

Community and genomic analysis
of the human small intestine microbiota

Bartholomeus van den Bogert

Thesis committee

Promotor

Prof. Dr M. Kleerebezem

Personal chair in the Host Microbe Interactomics Group
Wageningen University

Co-promotor

Dr E.G. Zoetendal

Assistant professor, Laboratory of Microbiology
Wageningen University

Other members

Prof. Dr T. Abee, Wageningen University

Dr J. Doré, National Institute for Agricultural Research, INRA, Jouy en Josas, France

Prof. Dr P. Hols, Catholic University of Leuven, Belgium

Dr A. Nauta, FrieslandCampina Research, Deventer, The Netherlands

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Community and genomic analysis
of the human small intestine microbiota

Bartholomeus van den Bogert

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr M.J. Kropff,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Wednesday 9 October 2013

at 4 p.m. in the Aula.

Bartholomeus van den Bogert

Community and genomic analysis of the human small intestine microbiota,

224 pages

PhD thesis, Wageningen University, Wageningen, NL (2013)

With references, with summaries in Dutch and English

ISBN 978-94-6173-662-8

Summary

Our intestinal tract is densely populated by different microbes, collectively called microbiota, of which the majority are bacteria. Research focusing on the intestinal microbiota often use fecal samples as a representative of the bacteria that inhabit the end of the large intestine. These studies revealed that the intestinal bacteria contribute to our health, which has stimulated the interest in understanding their dynamics and activities. However, bacterial communities in fecal samples are different compared to microbial communities at other locations in the intestinal tract, such as the small intestine. Despite that the small intestine is the first region where our food and intestinal microbiota meet, we know little about the bacteria in the small intestine and how they influence our overall well-being. This is mainly attributable to difficulties in obtaining samples with the small intestine being located between the stomach and the large intestine. Therefore, the work in this thesis aimed at providing a better understanding of the composition and dynamics of the human small intestinal microbiota and to provide insight in the metabolic potential as well as immunomodulatory properties of some of its typical commensal inhabitants. Small intestinal samples used in the work described in this thesis were collected from ileostomy subjects, individuals that had their large intestine surgically removed and the end of the small intestine connected to an abdominal stoma, providing access to luminal content of the small intestine.

Considering the importance of molecular techniques in contemporary ecological surveys of microbial communities, first of all, two technologies, barcoded pyrosequencing and phylogenetic microarray analysis were compared in terms of their capacity to determine the bacterial composition in fecal and small intestinal samples from human individuals. As PCR remains a crucial step in sample preparation for both techniques, the use of different primer pairs in the amplification step was assessed in terms of its impact on the outcome of microbial profiling. The analyses revealed that the different primer pairs and the two profiling technologies provide overall similar results for samples of fecal and terminal ileum origin. In contrast, the microbial profiles obtained for small intestinal samples by barcoded pyrosequencing and phylogenetic microarray analyses differed considerably. This is most likely attributable to the constraints that are intrinsic to the use of the microarray to enable the detection of predefined microbiota members only, which is due to the probe design that is largely based on large intestinal microbiota communities. However, the pyrosequencing technology also allows for identification of bacteria that were not in advance known to inhabit our intestinal tract.

The pyrosequencing technology was used as the method of choice to study the total and active small intestinal communities in ileostoma effluent samples from four different subjects through sequencing the 16S ribosomal RNA gene (rDNA) and ribosomal RNA (rRNA) content combined with metatranscriptome analysis by Illumina sequencing of cDNA derived from enriched mRNA of the same sample set to

investigate the activities of the small intestinal bacteria. The composition of the small intestinal bacterial communities as assessed from rDNA, rRNA, and mRNA patterns appeared to be similar, indicating that the dominant bacteria in the small intestine are also highly active in this ecosystem. *Streptococcus* spp. were among the bacterial species that were detected in each ileostoma effluent sample, albeit that their abundance varied greatly between samples from the same subject as well as samples from different subjects. *Veillonella* spp. frequently co-occurred with *Streptococcus* spp., indicating that the *Streptococcus* and *Veillonella* populations play a prominent role in the human small intestine ecosystem and their co-occurrence suggests a metabolic relation between these genera.

Therefore, cultivation and molecular typing methodologies were employed to zoom-in on these groups, which revealed that the richness of the small intestinal streptococci strongly exceeded the diversity that could be estimated on basis of 16S rRNA analyses, and could be extended to the genomic lineage level (anticipated to resemble strain-level). From ileostoma samples 3 different *Streptococcus* species were recovered belonging to the *S. mitis* group, *S. bovis* group, and *S. salivarius* group, which could be further divided in 7 genomic lineages. Notably, the *Streptococcus* lineages that were isolated displayed distinct carbohydrate utilization capacities, which may imply that their growth and relative community composition may respond quite strongly to differences in the dietary intake of simple carbohydrates over time. This notion is in good agreement with the observation that the *Streptococcus* lineage populations fluctuated in time with only one *Streptococcus* lineage being cultivated from both ileostoma samples collected in a one-year time frame. Conversely, the cultivated *Veillonella* isolates from samples during that same time-interval consistently encompassed a single lineage. Furthermore, this *Veillonella* lineage could be isolated from both the oral cavity as well as the ileostoma effluent. Analogously, three *Streptococcus* lineages that belong to a single phylotype also appeared to be present in bacterial communities from the oral cavity as well as the small intestine. These observations suggest the representatives of the *Veillonella* and *Streptococcus* genera that are encountered in the oral and small intestinal microbial ecosystems are closely related and indicate that the oral microbiota may serve as an inoculum for the upper GI tract.

The metabolic capacity of 6 small intestinal *Streptococcus* lineages, that were obtained from a single ileostoma effluent sample, was further investigated through the determination of genomic sequences of these lineages. The small-intestinal *Streptococcus* genomes were found to encode different carbohydrate transporters and the necessary enzymes to metabolize different sugars, which was in excellent agreement with what carbohydrates could be used by representative strains of the *Streptococcus* lineages.

To further our understanding how the different streptococci as representatives of the dominant small intestinal bacterial populations may influence our immune system, human dendritic cells were stimulated with strains of the different *Streptococcus*

lineages to study their immunomodulatory properties. The *Streptococcus* lineages differed significantly in their capacity to modulate cytokine responses of blood-monocyte derived immature dendritic cells. As *Streptococcus* and *Veillonella* frequently co-occur in the small intestinal ecosystem, pair-wise combinations of strains of these two species were also tested for their combined immunomodulatory properties. This resulted in considerably different cytokine responses as those that could be predicted from the stimulations with either *Streptococcus* or *Veillonella*, indicating that it is not trivial to predict gut mucosal associated immune responses and that the composition of the intestinal microbiota as a whole may have a distinct influence on an individual's immune status.

In conclusion, the work described in this thesis provides an expansion to the accumulating knowledge on the human intestine microbiota. Whereas most studies focus on the microbiota present in the distal regions of the intestinal tract, this study targeted the microbiota of the poorly proximal regions of the intestine and also addressed its capacity to interact with the local mucosal tissue. The data presented here can be exploited to guide the design of future studies that aim to elucidate the interplay between diet, microbiota and the mucosal tissues in the human small intestinal tract.

Table of contents

Chapter 1	General introduction and thesis outline	1
Chapter 2	Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples	21
Chapter 3	Congruency of phylogenetic composition and activity patterns in the small intestine	53
Chapter 4	Diversity of human small intestinal <i>Streptococcus</i> and <i>Veillonella</i> populations	87
Chapter 5	Comparative genomics analysis of <i>Streptococcus</i> isolates from the human small intestine reveals their adaptation to a highly dynamic ecosystem	111
Chapter 6	Immunomodulatory properties of <i>Streptococcus</i> and <i>Veillonella</i> isolates from the human small intestine microbiota	147
Chapter 7	General discussion	165
Appendices		
	References	186
	List of publications	204
	Co-author affiliations	205
	Samenvatting	206
	Acknowledgements – Dankwoord	209
	About the author	211
	Overview of completed training activities	212

Chapter 1

General introduction and thesis outline

A modified version of this chapter was published as:

Bartholomeus van den Bogert*, Milkha. M. Leimena*, Willem. M. de Vos, Erwin G. Zoetendal, and Michiel Kleerebezem. 2011. Functional Intestinal Metagenomics, p. 170-190. In F.J. de Bruin (ed.), Handbook of Molecular Microbial Ecology, Vol II: Metagenomics in Different Habitats. Wiley-Blackwell.

*These authors contributed equally to this work

This chapter also contains information adapted from:

Erwin G. Zoetendal*, Jeroen Raes*, Bartholmeus van den Bogert, Manimozhiyan Arumugam, Carien C. Booijink, Freddy J. Troost, Peer Bork, Michiel Wels, Willem M. de Vos, and Michiel Kleerebezem. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J, 2012. 6:1415-1426.

*These authors contributed equally to this work

Abstract

The mammalian gastrointestinal (GI) tract is inhabited by a myriad of microbes, dominated by bacteria. Different GI tract locations of the human gut harbor distinct bacterial communities, which increase in density and diversity along the longitudinal axes. The collective microbial community, also termed microbiota, is known to contribute to host health, which has stimulated the interest in understanding its dynamics and activities. Large-scale and in-depth characterization of the intestinal microbial community composition requires application of (high-throughput) 16S rRNA gene-based technologies. Further insights into the functional capabilities of the intestinal microbiota can be obtained through (functional) metagenomic approaches. Comprehensive interpretation of the data from these approaches requires advanced bioinformatic and systems biology approaches to decipher the role of the intestinal (micro)biological system in relation with human health and disease. This review describes recent developments and applications for quantitative and qualitative determination of the microbiota community composition in GI tract samples, as well as culture-based and metagenomic strategies to unravel their functional properties. This chapter ends with an outline of the thesis chapters that follow the General Introduction.

Introduction

The human gastrointestinal (GI) tract is inhabited by microorganisms that as a whole are referred to as the gastrointestinal microbiota (115, 203). Although the presence of Eukarya (289) and Archaea (74, 80) in the human GI tract has been reported, their absolute numbers as well as their diversity are relatively low. In addition, meta-analysis revealed a complex viral community of 1,200 genotypes in human feces obtained from adult subjects (34).

The highly diverse bacterial community residing in the human GI tract (Figure 1.1) is dominated by phylotypes belonging to the *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* (15, 80, 268). Although *Firmicutes* are found in the intestine of all mammals, each mammalian species harbors a distinct microbial composition (205). Clustering of the mammalian fecal microbiota composition based on dietary habits established distinct grouping of carnivores, omnivores and herbivores, which can be roughly characterized by increasing microbiota diversity, respectively. This indicates that bacterial diversity is at least partially mediated by the host's diet composition. Furthermore, the fecal microbiota composition could also be correlated to host phylogeny and intestinal morphology (e.g. fore- and hindgut herbivores), establishing the role of these factors in the determination of intestinal microbiota composition. Based on these observations it is not surprising that the human intestinal microbiota resembles that of omnivorous primates (205, 206).

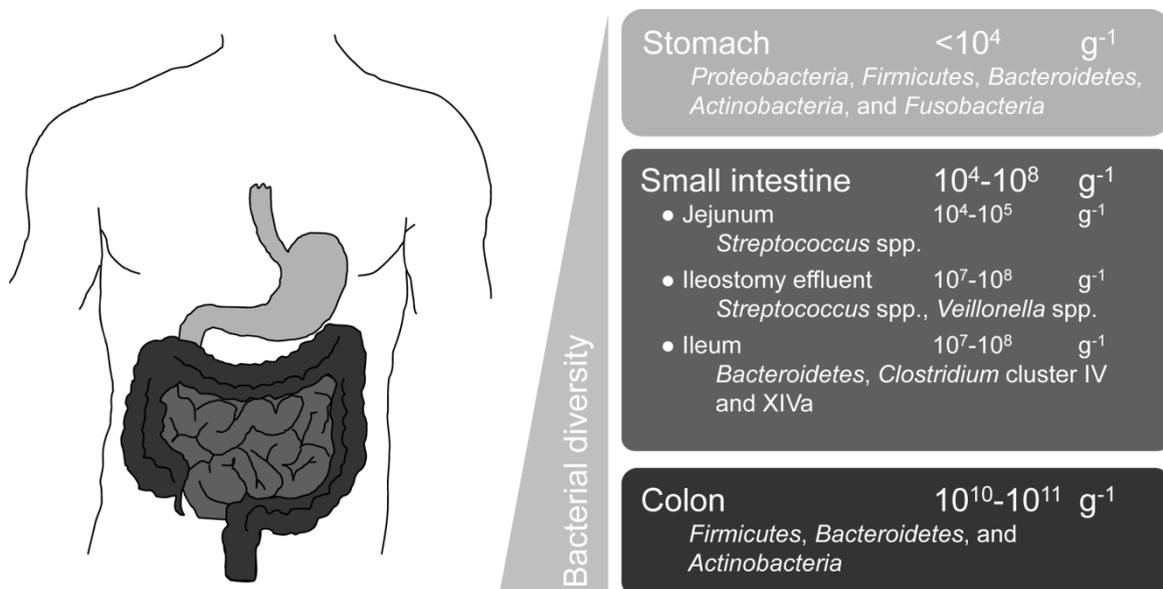


Figure 1.1. Schematic representation of the predominant microbiota composition and density at different regions in the human GI tract.

To eventually understand the factors that are involved in determination of intestinal microbial community composition will require advanced methodologies to assess this composition at phylogenetic as well as functional level. Here we discuss the recent developments in the qualitative and quantitative strategies to determine the composition of the inhabitants of the GI tract as well as their functional properties.

The emerging insights into the human intestinal microbiota in relation to health and disease may eventually provide microbiota-based diagnostic markers. Moreover, such insight may open avenues towards the rationalization of diet or microbial intervention strategies to prevent or treat human and animal diseases via modulations of the intestinal microbiota composition and/or function.

Characterization of the gastrointestinal microbiota

Up to two decades ago, studying microbial communities and identifying individual microbial inhabitants as well as their biological and physiological characterization depended on the ability to culture bacteria (8). *Escherichia coli*, originally named *Bacterium coli* by its discoverer Theodor Escherich, was the first bacterial isolate obtained from the human GI tract (95). Together with improved anaerobic culturing techniques by Hungate (148, 265), rich media facilitate cultivation of a plethora of bacteria species. However, most bacteria cultivated on these media are fast growing and belong to the readily cultivable fraction of complex microbial communities. Selective culture media enable the enrichment of the cultured fraction for bacteria of interest by hindering growth of unwanted bacteria. This cultivation strategy has been successfully used to isolate bacteria such as *Streptococcus* ((43, 340); Chapter 4), *Veillonella* (274, 275), *Enterococcus*, *Lactobacillus* (5), and *Bifidobacterium* (91) from environmental samples. This approach also led to the isolation of *Akkermansia muciniphila* from feces using anaerobic medium with gastric mucin as the carbon source (67). In addition, targeted cultivation can yield bacterial isolates of distinct bacterial groups, such as butyrate producers (78). Recently, gnotobiotic mice transplanted with bacterial communities from a human host were used to enrich for community members that are more suited to a certain diet. This microbial community could be clonally archived in a multi-well format to create collections of a subjects' microbiota for further phylogenetic and genetic analysis (110). Although cultivation approaches are time-consuming, laborious, and costly, analysis of obtained isolates yields an important source of knowledge on their physiological characteristics (67, 340, 380), functional properties through (comparative) analysis of their genomes (25, 175, 343) as well as provides the opportunity to study their immunomodulatory properties (233, 335, 336). During the last decades hundreds of gastrointestinal isolates were obtained (see (265) for a review), albeit that most of them originate from stool samples and the number of isolates from the upper GI tract is limited.

Recent innovations in the area of microbial cultivation, such as anopore-based microdish culture chips (152), may yield micro-colony isolates of previously uncultured bacteria. Microdroplet-based approaches (see (142) for a review) offers the means to compartmentalize cells in micro-droplets for subsequent (clonal) growth of micro-colonies and isolation of bacteria, including rare and so far uncultured phylotypes (374). Furthermore, confinement of multiple cells within the microdroplets offers avenues to investigate cell–cell interactions and facilitate isolation of obligate syntrophic bacteria (253, 374).

Despite these efforts to obtain cultured representatives of the intestinal microbiota, molecular technologies revealed that the majority of bacteria in the intestinal

ecosystem cannot readily be cultured under laboratory conditions (309, 382) and, therefore, there is a clear need for advanced culture-independent techniques for comprehensive characterization of the human intestinal microbiota.

Most culture-independent approaches are based on the universal bacterial phylogenetic marker, 16S ribosomal rRNA (rRNA) or its encoding gene. A variety of 16S rRNA approaches has been developed, including classical cloning and sequencing, DGGE/TGGE, FISH, and qPCR. These classical methods differ in terms of sensitivity, selectivity and phylogenetic resolution and have been employed to determine bacterial-community structures, or detect and/or quantify specific bacterial groups within a variety of samples derived from the human intestine (see (378) for a review). Over the past years, the 16S rRNA targeting methodologies for microbial profiling of ecosystems have evolved rapidly and technologies, such as phylogenetic microarray analysis and barcoded pyrosequencing, were developed to enable high-throughput and in-depth characterization of the microbiota (see (120) for a review).

Phylogenetic microarrays commonly contain 16S rRNA-targeting oligonucleotide probes immobilized on a carrier surface (in many cases glass slides) and enable high-throughput, semi-quantitative characterization of microbial communities (68, 251, 266). They have shown to be of value for assessing the composition and population dynamics of the microbiota (see (382) for a review). Phylogenetic array data-processing is not trivial and requires robust mathematical and statistical procedures to prevent artifacts (120). In addition, phylogenetic arrays are constrained to the detection of phylogenetic groups that are represented in the array design (252), which is illustrated by the incomplete coverage of the small intestinal microbiota by arrays designed on basis of predominantly fecal sequences (31). This shortcoming may be overcome by the use of flexible array design platforms (266), or by sequence based *de novo* community profiling technologies like barcoded illumina sequencing (39, 50) and 454 pyrosequencing (9). The latter technologies allow high-throughput analyses, provide phylogenetic resolution, and are not restricted to 16S rRNA gene sequences that are known beforehand. Importantly, deep pyrosequencing and phylogenetic microarray analysis of the microbial community of fecal samples generated comparable results (49, 339). Data interpretation of illumina sequencing and pyrosequencing is challenging due to the vast number of sequences obtained and requires stringent quality control of individual sequence reads (149) and effective sequences profiling tools, as well as advanced visualization and interpretation software suites to handle the datasets.

Moreover, the amount of the generated sequences determines the level of detection and for any quantification to be performed requires a sequence excess of the required depth. Sequence based community profiling technologies and any other PCR based approach, as well as FISH, require universal or selective oligonucleotide primers and probes for the bacterial group targeted, which can be based on the conserved and variable regions of the corresponding 16S rRNA gene sequences (42). However, the continuous expansion of 16S rRNA gene databases tends to outdate previously designed primers and/or probes, requiring constant updating of these molecular tools (16). This notion is exemplified by surveys (16, 138) and our

unpublished assessments of published universal 16S rRNA primers, demonstrating profound differences in their coverage of *Bacteria* (Figure 1.2) and/or other specific bacterial groups. Moreover, PCR based technologies that target highly conserved genes like the 16S rRNA gene may suffer from the formation of chimeric amplicons as an intrinsic artifact during amplification, which may contaminate the databases with biologically irrelevant sequences that are falsely assigned to specific bacterial groups (12). Therefore, it is imperative to screen obtained 16S rRNA gene sequences using appropriate tools (13, 82, 139) to identify chimeras and exclude them from analysis and deposition to the DNA databases.

While illumina based community profiling needs further optimization (50), pyrosequencing efforts in the recent years have made a considerable contribution in comprehensive characterization and comparison of microbial communities in a variety of body habitats (147), including the human oral cavity (169), throat, stomach (9), as well as in fecal samples (49, 69, 329). It is expected that illumina platform in the upcoming years will complement these compositional studies with greater depth of analysis due to the vast amount of sequence reads that are obtained with this technology.

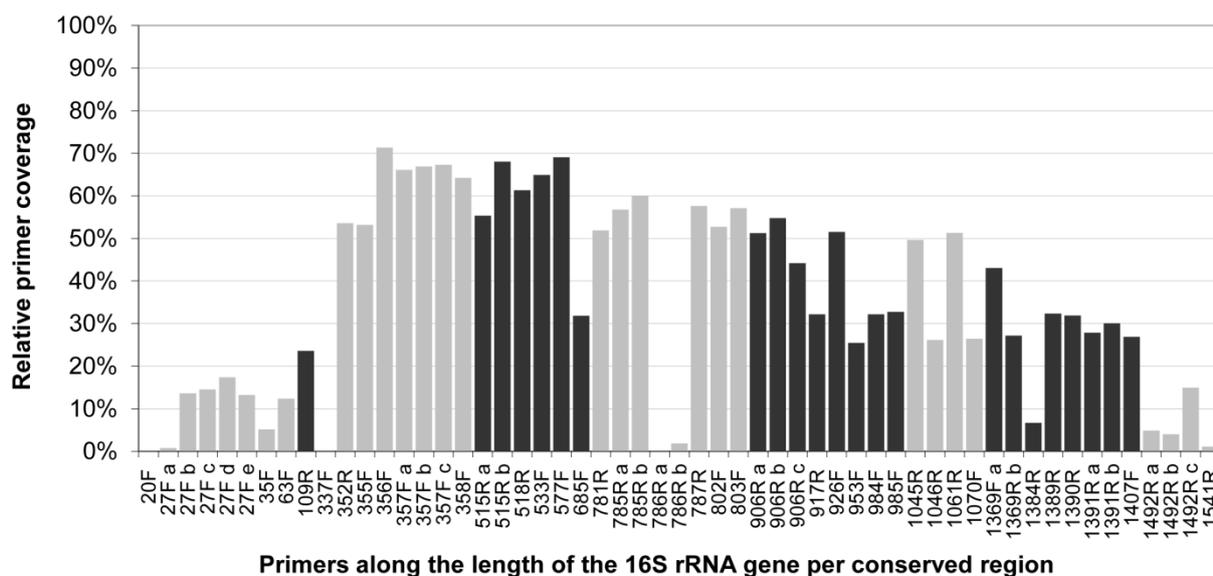


Figure 1.2. Overview of relative coverage of *Bacteria* by published universal 16S rRNA primers.

