[8] Turnbaugh P.J., Quince C., Faith J.J., McHardy A.C., Yatsunenko T., Niazi F., Affourtit J., Egholm M., Henrissat B., Knight R., Gordon J.I.: "**Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins**". *PNAS Early Edition* (www.pnas.org/cgi/doi/10.1073/pnas.1002355107), 1-6, (2010).

[9] 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., Egholm M., Henrissat B., Heath A.C., Knight R., Gordon J.I.: "**A core gut microbiome in obese and lean twins**". *Nature* (Epub – doi:10.1038/nature07540), 1-6, (2008).

[10] Zimmermann M.B., Chassard C., Rohner F., Goran E.K.N., Nindjin C., Dostal A., Utzinger J., Ghattas H., Lacroix C., Hurrell R.F.: "**The effects of iron fortification on the gut microbiota in African children: a randomized controlled trial in Côte d`Ivoire**". *Am J Clin Nutr*, Vol.92, 1406-1415, (2010).

[11] Ronaghi M., Uhlén M., Nyrén P.: "**A sequencing method based on real-time pyrophosphate**". *Science*, Vol.281, 363-365, (1998).

[12] Wu G.D., Lewis J.D., Hoffmann C., Chen YY, Knight R., Bittinger K., Hwang J., Chen J., Berkowsky R., Nessel L., Li H., Bushman F.D.: "**Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags**". *BMC Microbiology*, Vol.10:206, 1-14, (2010).

[13] Caporaso J.G. , Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R.: "**QIIME allows analysis of high-throughput community sequencing data**.", *Nature Methods,* Vol.7: 335–336, (2010)

[14] http://qiime.sourceforge.net/scripts/pick_otus.html (09.03.2011)

[15] Hamady M., Knight R.: ”**Microbial community profiling for human microbiome projects: Tools, techniques, and challenges”.,** *Genome Res*, Vol.19: 1141-1152, (2009)

[16] Cole J.R., Chai B., Marsh T.L., Farris R.J., Wang Q., Kulam S.A., Chandra S., McGarrell D.M., Schmidt T.M., Garrity G.M., Tiedje J.M.: “**The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy**”. *Nucleic Acids Research*, Vol.31/1, 442-443, (2003).

[17] Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T.: **"Cytoscape: a software environment for integrated models of biomolecular interaction networks**". *Genome Res*, Vol.13, 2498–2504, (2003).

[18] Gorkiewicz G., Thallinger G.G., Trajanoski S., Stocker G, Lackner S., Hinterleitner T., Gülly C., Högenauer C.: “**Alterations of the colonic microbiota in response to osmotic diarrhea**”. – in submission

[19] Dethlefsen L., Huse S., Sogin M.L., Relman D.A.: ”**The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing**”. *PLOS Biology*, Vol.6/11, 2383-2400, (2008)

[20] Wexler H.M.: ”**Bacteroides: the Good, the Bad, and the Nitty-Gritty**”. *Clinical Microbiology Reviews*, Vol.20/4, 593-621, (2007)

[21] Chen T., Yu W-H, Izard J., Baranova O.V., Lakshmanan A., Dewhirst F.E.: ” **The Human Oral Microbiome Database: a web accesible resource for investigating oral microbe taxonomic and genomic information**”. *Database*, Vol.2010, ArticleID baq013, (2010).

[22] Reiff C., Kelly D.: ”**Inflammatory bowel disease, gut bacteria and probiotic therapy**”. *International Journal of Medical Microbiology*, Vol.300, 25-33, (2010).

[23] Singhal S., Dian D., Keshavarzian A., Fogg L., Field J.Z., Farhadi A.: ”**The Role of Oral Hygiene in Inflammatory Bowel Disease**”. *Dig Dis Sci*, Vol.56, 170-175, (2011).

[24] Bures J., Kohoutova D., Förstl M., Rejchrt S., Kvetina J., Vorisek V., Kopacova M.: ”**Small intestinal bacterial growth syndrome**”. *World J Gastroenterol*, Vol.16/24, 2978-2990, (2010).

[25] Lazarevic V., Whiteson K., Huse S., Hernandez D., Farinelli L., Osteras M., Schrenzel J., Francois P.: “**Metagenomic study of the oral microbiota by Illumina high-throughput sequencing**”. *Journal of Microbiological Methods*, Vol.79, 266-271, (2009).

[26] Woodmansey E.J., McMurdo M.E.T., Macfarlane G.T., Macfarlane S.: “**Comparison of Compositions and Metabolic Activities of Fecal Microbiotas in Young Adults and in Antibiotic-Treated and Non-Antibiotic-Treated Elderly Subjects**”. *Applied and Environmental Microbiology*, Vol.70/10, 6113-6122, (2004).

[27] Wada R, Kawamura C, Inoue S, Watanabe K, Kaimori M, Yagihashi S.: “**Granulomatous colitis associated with botryomycosis of Propionibacterium acnes**”. *Arch Pathol Lab Med*, Vol.125/11, 1491-1493, (2001).

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

[29] Ohkusa T., Okayasu I., Ogihara T., Morita K., Ogawa M., Sato N.: ”**Induction of experimental ulcerative colitis by Fusobacterium varium isolated from colonic mucosa of patients with ulcerative colitis**”. *Gut*, Vol.52/1, 79-83, (2003).

[30] Basu S., Das P., Roy S., De S., Singh A.: "**Survey of gut colonisation with Stenotrophomonas maltiphilia among neonates**". *J Hops Infect*, Vol.72/2, 183-185, (2009).

[31] Mitchell J.: "**Streptococcus mitis: walking the line between commensalism and pathogenesis**". *Molecular Oral Microbiology*, Vol.26, 89-98, (2011).

[32] Filoche S., Wong L., Sissons C.H.: " **Oral Biofilms: Emerging Concepts in Microbial Ecology**". *J Dent Res*, Vol.89/1, 8-18, (2010)

Acknowledgments

I would like to thank

Ass.Prof. Dr.med.univ. Gregor Gorkiewicz for introducing me to the beautifully complex field of microbiomics

Dr. Slave Trajanoski for his support in various matters, many of them bioinformatical

the “Center for Medical Research” (ZMF) Graz and its head Mag. Dr.rer.nat. Christian Gülly for providing experience, nutrition and working space.

the “Core Facility Molecular Biology” of the ZMF Graz, led by Dr.in rer.nat Ingeborg Klymiuk for helping me become practical rather than theoretical.

Gabriele Michelitsch, Martina Hatz, Theresa Maierhofer , Carina Fischer, Karin Wagner, Karin Meister and Gerald Richter for being colleagues and friends.

Claire Fraser-Liggett, Ph.D., head of the Institue for Genome Sciences (IGS) in Baltimore for showing me a new dimension of science.

W. Florian Fricke, Ph.D. for his kindness, his knowledge of orioles and my call to Baltimore

the Graz University of Technology (TU Graz) for being a splendid Alma Mater and a place of both, professionalism and creativity.

Univ.-Prof. Dipl.-Biol. Dr.rer.nat. Gabriele Berg for her advice in studies and life

and

my family and friends.