The human gastrointestinal microbiota

The bacterial community is not evenly distributed in the GI tract and increases in density along the longitudinal axes. Bacterial colonization density in the human stomach is low, i.e., less than 10^4 bacteria per gram of contents, which is due to the stringent acidic conditions that kill most bacteria (115, 203). Additionally, the diversity of the stomach microbiota is relatively low, featuring merely 128 phylotypes in a collection of gastric biopsy specimens obtained from 23 human individuals (23).

Bacterial numbers increase along the small intestine from 10^4 bacteria in the duodenum and jejunum to approximately 10^8 bacteria per gram of contents in the ileum (115, 203). Despite its relatively low community density (relative to the large

intestinal microbiota, see below), this microbial community may play an important role in the bioavailability and conversion of dietary components, especially since the major part of food digestion and absorption occurs in the small intestine. Moreover, the small intestine is the dominant intestinal site to study interactions between the intestinal microbiota and the host immune system, due to the presence of unique elements that are involved in confinement of the microbiota to the intestinal lumen and luminal sampling of bacteria (see (75) for a review). Paneth cells, for example, secrete antimicrobial peptides (22), which diffuse in the mucus layer produced by goblet cells. This layer controls contact between the luminal microbiota and the underlying epithelium (75). Other important elements, primarily localized to the small intestine, include Peyer's patches that are involved in sampling luminal bacteria through overlying Microfold cells (M cells) (75, 217, 242). Bacteria are transported via transcytosis to the peyer's patches and loaded on dendritic cells (DC) that play a key role in bacterial handling and subsequent immune responses (see (55) for a review). Otherwise, there is evidence that DCs can sample luminal bacteria directly by establishing tight-junction-like structures with intestinal epithelial cells and directly sample bacteria by protruding dendrites outside the epithelium (270).

Considering the significance of the processes taking place in the small intestine, its microbiota can be anticipated to have an important influence on host physiology (381). However, the small intestinal microbiota is relatively unexplored, attributable to sampling difficulties of this poorly accessible region of the GI tract. Therefore, elucidation of this bacterial community is imperative for a better understanding of the microbial interactions and metabolic processes as well as the microbiota-immune system interactions that occur in the small intestine (32, 203). Efforts to study the small intestinal microbiota are dependent on biopsy specimens obtained during (emergency) surgery (3) or samples collected from sudden death victims at autopsy (128). Microbial analysis of biopsies from the jejunum and the distal ileum revealed a relatively low bacterial diversity in the jejunum mucosa with a microbial community dominated by *Streptococcus* spp. while a predominance of *Bacteroidetes* and *Clostridium* clusters IV and XIVa (according to the phylogeny proposed in (54)) was identified for the distal ileum (354). However, as a consequence of the relatively extensive procedures required for obtaining these samples, they may not represent the true small intestinal microbiota of a healthy individual and do not provide insights into population dynamics (32).

One alternative to obtain small intestinal samples, which circumvents the sampling difficulties associated with the small intestine, makes use of individuals that underwent colon resection due to cancer or inflammatory bowel disease (IBD) and as a result have the terminal ileum connected to a stoma. This stoma provides a unique opportunity to non-invasively and repetitively sample the contents of the terminal ileum (31, 32). A recent study indicated that the microbiota in the effluent samples from these ileostomy subjects do not represent that of the terminal ileum in healthy subjects due to the penetration of oxygen (125). This study, however, seems to contradict with recent findings by Booiijk, et al. that showed high abundance of strict anaerobes in ileostoma effluent samples (31). This may be explained by the fact that

the ileostomy subjects in the latter study had their colon removed at least 5 years before enrolment in the study during which anaerobic conditions in the small intestine are restored. The study by Booijink, et al. revealed that the small intestinal microbiota is different from that of the lower GI tract with a less diverse microbial community and greater temporal fluctuations in composition (31). Furthermore, the small intestinal microbial communities are enriched with *Streptococcus* and *Veillonella* (31, 381). In addition to the *Streptococcus* spp. encountered in the small intestine other members of the Lactic Acid Bacteria are detected, albeit that their inter-individual abundances fluctuate considerably (see (308) for a recent review). This is most likely due to the variability in available nutrients as a consequence of food consumption (308, 340). To assess the resemblance of ileostoma effluent to the true small intestinal microbiota, preliminary investigations in our laboratory employed an orally introduced catheter to collect small intestinal samples. This revealed that the microbial communities were enriched in *Streptococcus* and *Veillonella* (belonging to Bacilli and *Clostridium* cluster IX, respectively) in jejunal and proximal-ileum samples, while *Bacteroidetes* and *Clostridium* cluster XIVa were dominating in the more distal- or terminal-ileum (Figure 1.3), resembling the microbiota in ileostoma effluent and the colon, respectively.

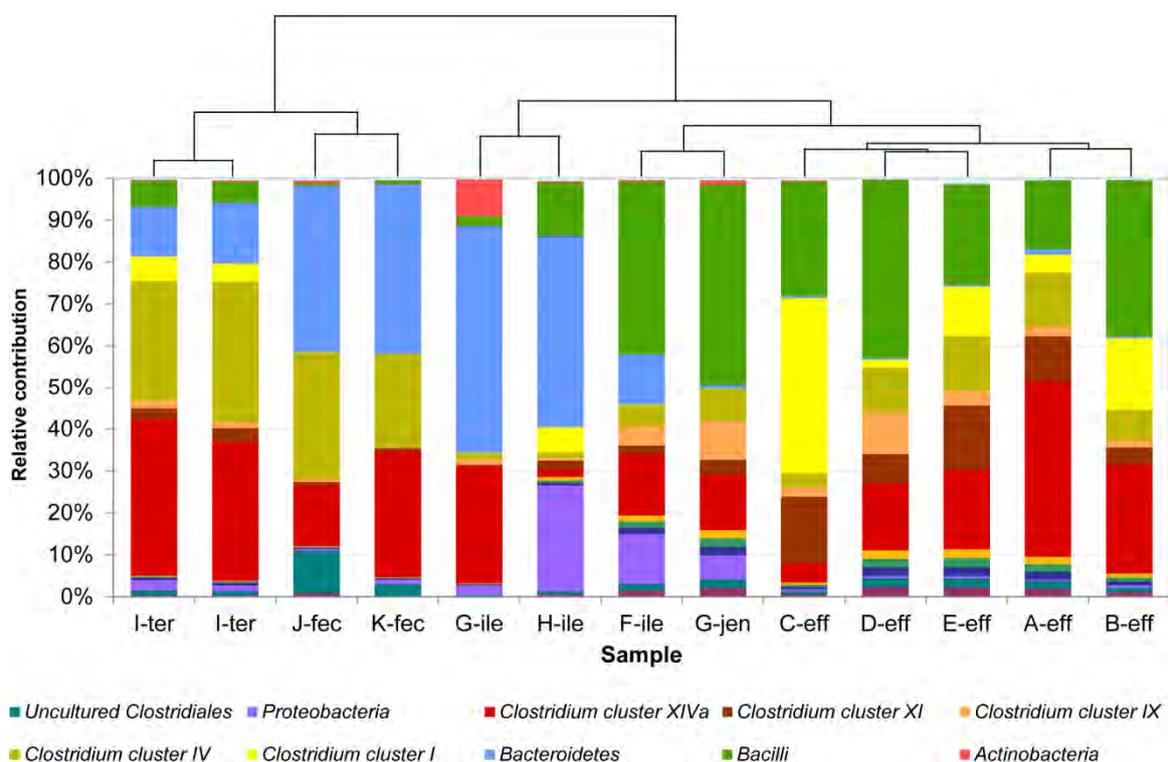


Figure 1.3. Relative contribution of detected phylogenetic groups present in samples derived from the small intestine and feces of four and two healthy individuals, respectively, and those from five healthy ileostomists. Profiles were generated by phylogenetic microarray analysis using the Human Intestinal Tract Chip (HITChip) (266). The tree represents the Euclidian clustering of the HITChip probe profiles. A–K encode the subjects; ileostoma effluent (eff), jejunum (jen), ileum (ile), terminal ileum (ter), feces (fec). In the legend, phylogenetic groups that contribute at least 2.5% to one of the profiles are indicated.

These results suggest that ileostoma effluent microbiota is not a suitable reflection of the community present in the lumen of the terminal ileum, but resembles more the community that is normally encountered in the proximal regions of the small intestine (381). One plausible explanation for this observation is the absence of colonic refluxes in the terminal ileum of ileostomy subjects.

Bacterial numbers increase tremendously in the large intestine, reaching 10^{11} bacteria per gram of contents (32, 203, 378). Often fecal samples are used to study the microbiota in the colon (151). However, it should be noted that the bacterial composition in these samples can differ significantly from that of the colonic mucosa (80, 201, 383). Analogous to the small intestine, the large intestine appears to contain a unique microbiota per person (382). In contrast to the small intestinal microbiota, the large intestinal microbiota of healthy individuals seems to remain relatively stable over time (229, 266, 376), and is much more diverse as compared to that of the small intestinal, with an estimated complexity that encompasses at least 500 phylotypes (80).

Homeostasis between the human host and the intestinal microbiota

A healthy human host and its intestinal microbiota coexist in a homeostatic relationship (Figure 1.4) (135, 203, 217). The intestinal microbiota benefits from a stable environment and nutrient supply that are provided in the intestinal tract, while the host gains products from microbial metabolic activity and protection against potential pathogens (115, 203, 217, 242). The human host is only able to utilize simple sugars, disaccharides, and starch. The benefits of metabolic activity of the intestinal microbiota on the host consist in fermentation of the remaining non-digestible dietary carbohydrates (136). The products of fermentation are gases (CO_2 , CH_4 , and H_2) and the Short Chain Fatty Acids (SCFA) that are dominated by butyrate, propionate, and acetate, with butyrate being the most preferred energy source of the colonic epithelium (136, 203, 366). Approximately 10% of the total energy requirement for humans is derived from SCFA (136, 203). In addition, vitamins K, B_{12} , riboflavin, biotin, folic acid, and pantothenate are synthesized by the intestinal microbial population. (115, 136, 282).

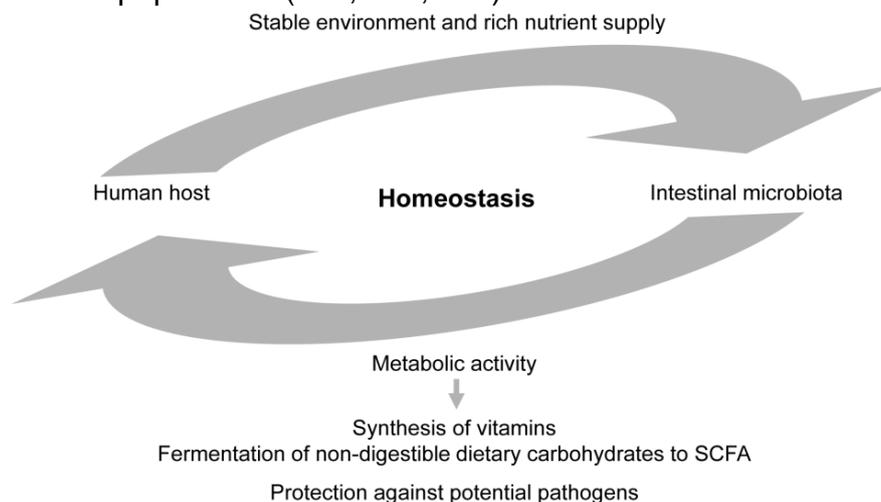


Figure 1.4. Homeostatic relationship between the human host and its intestinal microbiota.

This functional uniformity of the microbiota in individuals is in apparent contradiction with their distinct phylogenetic composition. Nevertheless, it has been suggested that a phylogenetic core may be present, which may account in part for this common metabolic activity. However, the definition of a phylogenetic core varies between studies and is dependent on analysis factors, such as the techniques employed to determine the composition of intestinal microbiota, subject health status, phylogenetic depth, and the prevalence threshold (48, 120, 156, 262, 284, 317, 329). The phylogenetic core does not appear to include universally conserved phylotypes (317, 329), but can be defined on basis of phylotypes (total of only 66 OTUs) prevalent in a large proportion of the subjects studied (more than 50 %), and accounting for a large fraction of the overall community (~35 %) (317). Analogously, fecal microbiota profiling using phylogenetic microarrays, revealed a set of responding probes that were shared among the individuals (266). Inversely, the vast majority (~80 %) of the detected phylotypes appears to be host specific (317).

Contrary to the individual-specific phylogenetic composition of the intestinal microbiota, recent random sequencing analyses of microbiota-derived DNA illustrated that functional grouping of sequence reads displayed remarkable conservation between individuals. Again, supporting the concept of a functional core rather than a phylogenetic core within the human intestine microbiota (188, 329). Furthermore, the study by Arumugam, et al. shows a remarkable high level of congruency in the total and core functional composition of fecal microbial communities among subjects of distinct geographic origins (11).

Besides being a mutualistic partner, the human intestinal microbiota has also been associated with human diseases. IBD and Irritable Bowel Syndrome (IBS) are relatively common intestinal disorders with unclear etiology. Although these disorders are multifactorial, recent studies suggest that they are associated with a deregulation of the homeostasis between the host and its intestinal microbiota (75, 135, 217). Moreover, an increasing number of diseases appears associated with an aberrant microbiota composition, including allergic (atopic) diseases, obesity, type I and II diabetes, and autism (203, 265, 322). The mechanisms that underlie the observed associations between health and the GI tract microbiota are not yet understood. Nevertheless, the observed associations suggest that the intestinal microbiota may be an important modulator of our overall well-being and could be a potential diagnostic and/or therapeutic target in human health and disease. This notion is exemplified by a recent study by Vrieze, et al. showing that infusion of male patients with metabolic syndrome with the intestinal microbiota from lean individuals leads to an improvement in insulin sensitivity (352).

Functional analysis of intestinal of microbial communities

Although 16S rRNA gene-based technologies assess the (relative) number and identity of microorganisms in microbial communities, they do not provide further insights into their functional properties. Functional insight into the microbiota can be achieved by studying the genomes of bacterial isolates. The genomics era was initiated with the complete sequencing of the genome of *Haemophilus influenza* in

1995 (96). Since then there has been an exponential increase in the number sequenced bacterial genomes (24, 179) with over 3000 currently (<http://www.ncbi.nlm.nih.gov/genomes>; as on October 15th, 2012) available in public databases. While initially employing 'classical' Sanger sequencing technologies, the genomics field was drastically accelerated by the development and implementation of next generation sequencing platforms such as 454 (Roche), SOLEXA (Illumina), SOLiD (Applied Biosystems), and Helicos (Heliscope), which produce continuously increasing amounts of sequence information at a constantly decreasing price per base (see (234) for a recent review).

Initially, bacterial genomics efforts focused on clinically relevant bacteria, such as *Salmonella* (254), *Streptococcus pyogenes* (92), and *Helicobacter pylori* (324), aiming to improve our understanding of their virulence factors. More recently, the scope has expanded to bacterial genomes of isolates from other food or intestinal sources, including the food microbes that are considered to confer a health benefit to their host when they reside in the intestinal tract like *Lactobacillus* (175) and *Bifidobacterium* species (see (176) for a recent review). Furthermore, the Human Microbiome Project (<http://www.hmpdacc.org>) currently sequences the genomes of cultured and uncultured bacteria from different human body habitats and plans to sequence a total of 3000 genomes, making a considerable contribution to the number of bacterial genomes (145). With the staggering number of sequenced genomes that are currently available, comparative genomics provides the means to understand the evolutionary relationship between bacteria (166, 344), albeit that these analyses may to some extent be confounded by lateral gene transfer between bacterial species (179).

Although genomics has provided a wealth of knowledge, this field of study depends heavily on having pure bacterial cultures as starting material for DNA extraction and sequencing. As most bacteria are not readily cultivable (see above), their genetic content is barely accessible for culture dependent genome sequencing approaches, thereby limiting whole genome analysis of these organisms. To this end, a lot of effort is dedicated to the development and use of strategies that enable single cell genomics (47, 367), which circumvent cultivation, but to date remain technologically highly challenging and generally provide a relatively poor coverage of the single cell genomes that are determined. Nonetheless, single cell genomics may provide molecular clues that could enable the design of culture conditions that could selectively enrich for so far uncultured bacteria.

Genomic analysis of the microbial community inhabiting a common environment can be achieved by applying metagenomics, which also bypasses the need for isolating and cultivating the individual species (44, 122). Metagenomic libraries of microbial communities contain information about the functional capabilities and the phylogenetic distribution of an ecosystem (123). Depending on the research purpose, metagenomic libraries can be built from short or long insert fragments. For a metagenomic library with short insert sizes (1-10 kb), harsh DNA extraction methods that shear the DNA can be used. However, depending on the microbial complexity of the environmental sample, a large number of clones is required to obtain adequate

1

coverage of the complete metagenome. Short insert libraries can be used for individual gene-function screening, e.g., identifying specific enzymatic activities (130). Methods for high molecular weight DNA extraction (21, 130, 276) enable the construction of large insert libraries (20-100 kb), which favors the discovery of multiple genes or operons that confer specific functional pathways to the cloning host. In addition, these large inserts may contain more reliable phylogenetic markers within the same insert as compared to the confidence of phylogenetic assignment that can be achieved using small sequence inserts (123). To prevent expression of unfavorable components in the cloning host, which is frequently *Escherichia coli*, many of the metagenomic expression libraries employ inducible promoters and are based on low copy number vectors (123). Still, cloning bias may significantly interfere with the appropriate coverage of potential counter-selected genetic elements that are encompassed within the metagenome of the targeted ecosystem. To circumvent the cloning step, next generation sequencing technologies (see above) are employed that allow for direct sequencing of environmental DNA extracts (296, 365). The added advantage of using next generation sequencing technologies is that the obtained sequence abundance data enables uncovering of the taxonomic and functional composition within an ecosystem (11, 329).

Screening of a metagenomic library can be grouped into two main strategies: sequence-based screening and function-based screening (32, 122, 123, 291). Sequence-based screening employs large scale sequence determination and mining to unravel (niche specific) gene functions in order to describe the microbial community residing in the environment in gene-function terms (291), while function-based screening mainly focuses on the identification of sequences that represent a particular function by screening for such function using a heterologous expression host (Figure 1.5).

Worldwide, several metagenomic-initiatives target the human intestinal microbiota and/or other niches associated with the human body to eventually uncover the complete human microbiome (258, 330). The MetaHIT consortium, for example, provided an extensive gene catalogue (> 3 million non-redundant genes) of the human gut microbiota on basis of deep sequencing of the DNA extracted from fecal samples from more than 120 individuals (262), while a complementary and much smaller metagenome catalogue has recently been published for the microbiota of the human small intestine (381). Furthermore, the MetaHIT consortium also reported that the fecal microbiome composition of human individuals can be grouped in three distinct clusters called enterotypes, which is irrespective of their geographic origin and may eventually support stratification approaches of human individuals to predict their differential responsiveness to diet and medical treatments (11). The Human Microbiome Project (146) examined the microbial composition on multiple body sites in a large cohort of healthy human individuals to investigate the 'normal' microbiota and found that each body site harbors a distinct microbial community that remains stable over time (147). Nonetheless, functional analysis of samples from the same cohort revealed that there appears to be a small core of low abundant metabolic modules across different body habitats (2).

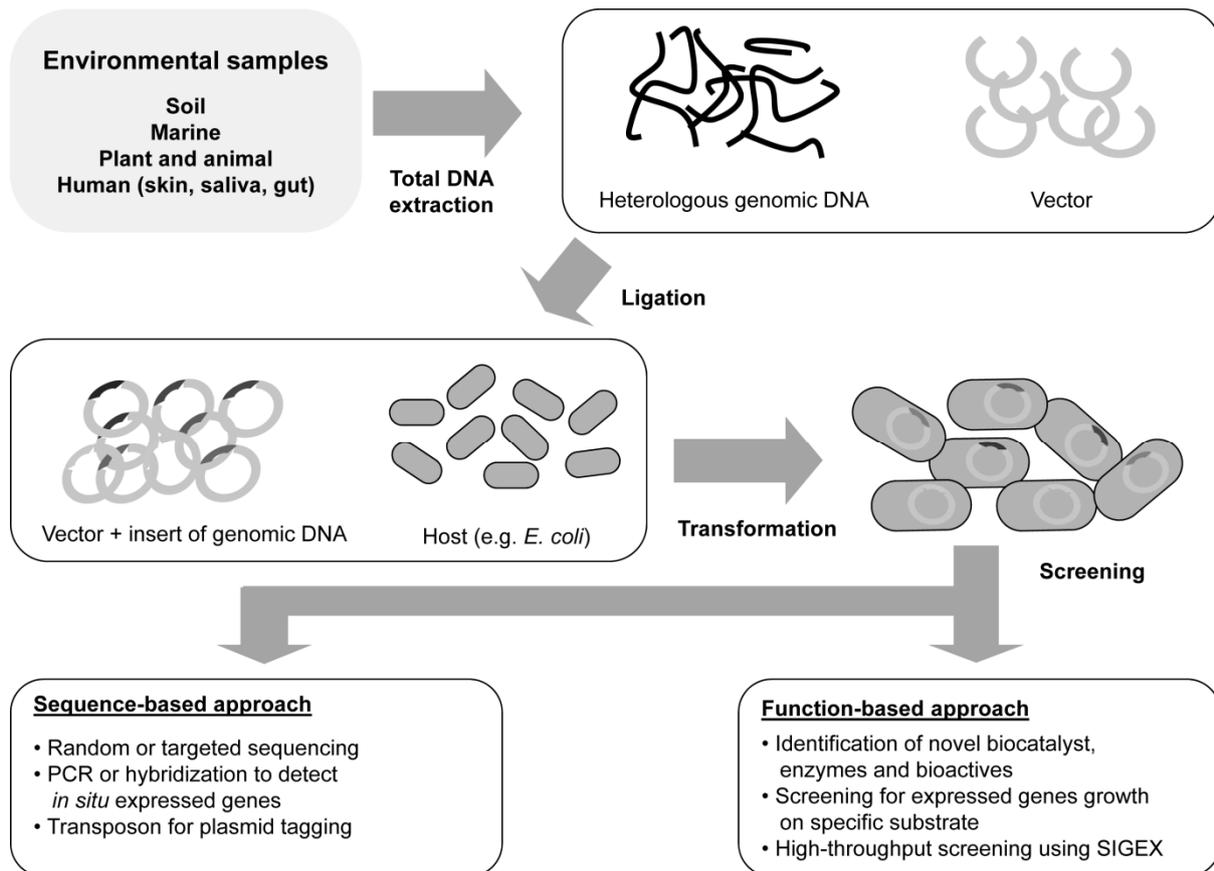


Figure 1.5. Schematic representation of metagenomic library construction and screening.

To elucidate the activity of microorganisms in their environment, a coupling of genotypic and phenotypic meta-analyses will be required. These include other functional metagenomic approaches, such as metatranscriptomics, metaproteomics and metabolomics, which use biomolecules (RNA, protein, and metabolites) of activity as targets (Figure 1.6).

Metatranscriptomics approach

Metatranscriptomics encompasses the (complete) transcriptome analysis of a microbial community, or of gene-subsets within that community, in a particular environmental niche (356). To detect expressed genes in environmental samples, a rapid, robust and direct RNA isolation method is required to overcome instability problems associated with bacterial RNA. For fecal material, several methods have been described that include the use of quenching protocols to limit RNA turn-over and degradation or employ chemicals like RNA-later to stabilize the microbial RNA present in the sample (377), yielding relatively good RNA quantity and quality. Total bacterial RNA extracted from the environment predominantly consists of ribosomal RNA (> 95%), while only a minor proportion reflects actual transcript activity (mRNA, commonly below 5% of total RNA). Therefore, enrichment steps have been developed to remove the bacterial ribosomal RNA or selectively capture the mRNA. The first is based on ribosomal RNA removal by selective hybridization to probes that target highly conserved regions of the ribosomal RNA and is available in readily

applicable commercial format (29), while the second method uses a selective and progressive 5' – 3' exonuclease reaction to digest rRNA on the basis of its 5' monophosphate terminus, which is absent in unprocessed mRNAs (356). The latter method requires multiple steps and as a consequence is technically challenging (29). Capture of mRNAs is enabled by polyadenylation of mRNAs using *Escherchia coli* poly(A)polymerase. Polyadenylated mRNAs can subsequently be captured using oligo(dT)probes or reverse transcribed using oligo(dT) primers (304).

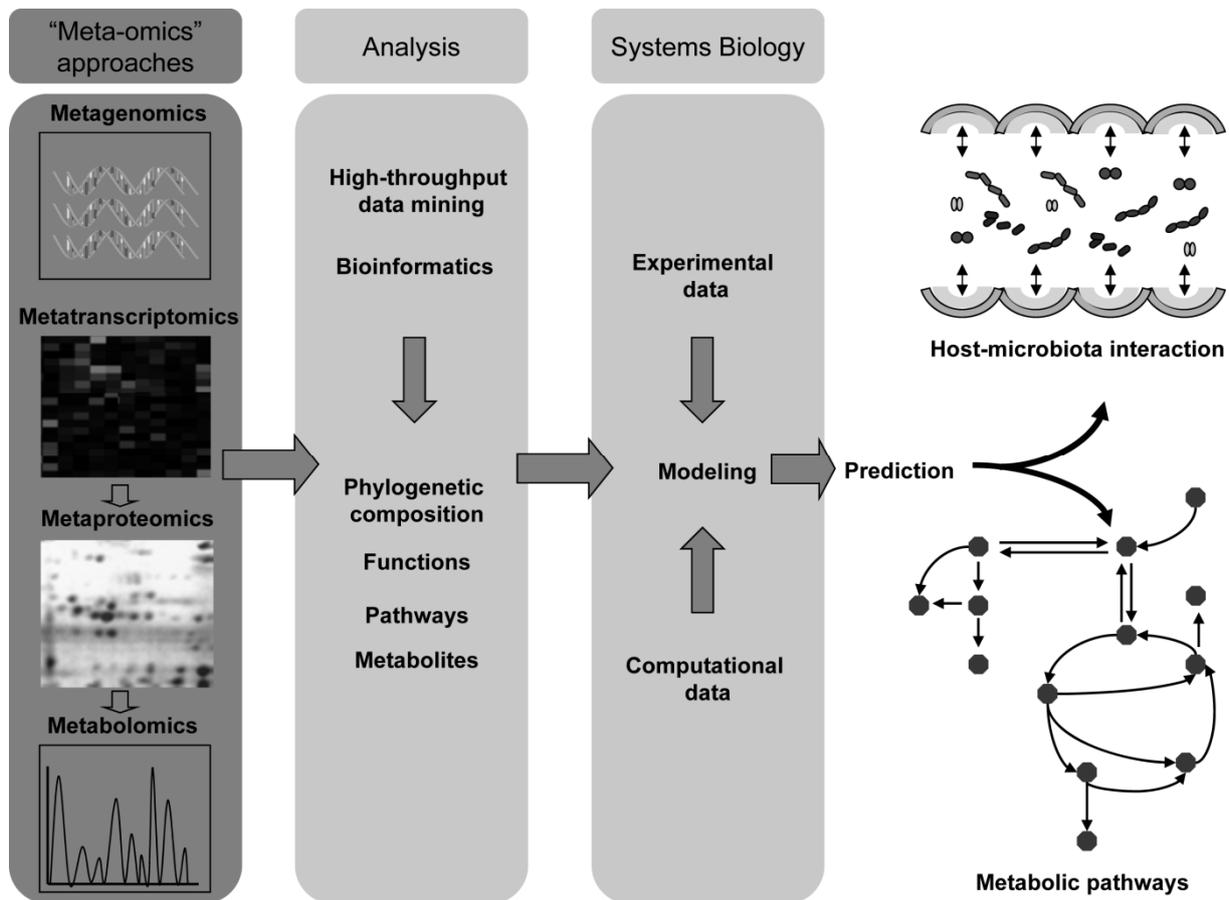


Figure 1.6. Combining “meta-omics” technologies with systems biology to study and model host-microbe interactions.

Transcriptome technologies for single organisms are well established and commonly use DNA microarrays that are designed on basis of the genome sequence and its annotation (61). As a consequence of the genetic diversity in many microbial communities, microarray design for complete metatranscriptome analyses is very challenging or even impossible. Nevertheless, gene catalogs of a certain ecosystem like the GI-tract microbiota gene-catalogue (262) may enable the design of ecosystem transcriptome analysis using high-density microarray designs (129). Alternatively, dedicated arrays may be designed that target specific gene-families of interest. To allow for comprehensive and quantitative data interpretation, such arrays require a probe design that covers the sequence-space of the targeted genes within the microbial community of interest, or may exploit highly conserved gene-sequence

elements that can detect clusters or families of genes corresponding with a particular function.

Sequencing of mRNA-derived cDNA has become an effective alternative for microarray based methods (218), especially enabling detection of transcript abundance in a metagenome pool of unknown composition. Development of next generation sequencing platforms (see above) produce an enormous abundance of short reads enabling more cost- and time-efficient transcriptome sequencing (87, 157, 238). Comparative transcriptomic sequencing for single microorganism such as *Lactobacillus plantarum* (198), *Escherichia coli* (323), *Bacillus anthracis* (255), and *Burkholderia cenocepacia* (371) under different growth conditions elucidate the response of an organism to its niche. Tartar, et al. (319) performed the first metatranscriptomic analysis in the more complex ecosystem of host-symbiont gene expression in the termite gut using an expressed sequence tag (EST) approach. Since then metatranscriptomic analysis using RNA sequencing of mRNA-derived cDNA clone libraries is employed to elucidate the active microbial populations in marine environments (99, 105), soil (17, 333), and the human GI tract (112, 331, 381).

Conversion of RNA into cDNA might introduce multiple biases that could interfere with the characterization and quantification of the transcriptome (52). Therefore, more accurate and high-throughput transcript analyses may be obtained by direct RNA sequencing (DRS), using only femtomoles of RNA without the need for cDNA synthesis (249). An optimized RNA sequencing protocol using HeliScope single-molecule sequencers has been presented for *Saccharomyces cerevisiae* as a model organism, employing polyadenylated and 3'-blocked RNA capture on a solid surface followed by labeling and sequencing (249, 318). Although DRS has only been demonstrated for a eukaryotic organisms to date (249, 294), it may provide avenues for analysis of microbial metatranscriptomes in complex environmental samples at high-throughput, low cost, and without bias.

Metaproteomics approach

Metaproteomics encompasses the study of the proteome produced by the community in a certain environment, addressing presence, relative abundance, and/or modification state of the community proteome. This approach may become a very powerful strategy for understanding overall microbial-ecosystem functioning (364). Classical proteome analyses employ two-dimensional (2D) gel electrophoresis (247), but this technology is not straightforward, labor intensive, and does not allow high-throughput analysis. Development of high-efficiency chromatographic peptide separation coupled with advanced mass spectrometry (MS) using improved ionization procedures, overcame initial difficulties in protein identification. This development in combination with advanced bioinformatics tools for reverse genetics and protein identification, including post-translational modifications, has paved the way for high-throughput protein identification using MS information (221).

Metaproteomic studies of microbiota in the human infant GI tract by combining 2D gel electrophoresis and matrix assisted laser desorption ionization-time of flight mass

1

spectrometry (MALDI-TOF MS) (173) demonstrated the application of this technique for complex microbial communities. Comparison with the database revealed a peptide sequence with high similarity to bifidobacterial transadolase. The development of novel high-throughput, gel-free (shotgun), non-targeted mass spectrometry (MS; see for a (298) recent review) enables a high level of protein identification. The principle is based on separation of peptides generated by enzymatic digestion using liquid chromatography and direct infusion into rapid scanning tandem mass spectrometers (2D-LC-MS/MS) through electrospray ionization followed by sequence comparison to the database. This approach has, for example, been used to detect and identify all proteins in the human distal gut microbiota without gel-based separation and revealed a high abundance of microbial proteins involved in translation, energy production and carbohydrate metabolism. Moreover, microbial activity interpretation in functional pathways demonstrated a complex interplay between the human host and its associated microbes (348).

To improve the accuracy of peptide identification, Rooijers, et al. demonstrated an iterative workflow with a combination of a synthetic metagenome, from known gut inhabitants, and metagenomic databases as reference for MS/MS spectra. Using this approach, the proteome of *A. muciniphila*-like bacteria in a fecal metaproteome was explored, identifying the predicted role in mucin degradation of these bacteria in the intestinal tract (277). Recently, Kolmeder and colleagues developed a new high-throughput metaproteomics analysis pipeline. Application of this pipeline to study the fecal metaproteome of three healthy individuals revealed that the metaproteome is host specific and remains relatively stable over time, analogous to what is observed for the intestinal microbiota composition (see above) (178).

For complex protein mixtures or limited protein samples, Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) improves metaproteomic analysis by providing high accuracy mass measurement, with unprecedented sensitivity, and resolution power (208). This technology recently enabled comprehensive analysis of the human breast milk proteome, identifying over 300 proteins that could be grouped to different functional categories. Despite the high accuracy with FTICR-MS, low abundant polypeptides were not characterized, indicating that the human milk proteome is not fully elucidated (259).

Metabolomics – metabonomics approach

Metabolite profiling can be divided into two areas, namely metabolomics and metabonomics. Metabolomics aims to characterize and quantify small metabolite molecules in complex biological samples, while metabonomics broadly aims to measure the global, dynamic metabolic response of multicellular biological systems to environmental factors or genetic manipulations (244) (Figure 1.6). Metabolomic and metabonomic studies employ the same analytical procedures which can be divided into targeted and non-targeted approaches. Targeted approaches focus on known metabolites while non-targeted approaches investigate all possible metabolites, covering changes in core metabolome as well as detection of unknown metabolites (248). The analytical platforms used in these studies, include nuclear

magnetic resonance (NMR) spectroscopy and gas chromatography-MS platforms, which are suited for mapping global biochemical changes in a non-targeted approach, and/or LC-electrochemistry arrays (LCECA) for mapping neurotransmitter pathways and oxidative stress pathways in both targeted and non-targeted studies (161). Variation in human metabolic phenotypes is associated to several factors, such as host genotype, age, sex, lifestyle, nutrition and their commensal microbial communities (248). Correlation analysis of symbiotic gut microbiota composition with the variation in metabolic phenotypes measured in human fecal and urinary samples of seven Chinese individuals suggested that there are profound host-microbiota symbiotic associations. These associations have an important influence on the global metabolism, regardless of the genetic background across a range of pathways or environmental conditions of the host (207). The effect of the gut microbiota on the host blood-metabolome was investigated in germ-free mice by means of untargeted MS, which revealed a direct impact of the microbiota on the drug metabolism capacity of the host (363). A metabonomic study using proton-NMR profiling of urine and plasma to characterize and compare the metabolic response of conventional and germ-free rats to hydrazine exemplified a reduced toxicity in conventional rats, indicating that the microbiota specifically alters the host hydrazine toxicity response (314). Notably, microbiota modulations by pro-, pre-, and synbiotic supplementation could be coupled to alterations in major mammalian metabolic processes and interorgan cross-talk, which were speculated to have long-term health consequences to the host (223).

These studies illustrate the intimate relation between overall biochemistry of the host and its microbiota, which supports their mutualistic relationship and underlines the importance of host-microbe metabolic relations. Analogously, combined metabolome analyses of different biological fluid, such as fecal water, urine and plasma in animal models further evidenced microbe-host mutualism, and allowed the tracking of microbial metabolites from non-digestible food ingredients (155). Knowledge obtained from studies using animal models could further be tested in humans to aim for more comprehensive phenotyping, more accurate definition of health, including information of metabolic disorders, which may open avenues for manipulating human microbiomes to optimize their contribution to human health by disease prevention and/or treatment (328).

Research aims and thesis outline

The research described in this thesis aims to provide a better understanding of the composition and dynamics of the human small intestinal microbiota and to provide insight in the metabolic potential as well as immunomodulatory properties of its typical commensal inhabitants. To this end, ileostoma effluent samples were used to investigate the small intestinal microbial composition with classical cultivation methods combined with state of the art microbial profiling technologies and molecular typing methodologies.

Considering the importance of molecular techniques in contemporary ecology investigations that study microbial communities, **chapter 2** compares and contrasts

two technologies, barcoded pyrosequencing and phylogenetic microarray analysis, for the determination of the bacterial composition in fecal and small intestinal samples from human individuals. In addition, the use of different primer pairs in barcoded pyrosequencing is evaluated to address their impact on the outcome of PCR-based approaches in microbial profiling. The analyses presented in the chapter demonstrate that different primers and the two profiling technologies provide overall similar results for samples of fecal and the terminal ileum. In contrast, the profiles obtained by barcoded pyrosequencing and phylogenetic microarray analyses were substantially different for ileostoma effluent samples, which is most likely attributable to less complete coverage of the small intestinal microbiota by the probes present on the phylogenetic microarray that were designed on basis of known large intestinal 16S rRNA gene sequences. As pyrosequencing offers *de novo* community profiling, subsequent work employed this technologies as the method of choice for profiling of small intestinal communities. The 16S ribosomal RNA gene (rDNA) and ribosomal RNA (rRNA) content of ileostoma effluent samples from four different subjects was determined, while Illumina sequencing of cDNA derived from enriched mRNA was used to obtain metatranscriptome profiles of the same sample set (**Chapter 3**). This revealed that the composition of the small intestinal bacterial communities as assessed from rDNA, rRNA, and mRNA patterns appeared to be similar, indicating that the dominant bacteria in the small intestine are also highly active in this ecosystem. *Streptococcus* spp. were among the few bacterial species that were detected in each ileostoma effluent sample, albeit that their intra- and inter- individual abundances varied greatly. *Veillonella* spp. were frequently co-occurring with *Streptococcus*, indicating that the *Streptococcus* and *Veillonella* populations play a prominent role in the human small intestine ecosystem and their co-occurrence suggests a metabolic relation between these genera. Therefore, cultivation and molecular typing methodologies were employed to zoom-in on these groups (**Chapter 4**), which revealed that the richness of the small intestinal streptococci strongly exceeded the diversity that could be estimated on basis of 16S rRNA analyses, and could be extended to the genomic lineage level (anticipated to resemble strain-level). From ileostoma samples 3 different *Streptococcus* species were recovered belonging to the *S. mitis* group, *S. bovis* group, and *S. salivarius* group, which could be further divided in 7 genomic lineages. Notably, the *Streptococcus* lineages that were isolated displayed distinct carbohydrate utilization capacities, which may imply that their growth and relative community composition may respond quite strongly to differences in the dietary intake of simple carbohydrates over time. This notion is in good agreement with the observation that the *Streptococcus* lineage populations fluctuated in time with only one *Streptococcus* lineage being cultivated from both ileostoma samples collected in a one-year time frame. Conversely, the cultivated *Veillonella* isolates from samples during that same time-interval consistently encompassed a single lineage. Furthermore, this *Veillonella* lineage could be isolated from both the oral cavity as well as the ileostoma effluent. Analogously, three *Streptococcus* lineages that belong to a single phylotype also appeared to be present in both oral and small intestinal microbiotas. These

observations establish a high-level phylogenetic relatedness between the representatives of the *Veillonella* and *Streptococcus* genera that are encountered in the oral and small intestinal microbial ecosystems.

The metabolic capacity of 6 small intestinal *Streptococcus* lineages was further investigated through the determination of their genomic sequence (**Chapter 5**). The *Streptococcus* lineages were found to encode a different repertoire of carbohydrate transporters and the cognate enzymes to metabolize different sugars, which was in excellent agreement with the observed carbohydrate utilization capacity determined for these representative strains of the *Streptococcus* lineages (**Chapter 4**). Additionally, the *Streptococcus* genomes were mined for two component systems and genes encoding bacteriocins to gain further insights how these *Streptococcus* lineages perceive and respond to environmental stimuli. In **Chapter 6** dendritic cells were stimulated with strains of the different *Streptococcus* lineages to study their immunomodulatory properties. The *Streptococcus* lineages differed significantly in their capacity to modulate cytokine responses of blood-monocyte derived immature dendritic cells. Intriguingly, co-stimulation of dendritic cells with *Streptococcus* and *Veillonella* resulted in considerably different cytokine responses as those that could be predicted from the stimulations with either *Streptococcus* or *Veillonella*, indicating that it is not trivial to predict gut mucosal associated immune responses based on results from *in vitro* assays.

Chapter 7 completes this thesis with a summary and discussion of the results that are presented in the light of what they contribute to the current state of knowledge. In addition, it discusses and proposes several research directions that may be developed in the future to further our knowledge of the microbial ecosystem in the human small intestine and its role in health and disease.



Chapter 2

Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples

Bartholomeus van den Bogert, Willem M. de Vos, Erwin G. Zoetendal, and Michiel Kleerebezem

Appl Environ Microbiol, 2011. 77: 2071-2080

Abstract

Large-scale and in-depth characterization of the intestinal microbiota necessitates application of high-throughput 16S rRNA gene-based technologies, such as barcoded pyrosequencing and phylogenetic microarray analysis. In this study, both techniques were compared and contrasted for analysis of the bacterial composition in three fecal and three small intestinal samples from human individuals. As PCR remains a crucial step in sample preparation for both techniques, different forward primers were used for amplification to assess their impact on microbial profiling results.

An average of 7,944 pyrosequences, spanning the V1 and V2 region of 16S rRNA genes, was obtained per sample. Although primer choice in barcoded pyrosequencing did not effect species richness and diversity estimates, detection of *Actinobacteria* strongly depended on the selected primer. Microbial profiles obtained by pyrosequencing and phylogenetic microarray analysis (HITChip) correlated strongly for fecal and ileal lumen samples, but were less concordant for ileostoma effluent. Quantitative PCR was employed to investigate the deviations in profiling between pyrosequencing and HITChip analysis. Since cloning and sequencing of random 16S rRNA genes from ileostoma effluent confirmed the presence of novel intestinal phylotypes detected by pyrosequencing, especially those belonging to the *Veillonella* group, the divergence between pyrosequencing and the HITChip is likely due to the relatively low number of available 16S rRNA gene sequences from small intestinal origin in the DNA databases that were used for HITChip probe design. Overall, this study demonstrated that equivalent biological conclusions are obtained by high throughput profiling of microbial communities, independent of technology or primer choice.

Introduction

The human gastrointestinal (GI) tract is inhabited by a microbiota that predominantly consists of bacteria and is dominated by the phyla *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* (268). This community increases in numbers as well as diversity along the longitudinal axes of the GI tract and ultimately reaches populations as high as 10^{11} bacteria per gram of contents in the large intestine (32, 203, 341). The diversity and population dynamics of the lower GI tract microbiota has been well documented (203, 266, 382). In contrast, the microbiota of the upper GI tract has been poorly described, which is mainly due to sampling difficulties (31, 203). Recently the human small intestinal microbiota was characterized using samples obtained from ileostomy subjects (31, 125), and samples from the small intestine of healthy individuals obtained with an orally introduced catheter (381).

Much emphasis has been placed on understanding the dynamics and activities of the intestinal bacterial communities (31, 266, 329). The means by which this research has been conducted, underwent a revolution from culture based approaches to molecular technologies during the last decades (see (341) and (378) for reviews). Molecular technologies based on 16S ribosomal RNA (rRNA) and its encoding gene, such as fluorescent *in situ* hybridization (FISH) (124), quantitative PCR (qPCR) (272), denaturing gradient gel electrophoresis (DGGE) (19), terminal-restriction fragment length polymorphism (T-RFLP) (83), as well as the classical 16S rRNA gene amplicon-cloning and sequencing approach (80) are commonly used for compositional studies of the intestinal microbial ecosystem (346, 378). However, these approaches are laborious, especially when one aims for in-depth microbial community profiling (378). Phylogenetic microarrays (252, 266) and pyrosequencing technology (220) have become popular methods since they principally allow high-throughput and in-depth monitoring of microbial communities. While the former relies on 16S rRNA gene-targeted oligonucleotide probes for detection of bacteria in environmental samples (266), the latter allows *de novo* community profiling by sequencing and subsequent identification of partial 16S rRNA gene amplicons (9).

Each of the above mentioned approaches for characterization of microbial communities is limited in its correct assessment of microbial abundances due to the unspecified partial phylogenetic coverage by primers or probes that are used during the initial stages of sample preparation (16, 98). This notion is exemplified by the underestimation of *Bifidobacterium* spp. abundances in fecal samples by the commonly used universal 27F primer (127). One constraint of phylogenetic arrays for microbial profiling is their limitation to detect phylogenetic groups for which probes are represented on the array, although higher taxonomic level probes can still provide information for those groups (252), whereas analysis of data from pyrosequencing is challenging due to the vast number of obtained sequences that may contain sequence errors that disturb appropriate data interpretation (120).

The aim of this research was to assess the accordance of barcoded pyrosequencing and a phylogenetic microarray, the Human Intestinal Tract Chip (HITChip) (266), for profiling of human fecal and small intestinal microbial communities. In addition, the

use of different primer pairs in barcoded pyrosequencing was evaluated to answer whether their use in PCR based approaches influences the outcome of microbial diversity estimates and profiling (Figure 2.1).

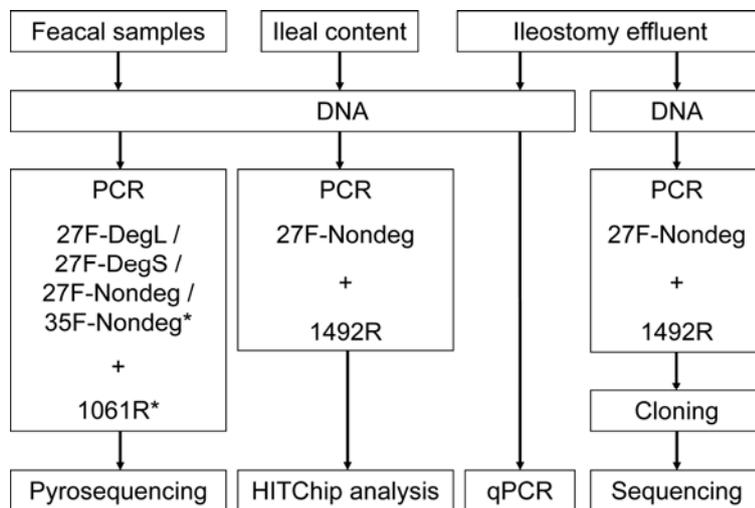


Figure 2.1. Schematic representation of the experimental set up for characterization of the microbial composition in fecal samples, ileal content and ileostoma effluent using molecular approaches. *: Forward primers and the reverse primer used for pyrosequencing were appended with adaptor A and adaptor B, respectively.

Materials and methods

Sample collection

Fecal samples (F1-F3) used in this study were collected at home from three healthy individuals (2 female, 1 male; aged 30-32 years), frozen in dry-ice immediately, and transported to the laboratory where they were kept at -80°C until further analysis.

Ileostoma effluent samples (S1 and S2) were previously collected (31) at least 3 h apart in the morning and afternoon, respectively, from a healthy 74 year old male ileostomist as part of a previous project, results of which are reported elsewhere (31). The volunteer collected the ileostoma effluent samples by emptying the ileostoma effluent in freezer baskets as soon as the bulk of ileostoma effluent was collected in a clean empty ileostoma bag. Samples were stored on dry-ice at approximately -80°C and were processed within three days after collection.

An ileum lumen sample (S3) was obtained from a 24 year old healthy female individual by using an orally introduced catheter, which passed to the ileum (120 cm distal to the pylorus) by peristalsis. Sampling was done under gastroenterologist-supervision, following flushing the ileum with 10 ml physiological salt solution through a port of the catheter, after which the sample was frozen and stored at -80°C .

Bacterial reference strains and culture conditions

Bifidobacterium longum (DSM 20219) was grown in ST medium as described in (293) with a substitution of proteose peptone for 1 g/l casitone (Becton Dickinson, Breda, the Netherlands) and meat extract for 3 g/l beef extract (Sigma, St. Louis, MO, USA). *E. coli* MC1061 was cultivated in Luria-Bertani (LB) broth at 37°C in a shaking incubator (135 RPM; Heidolph Instruments GmbH & Co, Schwabach, Germany). *Streptococcus thermophilus* (CNRZ 1066) was grown in M17 broth (Becton Dickinson) supplemented with 0.5% (w/v) glucose (Sigma) at 37°C . *Veillonella*

atypica (DSM 20739) was grown in *Veillonella* medium described in the DSMZ catalogue (Medium 136) under an N₂ atmosphere.

DNA extraction

Genomic DNA (gDNA) extractions from reference strains were performed using the fDNA[®] SPIN Kit for Soil (MP Biomedicals, Solon, OH, USA) with pelleted cells from 2 ml pure culture as starting material (data not shown).

Total DNA was extracted from 0.25 gram fecal sample and 0.25 ml ileal content, using the Repeated Bead Beating method described in (283), and from 0.2 gram ileostoma effluent as previously described (379) by using the QIAamp DNA Stool Mini Kit (Qiagen GmbH, Hilden, Germany). A recent study by Salonen, et al. (283) concluded that the difference in microbial composition between DNA extraction methods is relatively small in relation to that between subjects.

DNA was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop[®] Technologies, Wilmington, DE, USA) and adjusted to 10-20 ng/μl as template for subsequent 16S rRNA gene PCR amplification.

16S rRNA gene amplicon pyrosequencing

Amplicons from the V1-V6 region of 16S rRNA genes were generated by PCR using two degenerated (27F-DegL and 27F-DegS) and two non-degenerated (27F-Nondeg and 35F-Nondeg) primers in combination with a single reverse primer (1061R-Deg; Table 2.1) for each fecal and small intestinal DNA extraction.

To facilitate pyrosequencing using titanium chemistry, each forward primer was appended with the titanium sequencing adaptor A and a 'NNNN' barcode sequence (Table 2.1) at the 5' end, where NNNN is a sequence of four nucleotides that was unique for each sample and did not start with G nor contained a triplicate of identical bases. The reverse primer carried the titanium adaptor B at the 5' end.

PCRs were performed using a thermocycler GS0001 (Gene Technologies, Braintree, U.K.) in a total volume of 50 μl containing 1× PCR buffer, 1 μl PCR Grade Nucleotide Mix, 2.4 units of Faststart Taq DNA polymerase (Roche, Diagnostics GmbH, Mannheim, Germany), 200 nM of a forward and the reverse primer (Biologio BV, Nijmegen, The Netherlands), and 0.2-0.4 ng/μl of template DNA. The amplification program consisted of an initial denaturation step at 95°C for 5 min, 35 cycles of: denaturation at 95°C for 30 s, annealing at 56°C for 40 s and elongation at 72°C for 70 s, and a final extension step at 72°C for 10 minutes. The size of the PCR products was confirmed by gel electrophoresis using 1 μl of the reaction mixture on a 1% (w/v) agarose gel containing 0.4 μg/ml ethidium bromide (Bio-rad, Hercules, CA, USA). Control PCR reactions were performed alongside each separate amplification without addition of template, and consistently yielded no product. The optimal annealing temperature for primers (56°C) with attached adaptors and barcodes was determined by a 12-degree temperature gradient (49°C-61°C) PCR using DNA from fecal sample F2 (data not shown).

PCR products were purified with the ZR-96 DNA Clean and Concentrator kit (Zymo Research, Orange, USA) followed by DNA yield quantification using a NanoDrop ND-1000 spectrophotometer.

Purified PCR products were mixed in equimolar amounts with a final DNA concentration of 100 ng/μl. The pooled amplicons were pyrosequenced using a Genome Sequencer FLX in combination with titanium chemistry (GATC-Biotech, Konstanz, Germany). Sequencing occurred on a picotiterplate of which a quarter space was available for samples included in this study.

HITChip analysis

Microbial community profiling was also performed using the HITChip (266), which is a phylogenetic microarray, produced by Agilent Technologies (Palo Alto, CA) in 8×15K format, with over 4800 tiling oligonucleotides targeting the V1 or the V6 region of the 16S rRNA gene from 1,132 microbial phylotypes present in the human gastrointestinal tract (266). See the supplementary materials and methods section for the hybridization and analysis procedure.

Pyrosequence analysis and comparison with HITChip analysis

Pyrosequences were sorted per barcode. To the best of our knowledge, no recommendations for quality filtering of reads generated by pyrosequencing using titanium chemistry have been published to date, and therefore, we applied previously reported recommendations for quality filtering of pyrosequences generated by the GS 20 platform (149). This filtering was performed using an in-house perl script that passed sequences with exact matches to the forward primer, no ambiguous bases (N), and read-lengths not longer or shorter than 1 SD from the average sequence length (>87-157 and <314-359 nucleotides; See table S2.2 for the actual upper and lower read-length limit per sample). Additionally, primer sequences were removed from the pyrosequencing reads, and remaining sequences were analyzed. The number of operational taxonomic units (OTUs), rarefaction curves, and total species richness estimations (Abundance-based Coverage Estimators (ACE) and Chao1) (143) for the quality filtered sequences per sample were calculated using ESPRIT (310) with default settings (without removing low quality reads) at 0.02 distance level. Taxonomic classification of sequencing reads employed a locally installed version of the Ribosomal Database Project (RDP) Classifier (355), which by default produces classifications into the new higher-order taxonomy as proposed in Bergey's Taxonomic Outline of the Prokaryotes (102). The corresponding assignments differ from those that are produced by HITChip analysis, which has a standard output of the relative contributions of 1,132 microbial phylotypes at Level 1 (phylum-like with *Firmicutes* divided into classes or clusters), Level 2 (genus-like), and/or Level 3 (phylotypes based on >98% sequence identity) (266) to the overall microbial community per sample in the phylogeny as proposed by Collins, et al. (54). The 16S rRNA gene sequences of these phylotypes are present in a non-redundant ARB (215) database that was used for HITChip probe design: the human unique OTU database (266, 268). The sequences with corresponding assignments present in this

database were exported and used to train the RDP-classifier. This yielded a classifier that (in combination with a trial multiclassifier provided by the RDP staff) could classify pyrosequencing reads originating from different samples on a large scale with the same assignments as are produced by HITChip analysis. The confidence threshold used for classification of the pyrosequences was kept at 80%. Moreover, the multiclassifier summarized the assignments per taxon, which facilitated calculation of relative contributions and subsequent construction of microbial profiles for comparison with those that were generated by HITChip analysis.

Hierarchical cluster analysis of the microbial profiles was done in R version 2.9.2 by computing a distance matrix that was based on Pearson product-moment correlation coefficients (r) between pairs of profiles with level 2 community data. Visualization of hierarchical clustering was done by using the distance matrix in the `hclust` function in R with Ward's minimum variance method as agglomeration method. The Shannon diversity index was calculated in R using the `diversity` function with the level 2 community data from each sample. Screening sequences for exact matches with the HITChip probes or primers was performed using in-house perl scripts.

Quantification of bacterial community members by qPCR

All qPCR reactions were performed in 96 well PCR plates (Bio-rad) sealed with Microseal 'B' Film (Bio-rad) using a MyIQ Icyler with MyIQ software version 1.0.410 (Bio-rad). Each reaction was carried out in a total volume of 25 μ l using IQ SYBR Green Supermix (Bio-rad) according to manufacturer's instructions with 200 nM of forward and reverse primer in combination with 5 μ l template DNA.

From a literature survey, group-specific primers were chosen (Table 2.1) that were deemed optimal in their phylogenetic specificity and coverage (based on results of the 'probe match' tool offered in the Ribosomal Database Project <http://rdp.cme.msu.edu/> (53)) as well as the minimal tendency to form secondary structures, including hairpin loops, heterodimers, and homodimers (assessed using the IDTDNA Oligoanalyzer 3.1; <http://eu.idtdna.com/>) that may interfere with PCR efficiency (184). The optimal annealing temperature for each primer pair was determined by an 8-degree temperature (53°C-64°C) gradient PCR using gDNA from target bacterial reference strains as template (data not shown). The amplification program for most qPCR assays consisted of an initial denaturation step at 95°C for 5 min, 40 cycles of: denaturation at 95°C for 15 s, annealing at the optimal temperature for 30 s (with data acquisition) and elongation at 72°C for 30 s, and a final extension step at 72°C for 10 minutes. The elongation time for the *Streptococcus* qPCR assay was set at 20 s, whereas the denaturation and the elongation time for the *Bifidobacterium* qPCR assay were set at 30 s and 40 s, respectively. This was done, for practical reasons, to reduce the time to complete the *Streptococcus* qPCR assay and to provide sufficient time for denaturation and elongation of the relatively large amplicon (550 bp) produced during the *Bifidobacterium* qPCR assay. Melting curve analysis was carried out by incrementally increasing the temperature from 55°C to 95°C at 30 s per 0.5°C with continuous fluorescence collection.

Table 2.1. Adaptors and primers used in this study for 16S rRNA gene sequence PCR amplification for pyrosequencing, HITChip analysis, qPCR, and 16S rRNA gene cloning and sequencing

Target bacteria or bacterium	Primer ^a	Primer sequence (5'-3') ^b	Application ^c and PCR annealing temperature (°C) ^d	Reference
	Adaptor A	CCATCTCATCCCTGCGTGTCTC CGACTCAG	P	Provided by GATC-Biotech
	Adaptor B	BioTEG/CCTATCCCCTGTGTGC CTTGGCAGTCTCAG	P	Provided by GATC-Biotech
	27F-DegL	AGRGTTYGATYMTGGCTCAG	P(56)	(251)
	27F-DegS	GTTYGATYMTGGCTCAG	P(56)	This study ^e
	27F-Nondeg	GTTTGATCCTGGCTCAG	P(56) /H(52) ^f /C(52)	(266)
Total Bacteria	35F-Nondeg	CCTGGCTCAGGATGAACG	P(56)	(127)
	1061R-Deg	CRRCACGAGCTGACGAC	P(56)	(9)
	Uni-1492-rev	CGGCTACCTTGTTACGAC	H(52)/C(52)	(266)
	BACT1369F	CGGTGAATACGTTTCYCGG	Q(56)	(313)
	PROK1492R	GGWTACCTTGTTACGACTT		
<i>Bifidobacterium</i>	g-Bifid-F	CTCCTGAAAACGGGTGG	Q(55)	(228)
	g-Bifid-R	GGTGTCTTCCCGATATCTACA		
<i>Veillonella</i>	Veil-F-Rinttilä	AYCAACCTGCCCTTCAGA	Q(57)	(272)
	Veil-R-Rinttilä	CGTCCCGATTAACAGAGCTT		
<i>Streptococcus</i>	Strep-F-Rudney	AGATGGACCTGCGTTGT	Q(55)	(279)
	Stherm 08	GTGAACTTTCCACTCTCACAC		(100)
<i>Escherichia coli</i>	E.coli-F-Huijsdens	CATGCCGCGTGTATGAAGAA	Q(57)	(144)
	E.coli-R-Huijsdens	CGGGTAACGTCAATGAGCAAA		

^a Primer names may not correspond to original publication

^b M = A or C; R = A or G; W = A or T; Y = C or T

^c Pyrosequencing (P); HITChip analysis (H); qPCR (Q); 16S rRNA gene clone library construction and sequencing (C)

^d Annealing temperatures indicated in bold were determined as explained in Materials and methods

^e 5' end 3 nt trimmed version of the 27F-DegL primer

^f With a T7 promotor sequences appended on the 5' end

For each qPCR assay, a standard curve comprising 8 serial 10-fold dilutions of full-length 16S rRNA gene PCR products was generated from target gDNA preparations of the respective reference strains. For the total bacteria qPCR assay a standard curve was generated using *E. coli* MC1061 gDNA. The standard curves of each qPCR assay were used to determine the relative contribution of target bacterial groups to the total bacterial community in sample DNA preparations.

16S rRNA gene library construction and analysis

The 27F-Nondeg and Uni-1492-rev primers (Table 2.1) were used for PCR amplification of 16S rRNA gene sequences from the undiluted extracted DNA of sample S1 and S2. Each reaction was performed in quadruplicate in a total volume of 50 µl containing 1x PCR buffer (Promega), 200 nM of each primer (Biologio), 200 µM

of each deoxyribonucleotide triphosphate (Promega), 1.25 U GoTaq[®] DNA polymerase (Promega) and 1 µl of the extracted DNA. The amplification program was performed on a T1 thermocycler (Biometra, Göttingen, Germany) and consisted of an initial denaturation step at 94°C for 2 min, 35 cycles of: denaturation at 94°C for 30 s, annealing at 52°C for 40 s and elongation at 72°C for 90 s, and a final extension step at 72°C for 5 minutes. PCR products were verified by gel electrophoresis using 5 µl of the reaction mixture on a 1% (w/v) agarose gel containing 0.4 µg/ml ethidium bromide (Bio-rad). Quadruplicate PCR products from the same sample were pooled and subsequently purified with a High Pure Cleanup Micro Kit (Roche) using 10 µl Elution Buffer (Roche) for elution. The purified PCR products were diluted 10 times of which 1 µl was used for ligation into the pGEM-T easy vector (Promega) overnight at 4°C according to the manufacturer's instructions. XL1-blue competent cells (75 µl; Stratagene, La Jolla, CA, USA) were transformed with 2 µl ligation mixture according to manufacturer's instructions and subsequently plated on LB agar containing ampicillin (100 µg/ml; Sigma), isopropyl-β-D-thiogalactopyranoside (IPTG; 0.16 mM; Carl Roth GmbH, Karlsruhe, Germany), and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal, 100 µg/ml; Invitrogen, Carlsbad, CA, USA) for blue-white color screening. White colonies were randomly selected and separately cultured overnight at 37°C in LB medium containing ampicillin. Subsequently, 2×96 clones from each clone library were randomly selected and the cloned inserts were sequenced from both ends using the T7 and SP6 priming sites (GATC-Biotech, Konstanz, Germany). The obtained sequences per clone were assembled using Clone Manager 9 Professional Edition (Scientific & Educational Software, Cary, NC, USA) yielding near full-length 16S rRNA gene sequences, which were analyzed with DNA Baser v2.71.0 (HeracleSoftware, Lilienthal, Germany) to trim vector sequences. Subsequently, sequences were tested for chimeras using Mallard (13) following the instructions of the authors with the default settings to identify chimeric sequences. Putatively anomalous 16S rRNA gene sequences identified by Mallard were further analyzed following the anomaly confirmation protocol suggested by the authors and unambiguously anomalous sequences were excluded from further microbiome interpretations. Non-chimeric sequences were taxonomically classified using the in-house customized RDP classifier described above. Sequences for which no classification could be obtained above the 80% confidence threshold were classified using the locally installed version of the RDP classifier version 2.2 (355) with a default confidence threshold of 80%.

The non-chimeric 16S rRNA gene sequences from each clone library were aligned using the SILVA Webaligner (261) and subsequently imported into ARB. Each clone library was manually screened (using a neighbor joining distance matrix, employing no correction, generated in ARB) for sequences showing <98% identity to 16S rRNA gene sequences represented in the human unique OTU database that was used for HITChip probe design.

Nucleotide sequence accession numbers

The cloned 16S rRNA gene sequences from ileostoma effluent were deposited in the Genbank database and are available under accession numbers HQ176022-HQ176318.

Results

Analysis of pyrosequencing reads from 16S rRNA gene amplicons

Pyrosequencing of the 16S rRNA gene PCR amplicons from fecal samples (F1-F3), ileostoma effluent (S1 and S2), and ileal lumen content (S3) yielded in total 190,652 sequences with $7,944 \pm 2201$ sequences per sample. Quality filtering passed approximately 50% of the pyrosequencing reads, with an average length of 224 nt (Table 2.2; See Table S2.1 for a comprehensive overview of the characteristics of pyrosequencing reads before and after quality filtering). Detailed analysis revealed that the majority ($74.80\% \pm 4.74\%$) of the sequences that failed to pass quality filtering were due to sizes that were outside the sequence length thresholds (Table S2.2).

Microbial profiles, based on the level 1 (phylum-like; Figure 2.2) and level 2 (genus-like; Figure S2.1) taxonomic assignments of the pyrosequences that passed quality filtering, were constructed for all samples. As anticipated, most pyrosequences were assigned to the *Firmicutes* (85.6%), *Actinobacteria* (7.6%), and *Bacteroidetes* (2.9%; predominantly encountered in fecal samples F1 and F2 as well as in the ileal lumen content S3).

Notably, only 1.5% of all pyrosequences could not be classified using the confidence-threshold of 80% and are represented as unclassified_Human unique OTU (Figure 2.2). Furthermore, microbial profiles obtained using all pyrosequences (without quality filtering), revealed essentially the same profiles as those obtained with the quality-filtered sequences, albeit with a raised abundance of the unclassified_Human unique OTU (Figure S2.2), indicating that the quality filtering step does not drastically influence the reconstruction of the microbial community but predominantly eliminates noise.

Hierarchical clustering of the microbial profiles revealed separate grouping of fecal samples and ileum-lumen content from ileostoma effluent samples (Figure S2.3A). The divergence between these clusters was most apparent for phylogenetic groups belonging to the *Firmicutes* with fecal samples and ileum-lumen content being abundant in *Clostridium* clusters IV ($12.3\% \pm 7.3\%$), XIVa ($58.6\% \pm 8.6\%$), XVI ($4.4\% \pm 3.8\%$), and XVIII ($3.0\% \pm 1.8\%$), while Bacilli ($9.2\% \pm 4.5\%$) and *Clostridium* clusters I ($24.0\% \pm 11.5\%$), IX ($13.3\% \pm 7.1\%$), XI ($14.0\% \pm 14.1\%$), and XIVa ($27.2\% \pm 13.4\%$) were predominant in ileostoma effluent. Moreover, species richness, as reflected by Chao1 and ACE supported by the rarefaction curves, as well as the Shannon diversity index, was higher in fecal samples and ileum-lumen as compared to ileostoma effluent (Table 2.2; Figure S2.4).

Table 2.2. Characteristics of sequence analysis before and after quality filtering

Sample and primer	Characteristic before quality filtering			Characteristics after quality filtering							Shannon diversity index		
	No. of sequences	Sequence length (nt)		No. of sequences	Avg sequence length (nt)	% remaining quality filtered sequences	OTU ^a	Value	Chao1 ^a			ACE ^a	
		Avg	SD						Upper limit	Lower limit			
F1	27F-DegL	5051	248.77	105.78	2562	236.07	50.72	712	1029	1141	946	1049	3.139
	27F-DegS	7554	253.98	105.43	3767	244.11	49.87	993	1420	1553	1318	1341	3.168
	27F-Nondeg	5465	232.13	114.40	2652	227.05	48.53	775	1078	1181	1002	1118	2.981
	35F-Nondeg	7909	220.36	118.15	3747	219.15	47.38	1060	1572	1717	1460	1554	3.019
F2	27F-DegL	10425	247.38	98.65	5431	233.67	52.10	1199	1591	1704	1503	1587	3.269
	27F-DegS	11591	258.18	100.93	5803	245.61	50.06	1217	1618	1735	1527	1597	3.275
	27F-Nondeg	9524	230.07	110.93	4704	225.01	49.39	1139	1572	1697	1476	1565	3.231
	35F-Nondeg	7932	228.53	110.98	4071	223.96	51.32	1070	1569	1715	1457	1524	3.190
F3	27F-DegL	4782	230.79	100.25	2595	218.93	54.27	703	1103	1245	997	1058	2.845
	27F-DegS	5349	227.53	105.71	2784	220.06	52.05	709	1036	1156	949	1019	2.828
	27F-Nondeg	7430	204.91	110.82	3606	207.95	48.53	981	1424	1553	1324	1487	2.832
	35F-Nondeg	4584	201.40	113.57	2191	202.30	47.80	649	985	1111	894	950	2.774
S1	27F-DegL	10393	237.59	96.99	5016	217.23	48.26	687	910	1008	842	864	2.105
	27F-DegS	10320	243.42	96.71	5136	227.11	49.77	738	978	1076	908	960	2.084
	27F-Nondeg	9593	217.34	100.03	4660	203.31	48.58	664	838	914	785	844	1.914
	35F-Nondeg	11378	234.07	108.65	5822	227.00	51.17	842	1067	1154	1004	1056	2.190
S2	27F-DegL	7925	246.04	97.97	3993	227.30	50.38	722	964	1061	894	928	2.329
	27F-DegS	10053	241.03	101.26	4930	226.91	49.04	778	973	1051	917	980	2.385
	27F-Nondeg	7628	228.75	106.49	3752	216.33	49.19	669	852	929	797	862	2.160
	35F-Nondeg	10294	227.87	113.33	4509	216.00	43.80	758	965	1047	907	983	2.285
S3	27F-DegL	5924	247.74	100.82	3127	234.86	52.79	792	1079	1179	1005	1092	2.899
	27F-DegS	6501	240.44	105.29	3456	235.31	53.16	901	1301	1428	1204	1294	3.021
	27F-Nondeg	6759	218.29	114.20	3217	216.92	47.60	917	1374	1514	1267	1355	3.159
	35F-Nondeg	6288	218.97	114.26	3072	213.99	48.85	859	1196	1305	1114	1215	3.163

^a. Calculated at an 0.02 distance level.^b. 95% CI, 95% confidence interval.

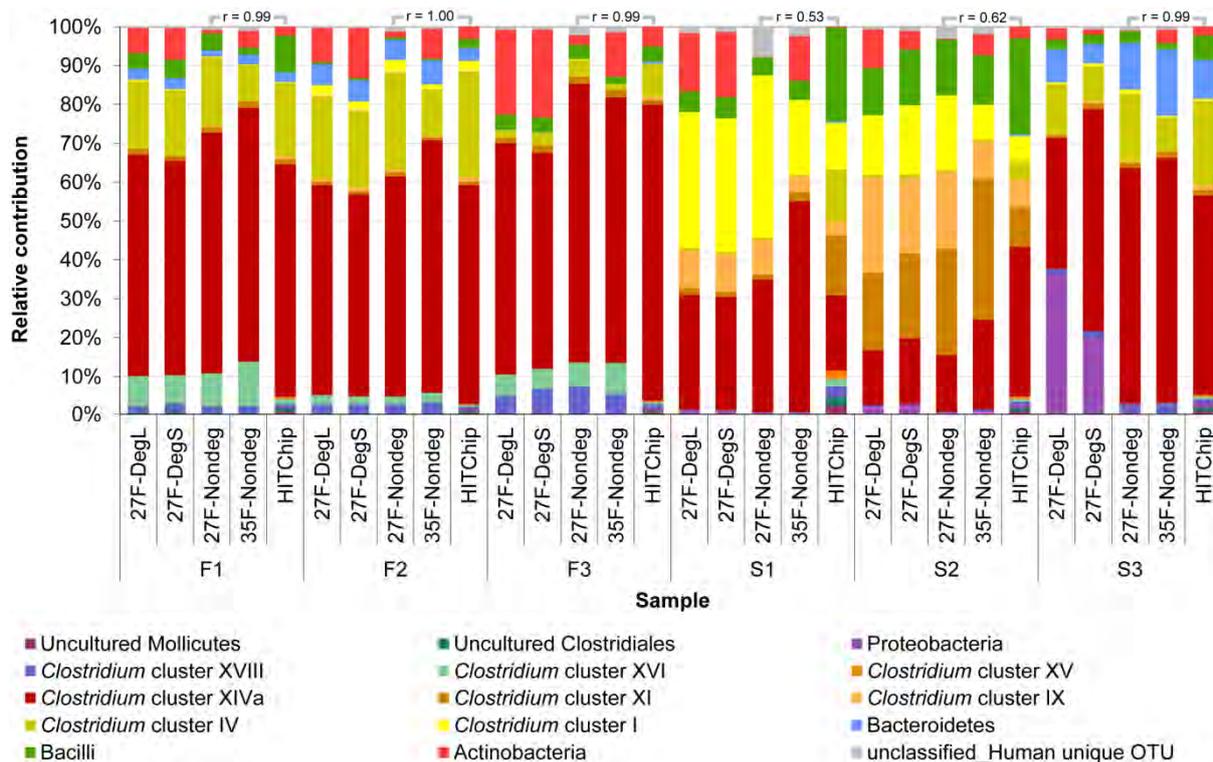


Figure 2.2. Relative contribution of detected bacterial phyla with pyrosequencing using four different forward primers and HITChip analysis for community data at level 1. Pearson product-moment correlation coefficients (r) between pairs of profiles are shown above the bars. The phyla Firmicutes was subdivided in Bacilli, *Clostridium* clusters, Uncultured Mollicutes, and Uncultured Clostridiales. Pyrosequences that could not be classified above the confidence threshold of 80% are grouped to Unclassified_Human unique OTU which is indicated in the microbial profiles with a shadow. Phylogenetic groups that contribute at least 1% to one of the profiles are indicated in the legend.

The effect of different forward primers on microbial profiling by barcoded pyrosequencing

To determine the effect of different primers on microbial profiling, 16S rRNA gene PCR amplicons were generated for each intestinal sample using four different forward primers (27F-DegL, 27F-DegS, 27F-Nondeg and 35F-Nondeg). Microbial profiles constructed on basis of pyrosequences per sample were highly correlated for the different primers (average r of 0.88 ± 0.14 at level 2 community data; data not shown). Furthermore, hierarchical clustering (Figure S2.3A) of the microbial profiles revealed distinct clusters of microbial profiles for each of the samples using the four forward primers, except for sample S3 as will be discussed below, indicating that the effect of primers on microbial profiling is smaller than the sample specific effect and supporting a high level of technical reproducibility of the pyrosequencing method. Analogously, comparison of rarefaction curves (Figure S2.4), species richness estimators Chao1 and ACE, as well as Shannon diversity indices (Table 2.2) did not reveal a particular primer giving consistently the highest or the lowest value for any of these ecological metrics. Nonetheless, qualitative comparison demonstrated that the microbial profiles deduced from pyrosequencing using amplicons generated with the 27F-DegL, 27F-DegS and 35F-Nondeg primers were notably more abundant in

Actinobacteria relative to those using the 27F-Nondeg primer (Figure 2.2), confirming previous reports on the underestimation of the *Actinobacteria* using the 27F-Nondeg primer (127).

Because the PCR annealing temperatures used for amplicon generation differed between pyrosequencing and HITChip analysis (56°C vs. 52°C, respectively) both annealing temperatures were employed for HITChip analysis of sample F3 to investigate the effect of different annealing temperatures on microbial profiling. Results demonstrated highly similar microbial profiles ($r \geq 0.99$; data not shown) and, therefore, comparison of microbiota profiling by pyrosequencing and HITChip analysis was not biased by different PCR annealing temperatures.

Comparison of barcoded pyrosequencing with HITChip analysis

The concordance between microbial profiling by pyrosequencing and phylogenetic microarray analysis was evaluated. Although the principle for classification and abundance estimations of microbial community members differ between these technologies, hierarchical clustering of the microbial profiles from both methods matched (Figure S2.3)

Microbial profiles as a result of HITChip analysis were compared to those from pyrosequencing using the 27F-Nondeg primer as this forward primer is used for amplicon generation in the HITChip analytic procedure. The resulting comparison of the community data at level 1 (phylum-like) showed a high correlation for the fecal samples (F1-F3; $r = 0.99-1.00$) and ileum-lumen content (S3; $r = 0.99$), while the correlation was lower for ileostoma effluent samples (S1-S2; $r = 0.53-0.62$; Figure 2.2). Correlations for the community data at level 2 (genus-like), were significantly lower, but remained highest for the fecal samples ($r = 0.63-0.78$) and ileum-lumen content ($r = 0.71$) and lowest for the ileostoma effluent samples ($r = 0.31-0.49$; Figure S2.1). This difference between ileostoma effluent and other intestinal samples is also demonstrated in the community data scatter plots at level 1 and 2 (Figure S2.5). Remarkably, numerous phylogenetic groups at level 2 showed significant abundances with HITChip analysis but were absent from the pyrosequence dataset (Figure S2.5 inset), suggesting that HITChip analysis enables detection of low abundance bacterial groups by its broader dynamic range compared to pyrosequencing.

To verify if suboptimal HITChip probe matches could potentially explain higher abundances per cluster with grouping of fecal samples and ileal lumen content (Cluster I) apart from ileostoma effluent samples (Cluster II) of some phylogenetic targets in the pyrosequence data relative to HITChip analysis (Figure S2.6), pyrosequences were screened for exact matches with the HITChip probes designed for detection of these phylogenetic groups (Figure S2.7). In general, both clusters approximately showed the same fraction (~90%) of pyrosequences that had a perfect match with at least one of the HITChip probes. For cluster I, *Eubacterium rectale* et rel. and *Ruminococcus obeum* et rel. contained most pyrosequences that lack a HITChip probe perfect-match, while *Veillonella* showed the highest number of sequences (19.3%) without a perfect-match for the HITChip probes for cluster II.

Phylogenetic groups for which the abundance estimates were higher in the HITChip analyses included *Streptococcus* spp. (Bacilli) for which a deviation in relative abundance estimations between the two profiling technologies was as high as ~7% (Figure S2.6). The abundance of this phylogenetic group and others was further investigated by means of qPCR as well as 16S rRNA gene cloning and sequencing (see below).

Quantification of bacterial groups by qPCR

To evaluate the performance of pyrosequencing and HITChip analysis in estimating relative abundances of microbial community members, the results of both techniques were compared with those obtained by means of group-specific qPCR, focusing on *Bifidobacterium*, *Veillonella*, *Streptococcus*, and *E. coli* (Figure 2.3). The estimated community proportion of *Streptococcus* spp. was consistently highest with HITChip analysis relative to qPCR assays and pyrosequencing, while *Bifidobacterium* abundance levels were expectedly low when determined with the HITChip as a result of using the 27F-Nondeg primer for the initial sample preparation. Surprisingly, *Bifidobacterium* abundances assessed with the qPCR assay were relatively low as well, whereas estimations by pyrosequencing using the 27F-DegL, 27F-DegS, and 35F-Nondeg primers showed relative abundances varying from 1.10% to 6.98% for samples F1, F2, F3 and S3 and from 4.04% to 16.37% for samples S1 and S2. For *Veillonella* and *E. coli* relative abundances were highest by pyrosequencing analysis, followed by intermediate values obtained from qPCR and lowest values assessed by HITChip analysis. Sample S3 was the only sample for which an *E. coli* abundance above 1.5% was detected and showed relative contributions as high as 31.8% and 17.4% in microbial profiles from pyrosequencing using the 27F-DegL and 27F-DegS primers, respectively.

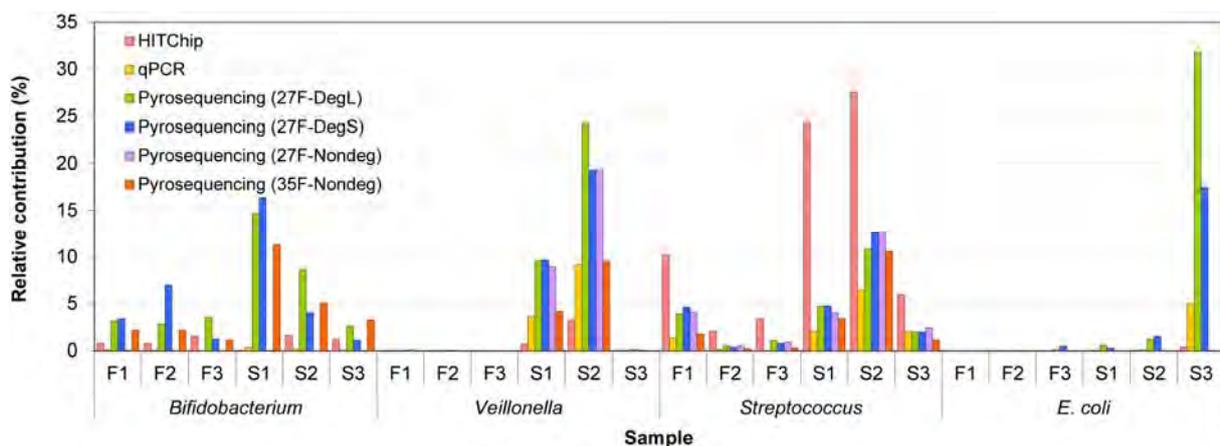


Figure 2.3. Comparison of the relative contribution as determined by means of HITChip, qPCR assays, and pyrosequencing for 4 phylogenetic groups in fecal (F) and small (S) intestinal samples. Relative contribution as assessed by qPCR assays were not determined for sample F3

Although the proportion of *E. coli* as assessed by the HITChip was considerable lower with an abundance of 0.44%, the *E. coli* specific qPCR assay revealed a contribution of 5% and confirms the abundant presence of *E. coli* in sample S3. Additionally, this observation suggests that microbial profiling using primers 27F-DegL and 27F-DegS results in a more accurate representation of the microbial composition in intestinal samples that contain higher proportions of *E. coli*.

Screening ileostoma effluent for novel bacterial phylotypes

Based on the prominent deviations between pyrosequencing and HITChip analysis for ileostoma effluent, these small intestinal samples are a potential source for a range of bacteria with 16S rRNA gene sequences that are absent from the human unique OTU database to date. To support this notion, 16S rRNA gene clone libraries for the two ileostoma effluent samples were constructed and analyzed (Figure 2.4). A total number of 139 and 158 cloned non-anomalous 16S rRNA gene sequences were obtained for the S1 and S2 16S rRNA gene clone libraries, respectively. Out of the total of 297 cloned sequences, 6 could not be classified at Level 2 assignments above the 80% confidence threshold using the RDP classifier trained with the human unique OTU database. Consequently, these sequences were grouped to “unclassified_” with the specific Level 1 assignment (Figure 2.4).

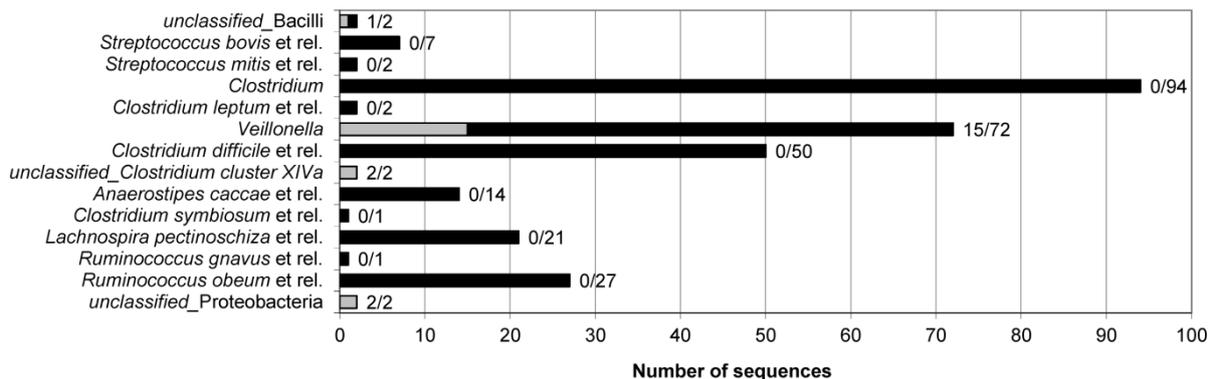


Figure 2.4. Total number of cloned sequences detected per phylogenetic group (black) and number of sequences showing <98% identity to the 16S rRNA gene sequences represented in the human unique OTU database (grey) for ileostomy sample S1 and S2. The ratio of number of sequences showing <98% identity to the 16S rRNA gene sequences represented in the human unique OTU database vs. total number of cloned sequences is provided for each level 2 group. Sequences that could not be classified above the confidence threshold of 80% are grouped to “unclassified_” at the specific rank per taxon.

To determine if the resolution of the identifications could be improved, sequences were re-classified using the standard RDP classifier (see Materials and methods), which is based on a more exhaustive set of 16S rRNA gene sequences compared to that of the human unique OTU database. This showed that both unclassified_Bacilli sequences were assigned to the order Lactobacillales, though deviating in their genus level classifications as *Granulicatella* for one and *Streptococcus* for the other. The sequence classified as the latter also showed <98% identity to 16S rRNA gene sequences represented in the human unique OTU database. Sequences of the

unclassified_*Clostridium* cluster XIVa group could be classified no further than the family Lachnospiraceae, whereas the unclassified_Proteobacteria were assigned to the genus *Variovorax* belonging to Betaproteobacteria.

Further analysis of the libraries revealed that 20 sequences, predominantly belonging to *Veillonella*, showed <98% identity to 16S rRNA gene sequences represented in the human unique OTU database and were therefore considered to be phylotypes not previously reported as associated with the human intestine (Figure 2.4). The finding of a relatively large proportion of these phylotypes among the *Veillonella* 16S rRNA gene sequences supports the suggestion that the HITChip probes display relatively poor sequence matches with these sequences (see above). Detailed analysis indeed showed that 5 out of 7 HITChip probes specific for *Veillonella* had more than 2 mismatches with sequences classified as *Veillonella*, whereas the remaining two probes at most had 1 mismatch (Figure S2.8).

The same screening strategy was applied to determine if the cloned sequences classified as *Veillonella* had exact matches with the forward and the reverse primer used for the *Veillonella* qPCR assay. Out of the 72 cloned sequences classified as *Veillonella*, only one sequence did not show a perfect match with the reverse primer. In contrast, 16 sequences had a single mismatch with the forward primer (Figure S2.8). Interestingly, 13 of these sequences were also identified as novel intestinal phylotypes (see above). This shows that the forward primer for the *Veillonella* qPCR assay is not in agreement with the 16S rRNA gene sequences of *Veillonella* and could also explain the lower *Veillonella* abundances as determined by qPCR in comparison with those estimated by pyrosequencing (Figure 2.3). Taken together, these results suggest that the human small intestine is inhabited by novel *Veillonella* phylotypes that previously have not been reported to inhabit this niche.

Discussion

In this study, the performance of two culture-independent techniques, barcoded pyrosequencing and phylogenetic microarray analysis using the HITChip, was compared and contrasted for profiling of human fecal and small intestinal microbial communities. Both techniques generated similar microbial composition profiles for fecal and terminal ileum-lumen samples, whereas more distinct profiles were obtained for ileostoma effluent samples. Ileostoma effluent, in comparison to fecal samples, contained less rich and diverse microbial communities, which were abundant in *Streptococcus* spp., *Veillonella* spp., and members of several *Clostridium* clusters. These findings are consistent with results published by Booiijink, et al. (31) as well as with the recent study by Zoetendal, et al. that concluded that the phylogenetic composition in ileostoma effluent is different from that of the ileum and resembles the microbiota in the proximal small intestine, i.e., the jejunum and proximal ileum (381).

The high comparability of the pyrosequencing- and HITChip-derived microbial profiles obtained for fecal samples is in agreement with previously published results (49). Thereby, this study confirms that both profiling technologies facilitate robust

microbial profiling and essentially generate equivalent biological conclusions regarding compositions of microbial communities. Nonetheless, abundance estimates for several phylogenetic groups deviated between pyrosequencing and HITChip analysis. Determining the exact cause for the technical divergence is not trivial, but possible reasons for this are: i) probe based vs. sequence based quantification; ii) failure to detect species that were not represented in the reference sequences used for probe design and cognate overestimation of relative abundances of the detected phylogenetic groups; iii) sequencing errors; iv) incorrect taxonomic classification of sequences; and/or v) difference in dynamic range between technologies. The latter was also apparent from comparison of HITChip analyses with very deep pyrosequencing resulting in a level of depth that is comparable to close to 200,000 reads per sample (49). Sequencing depth could be improved by employing the Illumina sequencing platform with which microbial community diversity is analyzed with increased depth relative to pyrosequencing (40, 50). However, to date, this approach is still challenging due to the limited phylogenetic resolution obtained from such short sequence reads and the increasing sequence-error rates for reads extended beyond 60 bp.

A challenging yet essential part of pyrosequencing analysis is quality control of the acquired dataset. Here, a strict quality filtering procedure was employed that eliminated approximately half of the pyrosequences, most of which were either deemed too short or too long. The proportion of sequences excluded from further analysis due to quality filtering was higher compared to other studies applying similar exclusion criteria. Those studies, however, employed the older GS 20 (9, 149) or 454 GS FLX (49) sequencing platforms, while in this study the Titanium method was applied.

HITChip probe design is based on a 16S rRNA gene database that predominantly contains sequences with a fecal or colonic origin. Since both profiling techniques correlated strongly for fecal samples and the HITChip offers a broader dynamic range of detection, HITChip analysis is preferred for profiling of the lower GI tract microbiota. However, HITChip coverage of the small intestinal microbiota appeared to be more incomplete (31) and, therefore, pyrosequencing would be the method of choice for *de novo* profiling of these microbial communities. Furthermore, this finding exemplifies the intrinsic constraint of microarray approaches that are limited to detection of phylogenetic groups for which sequences were included during array design. This may to a large extent explain the lower correlations between pyrosequencing and HITChip analysis that were obtained for the ileostoma effluent samples and suggests that ileostoma effluent harbors novel intestinal bacteria that have not been detected in feces or other large intestinal samples. To identify these phylotypes, pyrosequences were screened per phylogenetic group for perfect matches with the respective HITChip probes. All groups had multiple sequences without a probe match, which may indicate the presence of novel intestinal phylotypes, suggesting that HITChip probe-design may be improved by the addition of probes to detect this expanding community. The latter would require a flexible array design strategy as suggested by Rajilić-Stojanović, et al. (266). However,

despite quality filtering to improve overall pyrosequence dataset reliability, this technology still suffers from a relatively high sequence-error rate (~0.5% for the GS 20 and GS FLX platform (245)), which erroneously may contribute to the number of sequences that mismatch with the HITChip probes. Analogously, the pyrosequencing technology was reported to overestimate microbial diversity as a consequence of these sequencing errors (269).

Results from screening cloned 16S rRNA gene sequences from the ileostoma effluent samples showed that 7% of the cloned sequences represented novel intestinal phylotypes, which appears to be lower than might have been anticipated on basis of previous studies (31, 125). The novel phylotypes encountered here predominantly belonged to the *Veillonella* group, which is in agreement with the relatively large fraction of pyrosequences corresponding to this group that lacked a perfect-match HITChip probe. Other novel phylotypes from the clone libraries were identified as *Variovorax*, which have been cultured from soil (305) and have been detected in the rabbit cecum (235). However, to the best of our knowledge, *Variovorax* spp. have not been identified as inhabitants of the human GI tract to date. Therefore, the human small intestine contains a range of species that were not previously associated with this niche, and elucidating the role of these microorganisms, especially of the abundant *Veillonella*, in their environment is a task for the future.

Quantification of bacterial groups by means of qPCR was chosen as a benchmark technology to investigate the deviations in profiling between pyrosequencing and HITChip analysis. Discrepancies were observed between relative abundances as determined by qPCR and the two profiling techniques. This can at least in part be attributable to the difference in abundance calculations, which is based on a separate total bacteria assessment for qPCR, whereas for the HITChip and pyrosequencing relative abundances are calculated based on, respectively, probe signals or number of pyrosequences per phylogenetic group as part of the total. *Streptococcus* spp. abundance levels were consistently highest with HITChip analysis relative to qPCR assays and pyrosequencing. This observation corroborates the results previously published by Rajilić-Stojanović and colleagues (266), who reported a significantly higher relative abundance estimate of *Streptococcus* spp. in fecal samples based on HITChip analysis as compared to group-specific FISH analysis.

PCR amplification was performed using four different forward primers for each intestinal sample to assess the impact of PCR primer choice on microbial profiling by means of pyrosequencing. With the exception of sample S3, microbiota compositions per sample were highly similar for the different primers. This underpins the degree of reproducibility of microbial profiling by means of pyrosequencing and suggests that correlations between technical replicates can be expected to be even higher. Moreover, primer choice did not profoundly affect species richness and diversity estimates, which is in agreement with a recent study that showed consistent species evenness estimates when using different primer pairs targeting the same region of the 16S rRNA gene (88). Qualitative analysis of the microbial profiles, however, clearly revealed a lower abundance of Actinobacteria by the 27F-Nondeg primer,

which confirms the observations by Hayashi, et al. (127) showing that the 27F-Nondeg primer is incomplete in its coverage of *Bifidobacterium* spp.

In conclusion, this paper demonstrates that different primers and high-throughput 16S rRNA profiling technologies like barcoded pyrosequencing and HITChip analysis provide overall similar results. However, this similarity is dependent on the origin of the samples, which relates to the sequences used during array design and may thus be influenced by updated microarray design to accommodate novel sequences. Nonetheless, based on the results described here, it is our recommendation to use either the 27F-DegL or the 27F-DegS primer, since both these multiple-degenerate primers appear to provide a more complete assessment of Actinobacteria and *E. coli* abundances.

Acknowledgements

We appreciated the help of Carien Booijink and Freddy Troost for providing ileostoma effluent and ileum-lumen samples and isolated DNA thereof. We thank Sebastian Tims and Hans Heilig for providing fecal sample DNA and assistance with the HITChip analyses, Muriel Derrien and Odette Pérez Gutiérrez for their help in phylogenetic analysis of cloned sequences as well as Hauke Smidt and Mirjana Rajilić-Stojanović for critical reading of the manuscript. The authors would like to thank Christopher Bauser, Andrea Bolte, and Manuela Hinz of GATC-Biotech (Konstanz, Germany) for assistance in the set-up of the pyrosequencing experiments, and Benli Chai from RDP (East Lansing, MI, USA) for his assistance for customizing the RDP-classifier. Finally, this project was supported by the Netherlands Bioinformatics Centre (NBIC).

Supplementary information

Supplementary materials and methods - HITChip analysis

Microbial community profiling was also performed using the HITChip (266), which is a phylogenetic microarray, produced by Agilent Technologies (Palo Alto, CA) in 8×15K format, with over 4800 tiling oligonucleotides targeting the V1 or the V6 region of the 16S rRNA gene from 1,132 microbial phylotypes present in the human gastrointestinal tract (266). The HITChip has been validated by using FISH and qPCR on abundant species while it has been validated with artificial mixtures of 16S rRNA gene amplicons showing a linear dose-response in a large dynamic range of over 1000-fold (266). Moreover, the HITChip analyses are based on hybridization to oligonucleotides that are spotted at least twice on the arrays and all analyses are repeated in duplicate with a dye swap providing almost perfect reproducibility ($r \geq 0.98$).

The procedure for HITChip analysis was performed as described previously (266). In short, near full-length 16S rRNA gene amplicons were generated by PCR using the T7promotor-27F-Nondeg primer in combination with the Uni-1492-Rev primer (Table 2.1). PCR conditions were generally the same as for amplicon generation for pyrosequencing, except that PCRs were performed in duplicate for each sample, annealing was set at 52°C, and elongation time was set at 90 s. PCR products were pooled and purified with a High Pure Cleanup Micro Kit (Roche) using 15 µl Elution Buffer (Roche) for elution. Subsequently, PCR products were *in vitro* transcribed using the Riboprobe System (Promega, Madison, WI, USA) including a mix containing aminoallyl-rUTP (Ambion, Austin, TX, USA). DNA was digested by treatment with RNase free DNase (Promega), followed by RNA purification using RNeasy MinElute Cleanup Kit (Qiagen) and labeling with Cy3 and Cy5 (Amersham Biosciences, Little Chalfont, UK). Labeled RNA was purified as described above and fragmented with 10× fragmentation reagent (Ambion) Hybridization of labeled RNA on the arrays, washing of the slides followed by microarray data extraction and analysis was performed as described previously (266). Microbial profiles of the microbiota of each sample generated in this fashion were used for comparison with those from pyrosequencing.

Supplementary figures

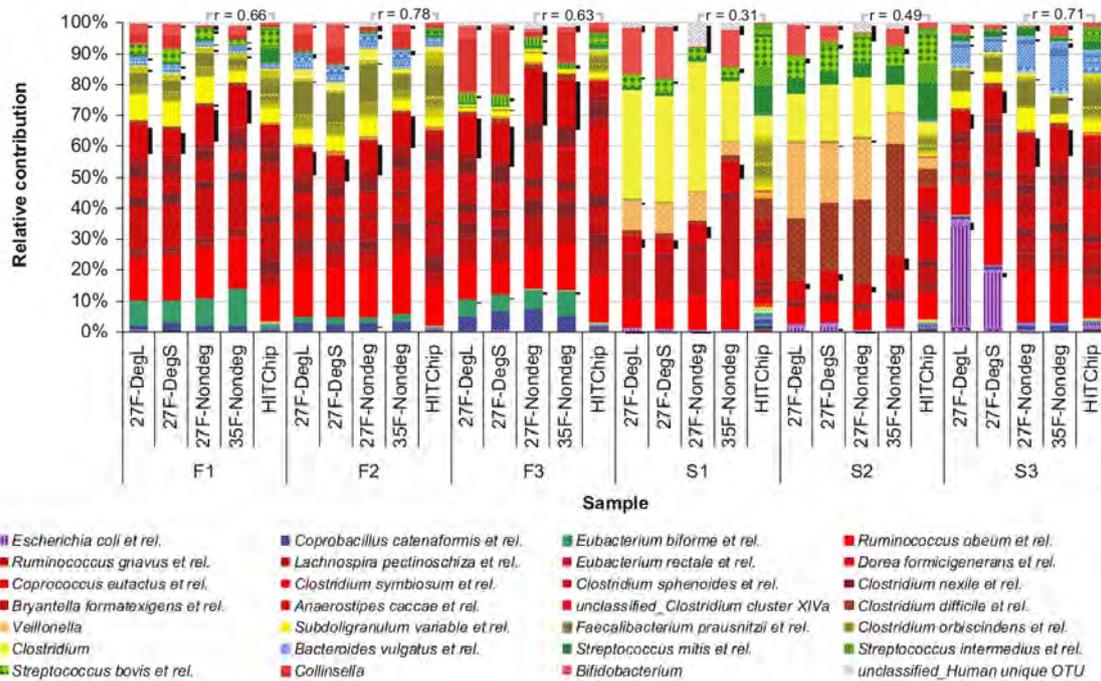


Figure S2.1. Relative contribution of detected bacterial phyla with pyrosequencing using four different forward primers and HITChip analysis for community data at level 2. Pyrosequences that could not be classified above the confidence threshold of 80% are grouped to “unclassified_” at the specific rank per taxon and are indicated in the microbial profiles with a shadow. Phylogenetic groups that contribute at least 5% to one of the profiles are indicated in the legend.

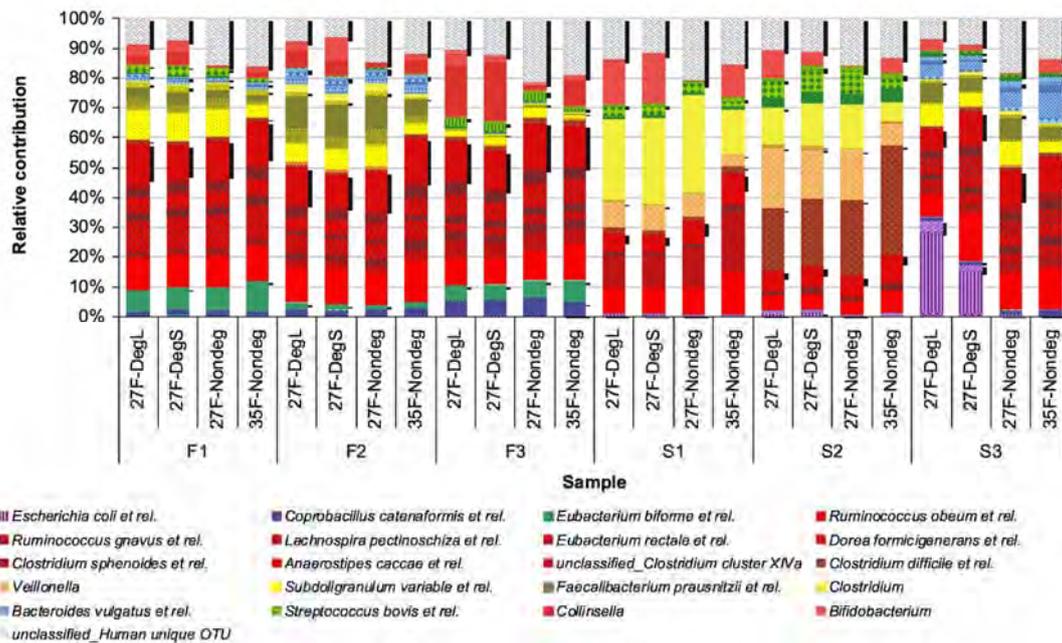


Figure S2.2. Relative contribution of detected bacterial phyla with pyrosequencing using four different forward primers for community data at level 2 after bypassing the quality filtering step.

Pyrosequences that could not be classified above the confidence threshold of 80% are grouped to “unclassified_” at the specific rank per taxon and are indicated in the microbial profiles with a shadow. Phylogenetic groups that contribute at least 5% to one of the profiles are indicated in the legend.

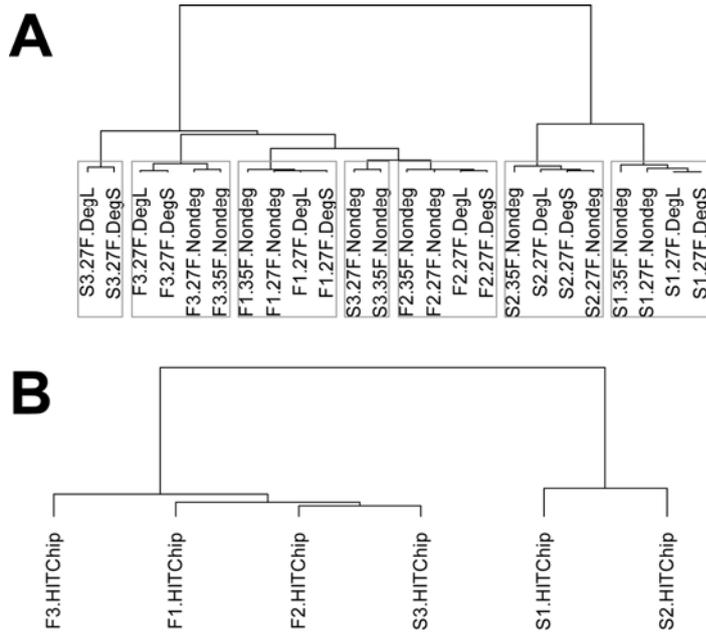


Figure S2.3. Hierarchical clustering of the microbial profiles from (A) pyrosequencing using four different forward primers and (B) HITChip analysis.

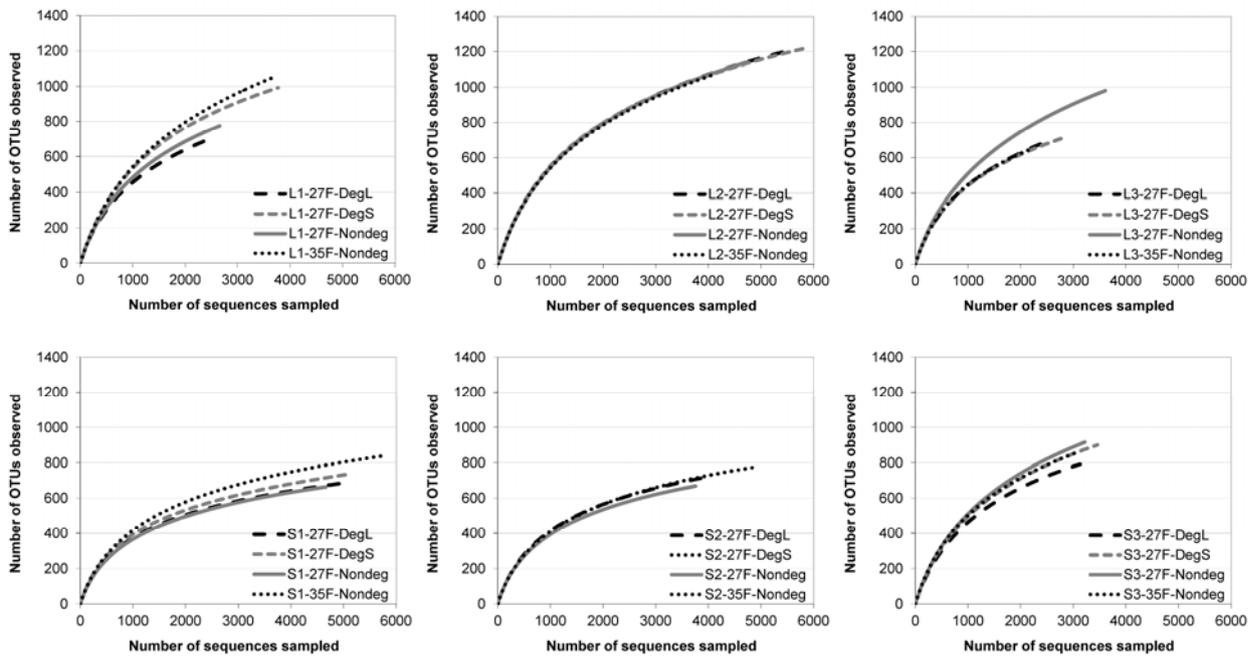


Figure S2.4. Rarefaction curves at 0.02 distance level generated for Fecal samples (F1-F3), ileostoma effluent (S1-S2), and ileal lumen content (S3). The forward primers used for the respective PCR are indicated with different colors.

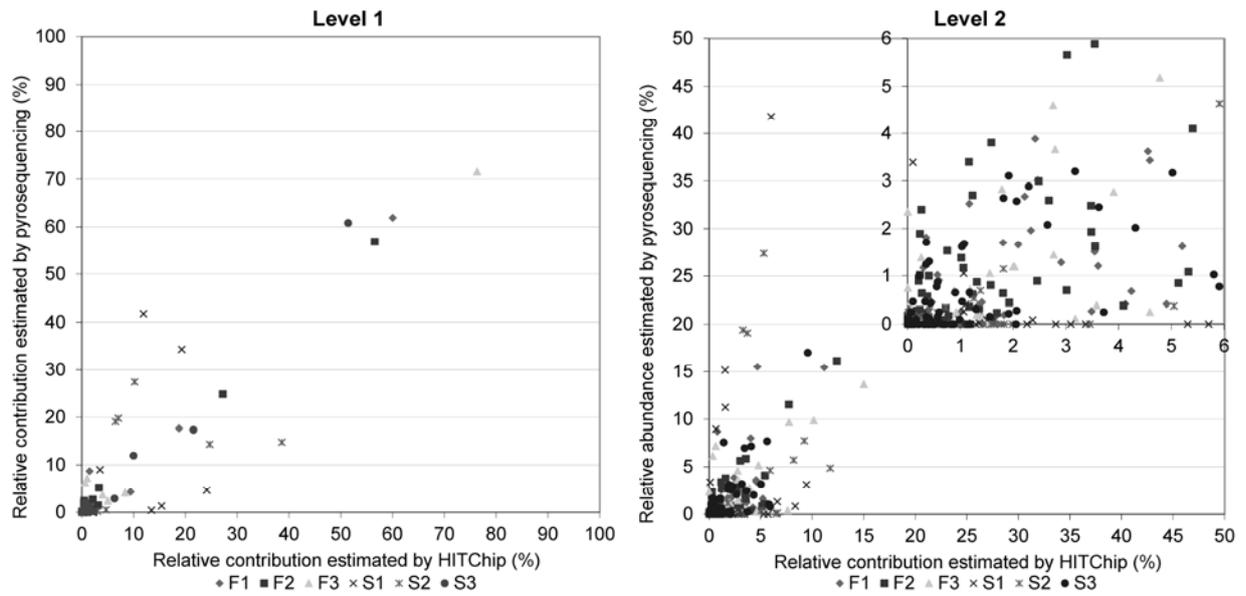


Figure S2.5. Comparison of relative contributions estimated by pyrosequencing versus HITChip analysis for community data of level 1 and 2. Data points for groupings of pyrosequences that could not be classified above the confidence threshold of 80% (Unclassified_) were removed because these groups are not represented in the output of HITChip analysis. The inset shows relative contributions for community level 2 from 0 to 1%.

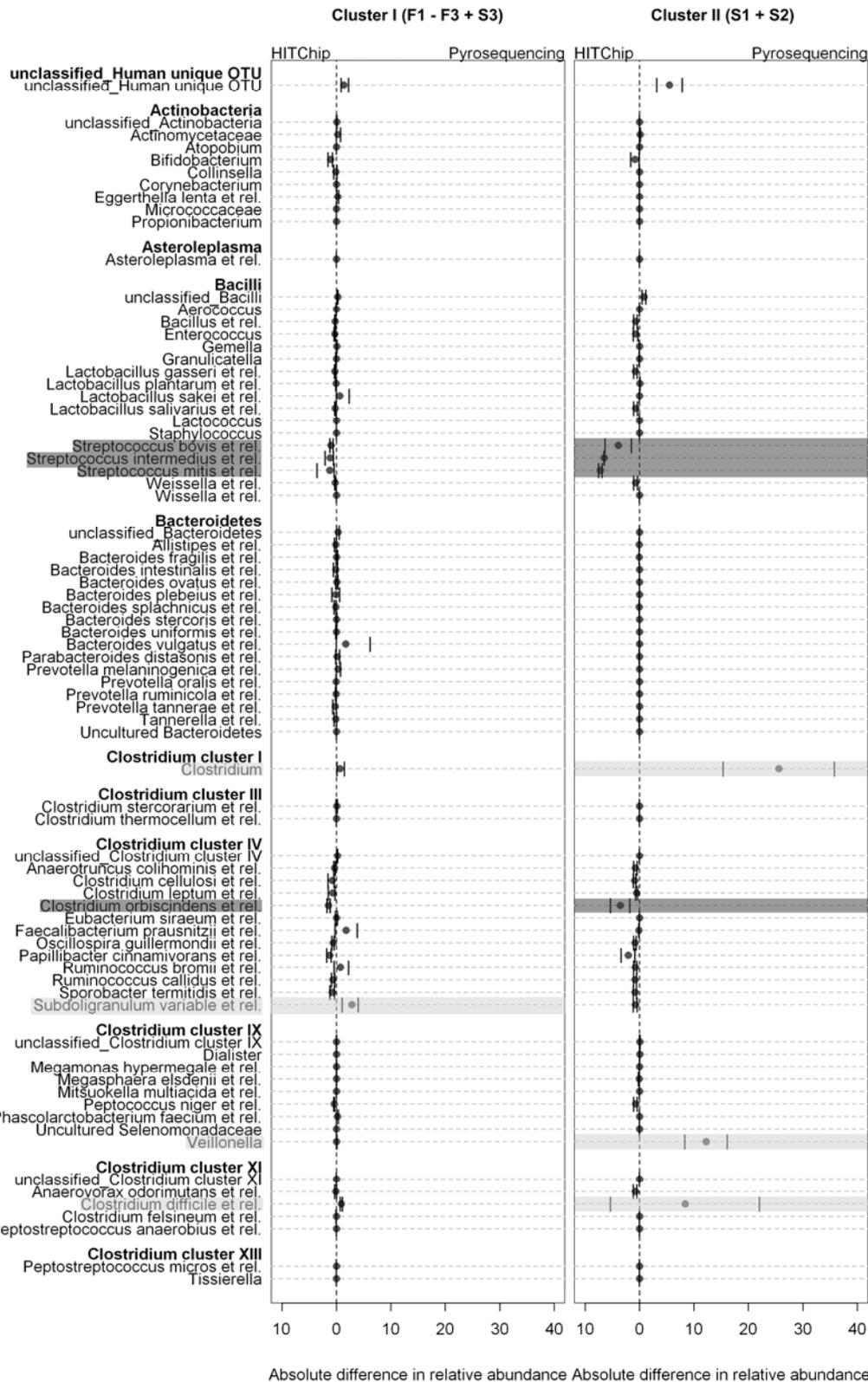


Figure S2.6. Difference in relative abundances of phylogenetic groups as estimated by pyrosequencing and HITChip analysis in cluster I and II. Blue circles represent the average difference per phylogenetic group with the highest and lowest observed difference values from each cluster indicated with vertical lines. Phylogenies highlighted in lightgrey and darkgrey represent groups for which relative abundances were estimated >2.5% higher by pyrosequencing and HITChip analysis, respectively.

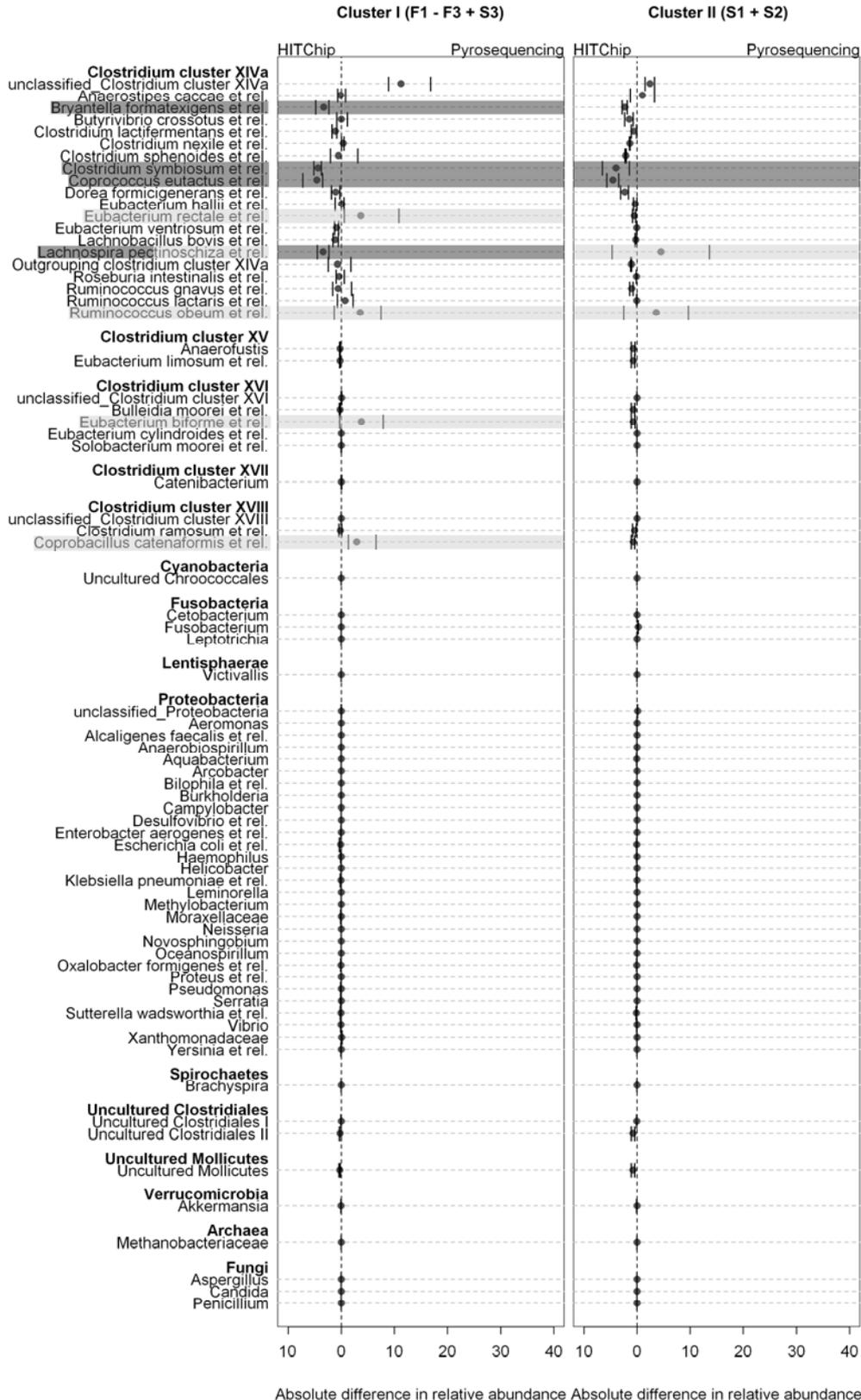


Figure S2.6 (Continued). Difference in relative abundances of phylogenetic groups as estimated by pyrosequencing and HITChip analysis in cluster I and II. Blue circles represent the average difference per phylogenetic group with the highest and lowest observed difference values from each cluster indicated with vertical lines. Phylogenies highlighted in lightgrey and darkgrey represent groups for which relative abundances were estimated >2.5% higher by pyrosequencing and HITChip analysis, respectively.

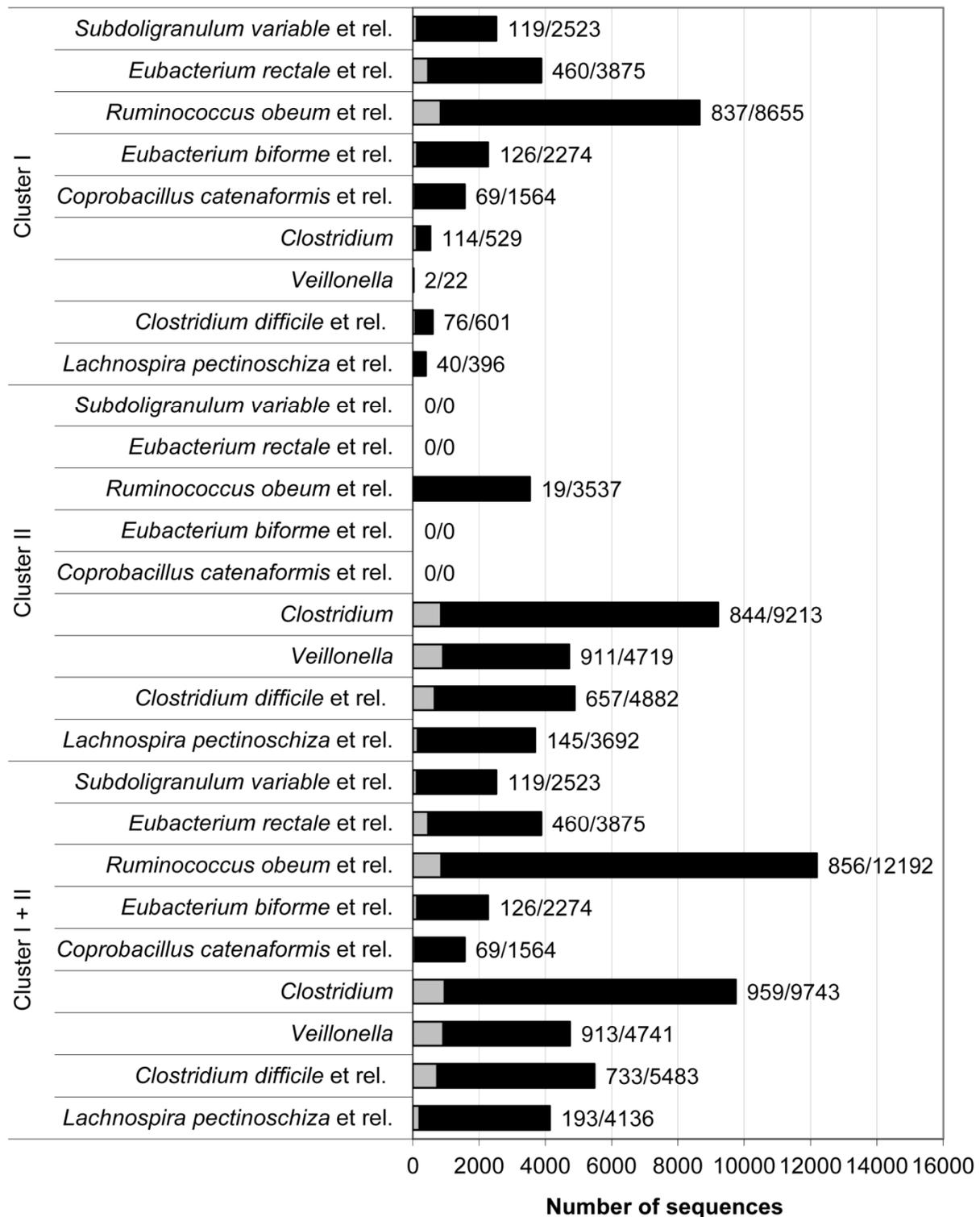


Figure S2.7 Total number of pyrosequences with (black) or without (grey) exact matches with HITChip probes per phylogenetic group. The ratio of number of sequences without exact matches with HITChip probes vs. total number of pyrosequences is indicated for each specific level 2 group.

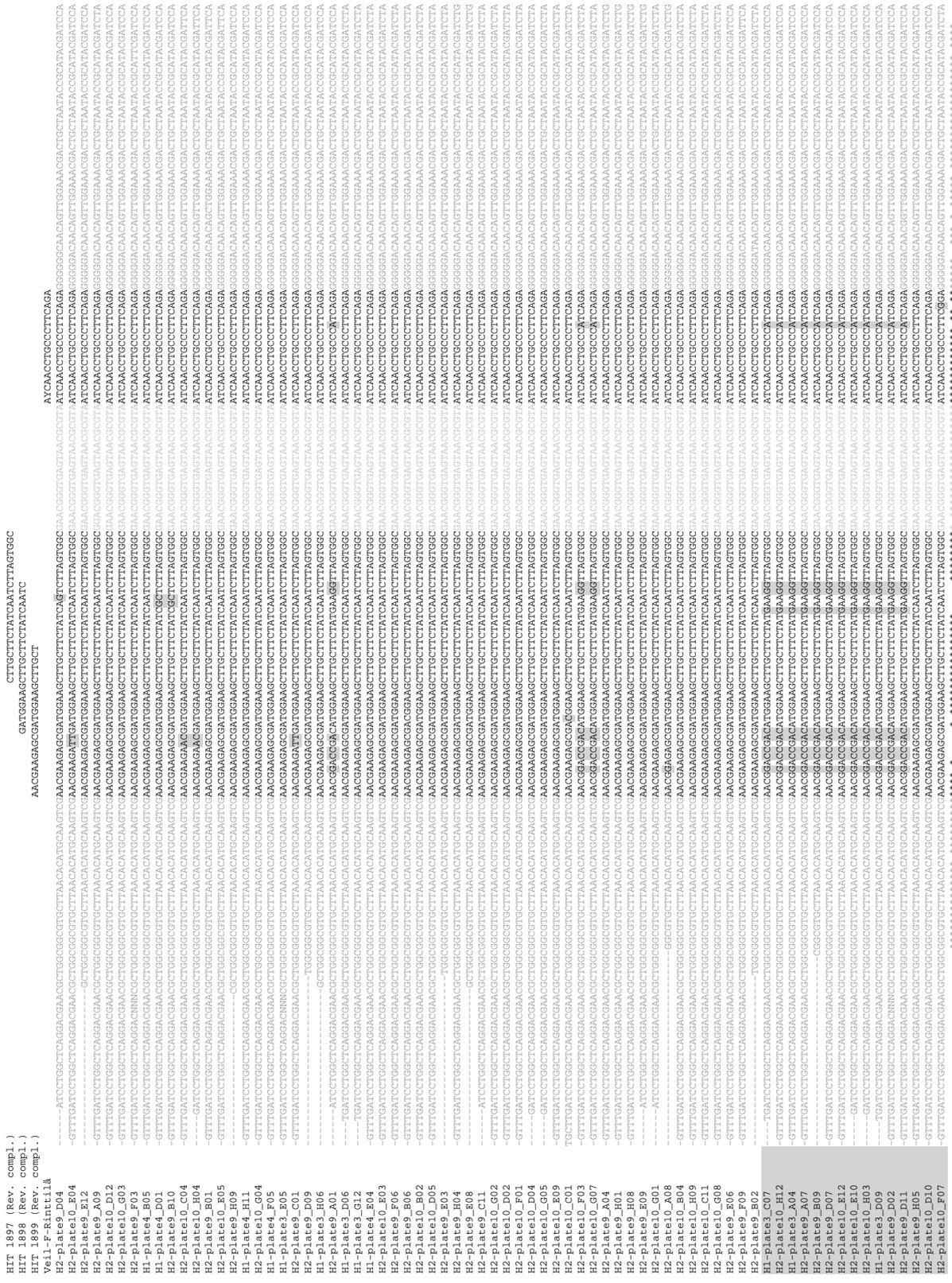


Figure S8a: Alignment of cloned 16S rRNA gene sequences classified as *Veillonella* from ileostomy sample S1 and S2 with binding sites of HITChip probe HIT 1897, HIT 1898, HIT 1899 (reverse complement), and qPCR primer Veil-F-Rinttilä. Nucleotides that differ from the probes/primers have been highlighted in grey. Sequences highlighted in yellow represent sequences with <98% identity to the 16S rRNA gene sequences represented in the human unique OTU database.



HIT 737 (Rev. compl.) CAGAACAGAGATGGTTCCCTTTC

H2-plate9_D04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_E04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_E12 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_A09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_D12 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G03 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_F03 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate4_B05 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate4_D01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_B10 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_C04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_H04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_B01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_E05 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_H09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate4_H11 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate4_F05 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate3_E05 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_C01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_D09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate3_H06 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_A01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate3_D06 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate3_G12 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate4_E04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_E03 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_F06 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_B06 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_B02 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_D05 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_E03 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_H04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_E08 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_C11 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G02 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_D02 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_F01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_D04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G05 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_E09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_C01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_F03 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G07 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_A04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_H01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_H08 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_E09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G01 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_A08 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_B04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_H09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_C11 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate10_G08 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_E06 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_F09 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_B02 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H1-plate3_C07 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate10_H12 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H1-plate3_A04 ACCTTACCAGGTCTTGACATTGATGGA CAGAACTAGAGATAGTTCCCTTTC TCGGAAGC

H2-plate9_A07 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate9_B09 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate9_D07 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate10_E12 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate10_E10 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate10_H03 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H1-plate3_D09 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate9_D02 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate9_D11 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate9_H05 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate10_D10 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

H2-plate10_F07 ACCTTACCAGGTCTTGACATTGATGGA CAGGTCCAGAGATGGACTCTCTTC TCGGAAGC

Figure S8c: Alignment of cloned 16S rRNA gene sequences classified as *Veillonella* from ileostomy sample S1 and S2 with binding sites of HITChip probe HIT 737 (reverse complement). Nucleotides that differ from the probes/primers have been highlighted in grey. Sequences highlighted in yellow represent sequences with <98% identity to the 16S rRNA gene sequences represented in the human unique OTU database.



Veil-R-Rinttilä (Rev. compl.) AAGCTCTGTAAATCGGGACG

H2-plate9_D04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_E04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_E12 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_A09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_D12 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G03 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_F03 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate4_B05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate4_D01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_B10 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_C04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_H04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_B01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_E05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_H09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate4_H11 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate4_F05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate9_E05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_C01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_D09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate3_H06 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_A01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate3_D06 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate3_G12 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate4_E04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_E03 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_F06 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_B06 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_B02 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_D05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_E03 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_H04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_E08 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_C11 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G02 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_D02 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_F01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_D04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_E09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_C01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_F03 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G07 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_A04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_H01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_H08 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_E09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G01 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_A08 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_B04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_H09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_C11 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_G08 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_E06 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_F09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_B02 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate3_C07 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_H12 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate3_A04 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_A07 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_B09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_D07 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_E12 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_E10 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_H03 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H1-plate3_D09 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_D02 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_D11 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate9_H05 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_D10 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

H2-plate10_F07 ATGACGGCCTTCGGGTT-GTA AAGCTCTGTAAATCGGGACG AATGGTCTCTTGCGGAA-T

Figure S8d: Alignment of cloned 16S rRNA gene sequences classified as *Veillonella* from ileostomy sample S1 and S2 with binding sites of and qPCR primer Veil-R-Rinttilä. Nucleotides that differ from the probes/primers have been highlighted in grey. Sequences highlighted in yellow represent sequences with <98% identity to the 16S rRNA gene sequences represented in the human unique OTU database.

Supplementary tables

Table S2.1. Characteristics of sequence analysis before and after quality filtering

Sample and primer	Characteristic before quality filtering					Characteristics after quality filtering										
	No. of sequences	Total number of bp	Sequence length (nt)			No. of sequences	Total Number of bp	Avg sequence length (nt)	% remaining quality filtered sequences	OTU ^a	Chao1 ^a			Shannon diversity index		
			Avg	SD	Upper limit						Lower limit	Value	Upper limit		Lower limit	
F1																
27F-DegL	5051	1256543	248.77	105.78	354.55	143.00	2562	604812	236.07	50.72	712	1029	1141	946	1049	3.139
27F-DegS	7554	1918545	253.98	105.43	359.41	148.55	3767	919552	244.11	49.87	993	1420	1553	1318	1341	3.168
27F-Nondeg	5465	1268611	232.13	114.40	346.54	117.73	2652	602139	227.05	48.53	775	1078	1181	1002	1118	2.981
35F-Nondeg	7909	1742813	220.36	118.15	338.51	102.21	3747	821168	219.15	47.38	1060	1572	1717	1460	1554	3.019
F2																
27F-DegL	10425	2578982	247.38	98.65	346.03	148.74	5431	1269039	233.67	52.10	1199	1591	1704	1503	1587	3.269
27F-DegS	11591	2992590	258.18	100.93	359.11	157.25	5803	1425282	245.61	50.06	1217	1618	1735	1527	1597	3.275
27F-Nondeg	9524	2191190	230.07	110.93	341.00	119.14	4704	1058430	225.01	49.39	1139	1572	1697	1476	1565	3.231
35F-Nondeg	7932	1812673	228.53	110.98	339.51	117.54	4071	911738	223.96	51.32	1070	1569	1715	1457	1524	3.190
F3																
27F-DegL	4782	1103655	230.79	100.25	331.04	130.54	2595	568125	218.93	54.27	703	1103	1245	997	1058	2.845
27F-DegS	5349	1217057	227.53	105.71	333.24	121.82	2784	612636	220.06	52.05	709	1036	1156	949	1019	2.828
27F-Nondeg	7430	1522469	204.91	110.82	315.73	94.09	3606	749871	207.95	48.53	981	1424	1553	1324	1487	2.832
35F-Nondeg	4584	923238	201.4	113.57	314.97	87.84	2191	443231	202.30	47.80	649	985	1111	894	950	2.774
S1																
27F-DegL	10393	2469261	237.59	96.99	334.58	140.60	5016	1089616	217.23	48.26	687	910	1008	842	864	2.105
27F-DegS	10320	2512124	243.42	96.71	340.13	146.72	5136	1166462	227.11	49.77	738	978	1076	908	960	2.084
27F-Nondeg	9593	2084989	217.34	100.03	317.38	117.31	4660	947405	203.31	48.58	664	838	914	785	844	1.914
35F-Nondeg	11378	2663271	234.07	108.65	342.72	125.42	5822	1321608	227.00	51.17	842	1067	1154	1004	1056	2.190
S2																
27F-DegL	7925	1949842	246.04	97.97	344.01	148.06	3993	907622	227.30	50.38	722	964	1061	894	928	2.329
27F-DegS	10053	2423069	241.03	101.26	342.29	139.77	4930	1118673	226.91	49.04	778	973	1051	917	980	2.385
27F-Nondeg	7628	1744928	228.75	106.49	335.24	122.26	3752	811675	216.33	49.19	669	852	929	797	862	2.160
35F-Nondeg	10294	2345647	227.87	113.33	341.19	114.54	4509	973935	216.00	43.80	758	965	1047	907	983	2.285
S3																
27F-DegL	5924	1467588	247.74	100.82	348.55	146.92	3127	734417	234.86	52.79	792	1079	1179	1005	1092	2.899
27F-DegS	6501	1563131	240.44	105.29	345.73	135.16	3456	813226	235.31	53.16	901	1301	1428	1204	1294	3.021
27F-Nondeg	6759	1475446	218.29	114.20	332.49	104.09	3217	697839	216.92	47.60	917	1374	1514	1267	1355	3.159
35F-Nondeg	6288	1376912	218.97	114.26	333.24	104.71	3072	657367	213.99	48.85	859	1196	1305	1114	1215	3.163

^a: Calculated at an 0.02 distance level.^b: 95% CI, 95% confidence interval.

Table S2.2. Percentage of sequences remaining after quality filtering applying all criteria or a single criterion

Sample and primer	Percentage of pyrosequences remaining after quality filtering			
	All criteria	Exact matches to the forward primer	No ambiguous bases (N)	Read-lengths no longer or shorter than 1 SD from the average sequence length
F1				
27F-DegL	50.72	91.84	85.19	64.36
27F-DegS	49.87	91.20	83.45	65.45
27F-Nondeg	48.53	95.08	85.53	60.49
35F-Nondeg	47.38	96.28	85.91	57.98
F2				
27F-DegL	52.10	92.29	84.45	66.40
27F-DegS	50.06	89.01	83.45	65.86
27F-Nondeg	49.39	95.86	84.48	60.90
35F-Nondeg	51.32	95.78	86.14	61.88
F3				
27F-DegL	54.27	92.74	88.12	65.68
27F-DegS	52.05	92.52	86.82	63.45
27F-Nondeg	48.53	96.26	88.17	56.33
35F-Nondeg	47.80	93.52	87.61	56.98
S1				
27F-DegL	48.26	92.36	83.40	64.36
27F-DegS	49.77	92.59	84.56	65.27
27F-Nondeg	48.58	96.21	83.14	62.82
35F-Nondeg	51.17	95.63	87.24	62.50
S2				
27F-DegL	50.38	92.62	84.68	64.69
27F-DegS	49.04	92.13	84.55	63.87
27F-Nondeg	49.19	96.26	84.86	61.60
35F-Nondeg	43.80	88.20	85.79	60.23
S3				
27F-DegL	52.79	91.29	85.08	66.17
27F-DegS	53.16	92.39	86.53	64.22
27F-Nondeg	47.60	94.35	86.45	57.35
35F-Nondeg	48.85	95.80	85.38	59.21

Chapter 3

Congruency of phylogenetic composition and activity patterns in the small intestine

Milkha M. Leimena*, Bartholomeus van den Bogert*, Jos Boekhorst, Eddy J. Smid, Erwin G. Zoetendal, and Michiel Kleerebezem

*These authors contributed equally to this work

Manuscript for publication in preparation

Abstract

The human small intestine is the main site for food absorption and digestion. Here, samples collected from four ileostomy subjects at four time points were used to study the phylogenetic dynamics of total and active fractions of the small intestinal microbiota through pyrosequencing of 16S ribosomal RNA gene (rDNA) and ribosomal RNA (rRNA) combined with elucidation of the specific activity patterns through metatranscriptomics. The community composition as assessed from rDNA, rRNA, and mRNA patterns appeared to be similar, indicating that the dominating microbial community is also highly active in this ecosystem. However, species richness and diversity metrics of the active ileostoma microbiota were reduced compared to the total microbiota, suggesting that some low abundance populations may represent less active members of the microbial community. Although each subject displayed a distinct microbial community, the genera *Streptococcus*, *Veillonella* as well as clostridial community members not only co-occurred in all ileostoma samples but also in most activity profiles, indicating that these are typical and active commensals of the human small intestine. While specific metabolic functions were assigned to distinct phylogenetic fractions in different samples, the overall activity patterns of the samples appeared relatively similar, demonstrating a high degree of functional redundancy within the ecosystem. This study integrates the complementary 16S rRNA and metatranscriptome based community reconstruction and provides insights into the functional biodiversity of the intestinal microbiota and enables the detection of metabolic interactions between its community members.

Introduction

The human gastrointestinal (GI) tract is populated with complex microbial communities, collectively referred to as microbiota, that predominantly consists of bacteria (341). During the last decades, 16S rRNA gene-based technologies have been extensively used to profile the human intestinal microbiota in health and disease (266, 322). More recently, metagenomics of this ecosystem has provided a wealth of knowledge, including a comprehensive gene catalogue of the human fecal microbiota (262), and clues about key microbiota members that could influence host health and disease (35, 147).

Metatranscriptomics may complement DNA-based metagenome analysis by identifying the active microbial members and unravelling their response to certain environmental conditions (106, 236, 356). Using next generation sequencing (NGS), in-depth metatranscriptomics of mRNA- or rRNA-derived cDNA sequences were successfully performed to investigate the activities of different complex microbial communities in marine (99, 105), soil (17, 333), and human gastrointestinal tract (112, 331) environments.

In general, fecal samples are often employed as representative samples to study the human gut microbiota. However, several studies have revealed that fecal microbial communities can differ considerably from the microbial community at other locations in the GI tract, notably the small intestine (31, 339, 381). Considering that the major part of nutrient digestion and absorption occurs in the small intestine (203) and that its mucosa is the main immune sampling mucosal surface of the human GI tract (see (75) for a review), this region of the GI tract can be anticipated to be of great importance to study interactions between the host and the intestinal microbiota (see (58) for a recent review). Investigations in our laboratory have used samples collected from ileostomy subjects to study the composition and the function of the human small intestine. Ileostomy subjects are characterized by surgical resection of their colon and hence, have their terminal ileum connected to an abdominal stoma. As a result, luminal content from the small intestine is excreted into an appliance, providing a unique opportunity for repetitive and non-invasive sampling of the luminal microbiota of the small intestine (31, 339, 381). Our studies have revealed that the ileostoma microbiota resembles that of the proximal small intestine in healthy individuals, which is characterized by a less diverse and highly fluctuating community with bacteria belonging to the genera *Streptococcus* and *Veillonella* being enriched (31, 339, 341, 381). These genera were previously reported to metabolically interact through lactate production and utilization by the *Streptococcus* and *Veillonella* species, respectively (84, 381). Therefore, it is not surprising that these genera also co-occur in the oral cavity (169), throat (9), stomach (23), and esophagus (256) and thus seem to be important commensals of the upper digestive tract. Characterization of cultured isolates of the *Streptococcus* and *Veillonella* genera from the small intestine community, revealed substantial richness within the *Streptococcus* lineages present in the ecosystem ((340), Chapter 4). The genomes of representative isolates belonging to these lineages contained a diverse genetic capacity to import and utilize

(simple) carbohydrate substrates for growth, which was in excellent agreement with their phenotypic characteristics (Chapter 5). Preliminary, low depth metatranscriptomic analysis of ileostoma effluent employing cDNA clone-libraries revealed high expression of streptococcal carbohydrate transport functions, supporting a major role for the *Streptococcus* population in uptake and fermentation of the available dietary carbohydrates in the human small intestine (381). However, the limited depth of analysis in these studies prevented a more complete and comprehensive reconstruction of the activity profile of the streptococcal population, let alone the less abundant *Veillonella* spp.

In this study we applied an integrated approach that combined microbial profiling by pyrosequencing the 16S ribosomal RNA gene (rDNA) and ribosomal RNA (rRNA) content and direct illumina sequencing of cDNA derived from enriched mRNA of four ileostomy subjects at four time points to elucidate the small intestinal microbiota dynamics at the level of population composition as well as its specific activity pattern.

Materials and methods

Ethics statement

The study was approved by the University Hospital Maastricht Ethical Committee, and was conducted in full accordance with the principles of the 'Declaration of Helsinki' (52nd WMA General Assembly, Edinburgh, Scotland, October 2000). Subjects were informed about the study orally and in writing and signed a written informed consent before participation.

Collection of ileostoma effluent

A total of 16 ileostoma effluent samples were collected from four ileostomy subjects (2 male and 2 female; $66 \pm$ standard deviation of 9.0 years) who were colectomized at least 5 years prior to the sampling period, are clinically considered to be healthy, and have a normally functioning small intestine: subjects did not report any complaints related to GI functioning for at least three years prior to testing, and were not following any treatment for GI-related symptoms or specific dietary regime. The subjects donated four samples each, collected on two distinct time points during the day (morning and afternoon), two days apart. The subjects collected the ileostoma effluent in a clean, empty ileostoma appliance, which was emptied in centrifuge bottles (Nalgene, Rochester, NY, USA) containing 100 ml RNA*later*® (Ambion, Austin, TX, USA), immediately after the bulk of the effluent flowed into the appliance. Samples were gently homogenized and stored for 4-10 hours at room temperature, after which the samples were frozen by transferring the tubes to dry ice. Frozen samples were transported to the laboratory, where they were kept at -80°C until further analysis.

RNA and DNA extraction

Cell pellets were obtained from the RNA*later*-effluent samples by adding four volumes of PBS, followed by centrifugation at 4600g for 10 minutes using a Heraeus

Multifuge 3 S-R Centrifuge (DJB Labcare Ltd., England, UK). The cell pellet was re-suspended in 500 µl ice-cold TE buffer (Tris-HCl pH 7.6, EDTA pH 8.0). Total RNA and DNA was extracted from the resuspended cell pellet according to the Macaloid-based RNA isolation protocol (377) with the use of Phase Lock Gel heavy (5 Prime GmbH, Hamburg) (240) during phase separation. The aqueous phase was split in two aliquots up to 300 µl, one for RNA and one for DNA isolation. For the RNA extraction, the aqueous phase was purified using the RNeasy mini kit (Qiagen, USA), including an on-column DNaseI (Roche, Germany) treatment as described previously (377). Total RNA was eluted in 30 µl ice-cold TE buffer and the RNA quantity and quality were assessed using NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, USA) and Experion RNA Stdsens (Biorad Laboratories Inc., USA), respectively.

Total DNA extraction was preceded by treatment of the sample (300 µl aqueous phase) with 3 µl RNase A (10 mg/ml; Qiagen GmbH, Hilden, Germany) at 37°C for 15 minutes. Subsequent steps employed a modified version of the QIAamp DNA Stool Mini Kit (Qiagen GmbH, Hilden, Germany) protocol. Initially, 22.5 µl proteinase K (20 mg/ml; Ambion) and 300 µl buffer AL from QIAamp kit were added to the sample followed by incubation at 70°C for 10 minutes. After addition of 300 µl ethanol (VWR, Amsterdam, The Netherlands), the sample was transferred to a QIAamp column and centrifuged (13,000g, 1 minute, at room temperature). DNA pellets were washed subsequently with the AW1 and AW2 buffers from QIAamp kit, according to manufacturer's instructions. Finally, the DNA was eluted with 30 µl Nuclease Free Water (Promega).

Profiling small intestinal populations

For 16S rDNA based microbial composition profiling and 16S rRNA based microbial activity profiling, barcoded amplicons from the V1-V2 region of 16S rDNA and 16S rRNA genes were generated by PCR and reverse transcription PCR (RT-PCR), respectively, using the 27F-DegS primer ((339); Chapter 2) that was appended with the titanium sequencing adaptor A and a 8 nt sample specific barcode (121) at the 5' end, and an equimolar mix of two reverse primers (338R I and II (117) based on three previously published probes EUB 338 I, II and III (60); Table 3.1), that were 5'-extended with the titanium adaptor B.

Table 3.1. Adaptors and primers used in this study

Primer ^a	Primer sequence (5'-3') ^b	Reference
Adaptor A	CCATCTCATCCCTGCGTGTCTCCGACTCAG	Provided by GATC-Biotech
Adaptor B	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG	Provided by GATC-Biotech
27F-DegS	GTTYGATYMTGGCTCAG	(339); Chapter 2
338R-I	GCWGCCTCCCGTAGGAGT	(60, 117)
338R-II	GCWGCCACCCGTAGGTGT	

^a Primer names may not correspond to original publication

^b M = A or C; R = A or G; W = A or T; Y = C or T

PCRs were performed in a total volume of 100 μ l containing 1 \times HF buffer (Finnzymes, Vantaa, Finland), 2 μ l PCR Grade Nucleotide Mix (Roche, Diagnostics GmbH, Mannheim, Germany), 2 U of Phusion® Hot Start II High-Fidelity DNA polymerase, 500 nM of a forward and the reverse primer mix (Biologio BV, Nijmegen, The Netherlands), and 0.2-0.4 ng/ μ l of template DNA. The PCRs were performed using a thermocycler GS0001 (Gene Technologies, Braintree, U.K.) using an amplification program that consisted of an initial denaturation at 98°C for 30 s, 30 cycles of: denaturation at 98°C for 10 s, annealing at 56°C for 20 s and elongation at 72°C for 20 s, and a final extension at 72°C for 10 minutes. RT-PCRs of total RNA were performed using a one-step RT-PCR system (Access Quick, Promega, Leiden, The Netherlands) according to the manufacturer protocol, albeit with 30 amplification cycles instead of 40 and modified amplification steps, which consist of denaturation at 94°C for 10s, annealing at 56°C for 20s and elongation at 68°C for 20s.

The size of the PCR and RT-PCR products (~375 bp) was confirmed by gel electrophoresis using 5 μ l of the amplification-reaction mixture on a 1% (w/v) agarose gel containing 1 \times SYBR® Safe (Invitrogen, Carlsbad, CA, USA). PCR products were purified with the High Pure Cleanup Micro Kit (Roche) using 10 μ l Nuclease Free Water for elution, and quantified using a NanoDrop ND-1000 spectrophotometer (Nano-Drop Technologies, Wilmington, DE). To determine the variation introduced by PCR amplification on the reproducibility of the pyrosequencing technique the RNA and DNA from the morning ileostoma sample collected on day 1 of ileostomy subject 2 was amplified twice. Purified (RT-)PCR products were mixed in equimolar amounts followed by running the amplicons on an agarose gel, band-excision, and purification by the DNA gel extraction kit (Millipore, Billerica, MA, USA). Purified amplicon pools were pyrosequenced using a Genome Sequencer FLX in combination with titanium chemistry (GATC-Biotech, Konstanz, Germany).

The pyrosequencing data analysis was carried out with a workflow employing the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (38) using settings as recommended in the QIIME 1.2 tutorial with the following exceptions: reads were filtered for chimeric sequences using Chimera Slayer (118); and OTU clustering was performed with an identity threshold of 97%, using parameters as recommended in the QIIME newsletter of December 17th 2010 (<http://qiime.wordpress.com/2010/12/17/new-default-parameters-for-uclust-otu-pickers/>). Additional data handling was done using in-house developed Python and Perl scripts. The Ribosomal Database Project (RDP) classifier version 2.2 (355) was used for taxonomic classifications up to the genus level.

Hierarchical clustering of samples was performed using UPGMA with weighted UniFrac as a distance measure. Robustness of the clustering was estimated using jackknifing analysis (20 replicates) as implemented in Qiime 1.2.

Pyrosequencing reads were deposited in the NCBI sequencing read archive (SRA) and are available under accession number SRP023505.

Metatranscriptome analysis of the human small intestinal microbiota

The metatranscriptome datasets were generated by Illumina sequencing of 16 cDNA libraries derived from mRNA enriched samples of the ileostoma effluent microbiota. The mRNA enrichment was performed by the removal of 16S and 23S rRNA using sequence-based capture probes attached to magnetic beads (MICROBExpress™, Ambion, Applied Biosystem, Nieuwerkerk a/d IJssel, The Netherlands) using the manufacturer's protocols (356). The enriched mRNA was quantified spectrophotometrically (NanoDrop) and its quality was assessed by microfluidics-based electrophoresis system (Experion RNA StdSens; Biorad Laboratories Inc., USA).

Double stranded cDNA was synthesized using the Invitrogen's SuperScript® Double-Stranded cDNA Synthesis kit (Invitrogen), with addition of SuperScript® III Reverse Transcriptase (Invitrogen) and random priming using random hexamers (Invitrogen) as described previously (198, 371) followed by RNase A (Roche, Germany) treatment, phenol-chloroform extraction, and ethanol precipitation. Double stranded cDNA was quantified using the NanoDrop 1000 spectrophotometer and verified by the sequencing provider (GATC Biotech, Konstanz, Germany) using an Agilent 2100 Bioanalyzer (Agilent technologies Inc., Waldbronn, Germany).

Sixteen Illumina sequencing libraries were constructed from double-stranded cDNA (ds cDNA) according to the ChiP protocol (292) with insert size between 200-300bp. Each sequencing library was barcoded and four libraries belonging to the same subject were pooled and sequenced on a single flow cell. Sequencing of four flow cells was performed using Illumina HiSeq2000, which generate in average of ~50 million reads for each sample.

After removal of low quality reads, the 16 Illumina sequencing datasets were subjected to a metatranscriptome analysis pipeline as described in (Leimena and Ramiro-Garcia, et al. Submitted). The mRNA reads were assigned to the NCBI prokaryote genome database (October, 2012). The metatranscriptome analysis pipeline was performed in four steps, which includes removal of ribosomal RNA derived sequences (180), assignment of taxonomic origin of mRNA derived sequences, classification of mRNA derived sequences as coding (mapped within an annotated gene) and non-coding (mapped within intergenic regions), and assignment of function prediction to coding sequence reads. All reads that passed the rRNA removal step were defined as mRNA reads.

Taxonomic classifications were performed by aligning the mRNA reads using MegaBLAST and BLASTN (7) to a prokaryotic genome database consisting of bacterial and archaeal full and draft genomes. Minimal bit scores thresholds of 148 and 110 were used for phylogenetic and functional assignments at genus and family level, respectively (Leimena and Ramiro-Garcia, et al. Submitted). Predicted gene products of identified protein encoding genes were assigned to the Clusters of Orthologous Groups (COGs) by BLASTP searches against the COG database (320) using an e-value $<10^{-6}$ for COG assignments and to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (165) using KEGG Automatic Annotation Server (KAAS; <http://www.genome.jp/tools/kaas/>). Principal component analysis

(PCA) was performed employing Canoco 5 software (202) to assess correlations between COG identifiers (IDs) and subject using the relative abundance of COG IDs as response variables and subjects as explanatory variables. The identified genes that have KEGG annotation were subjected to metabolic pathway mapping using iPath v2 (<http://pathways.embl.de/iPath2.cgi>) (368). Gene expression level of the metabolic pathways was indicated by the line width, which was determined from the log 2 value of the read count of KEGG annotated gene.

Multivariate statistical analysis for metatranscriptome and pyrosequencing datasets

Redundancy analysis (RDA) was performed using Canoco 5 software (202) to assess correlations between 16S rDNA, 16S rRNA and mRNA datasets and sample characteristics. COG relative abundances obtained from metatranscriptome dataset and OTU relative abundance obtained from pyrosequencing were used as response variables and subject origins as explanatory variables. RDA analysis was also performed using phylogenetic assignments of 16S rDNA, 16S rRNA, and mRNA sequences at genus level as biological variables. Unrestricted Permutation Test was used to assess significance of the variables.

Results

Analysis of pyrosequencing reads from 16S rDNA and rRNA amplicons

The composition of the whole microbial community and the active fraction in ileostoma effluent samples collected from four ileostomy subjects at four different time points was analysed through pyrosequencing of 16S rDNA and rRNA, respectively. A total 369,184 quality filtered sequences were obtained with an average of 10,255 (\pm standard deviation of 4,650) sequences per sample. Over 99% of retrieved sequenced reads were assigned above the 80% confidence threshold using the RDP classifier to order level, while lower fractions were assigned to family (91%) and genus (72%) level. The majority (>98%) of the total sequences were assigned to the phyla Firmicutes (77.6%) and Proteobacteria (20.7%; Figure S3.1). The former predominantly comprised of *Streptococcus* (30.2%) and members belonging to the order Clostridiales (33.2%), while the latter was dominated by the genus *Escherichia/Shigella* (17.6%).

The similarity between the microbial composition in each sample was assessed using the unweighted (based on presence/absence) and weighted (based on relative abundances) UniFrac distance metrics (210, 211). Independent amplicon production from independent nucleic acid extractions of the morning ileostoma sample collected on day 1 from ileostomy subject 2, yielded highly similar microbial profiles (Figure S3.2) that tightly clustered together (Figure 3.1) and were generally more similar compared to the intra- and inter-subject similarity between microbial composition or activity profiles (Figure S3.3). Comparison of OTU numbers, Chao1 richness estimations, and Shannon diversity from different sample preparations showed less variation in these ecological metrics compared to that between samples collected

from the same subject (Figure S3.4). These results imply that PCR amplification and nucleic acid extraction dependent variation is strongly exceeded by the sample-specific variation, and underpins the technical reproducibility of these analyses.

Clustering of the weighted UniFrac distances showed that morning and afternoon ileostoma effluent samples collected over a period of three days from subject 3 grouped first by morning or afternoon and then by day of sampling (Figure 3.1), which is in agreement with previous observations by Booiijink, et al. (31). However, samples obtained from the other subjects did not exhibit similar clustering consistency (Figure 3.1), indicating that the degree of compositional fluctuation of the small intestinal microbiota over time differs between individuals. Furthermore, qualitative analysis of the microbial profiles revealed considerable changes in the abundances estimated for different phylogenetic groups within a 72 hour time-frame and even between morning and afternoon samples obtained during the same day (Figure 3.2A). The fluctuations of small intestine microbial community composition and activity profiles confirms that these ecosystem dynamics occur within a single day's timeframe in the ileostoma microbiota (31), which possibly relates to variations in the composition of the subject's diet intake that is expected to impact strongly on the microbiota of the small intestine.

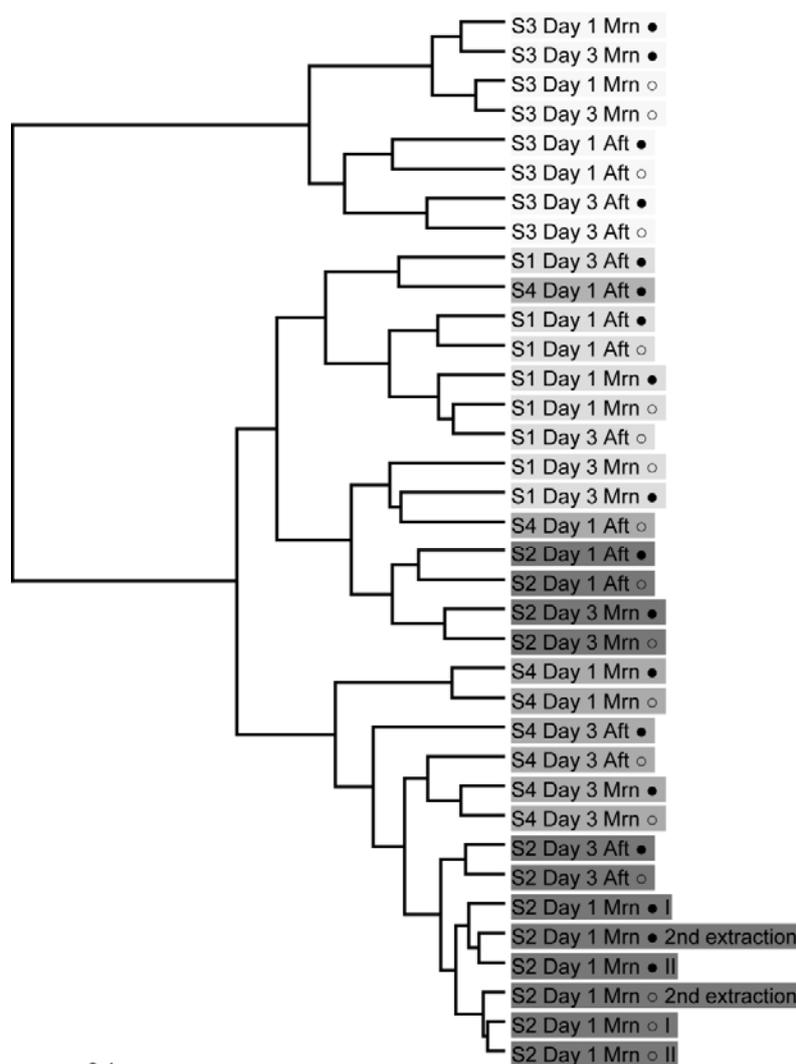


Figure 3.1. Hierarchical clustering of the microbial composition profiles derived from pyrosequencing of 16S rRNA amplicons representing the total microbial community (rDNA) and its active fraction (rRNA) in morning (Mrn) and afternoon (Aft) ileostoma samples. Leaves are colored according to subject. Closed circles represent total microbiota and open circles represent the active fraction of the small intestinal microbiota. Independent amplicon production (I and II) and nucleic acid extractions (I and 2nd extraction) were performed for DNA and RNA from the morning ileostoma sample collected on day 1 from ileostomy subject.

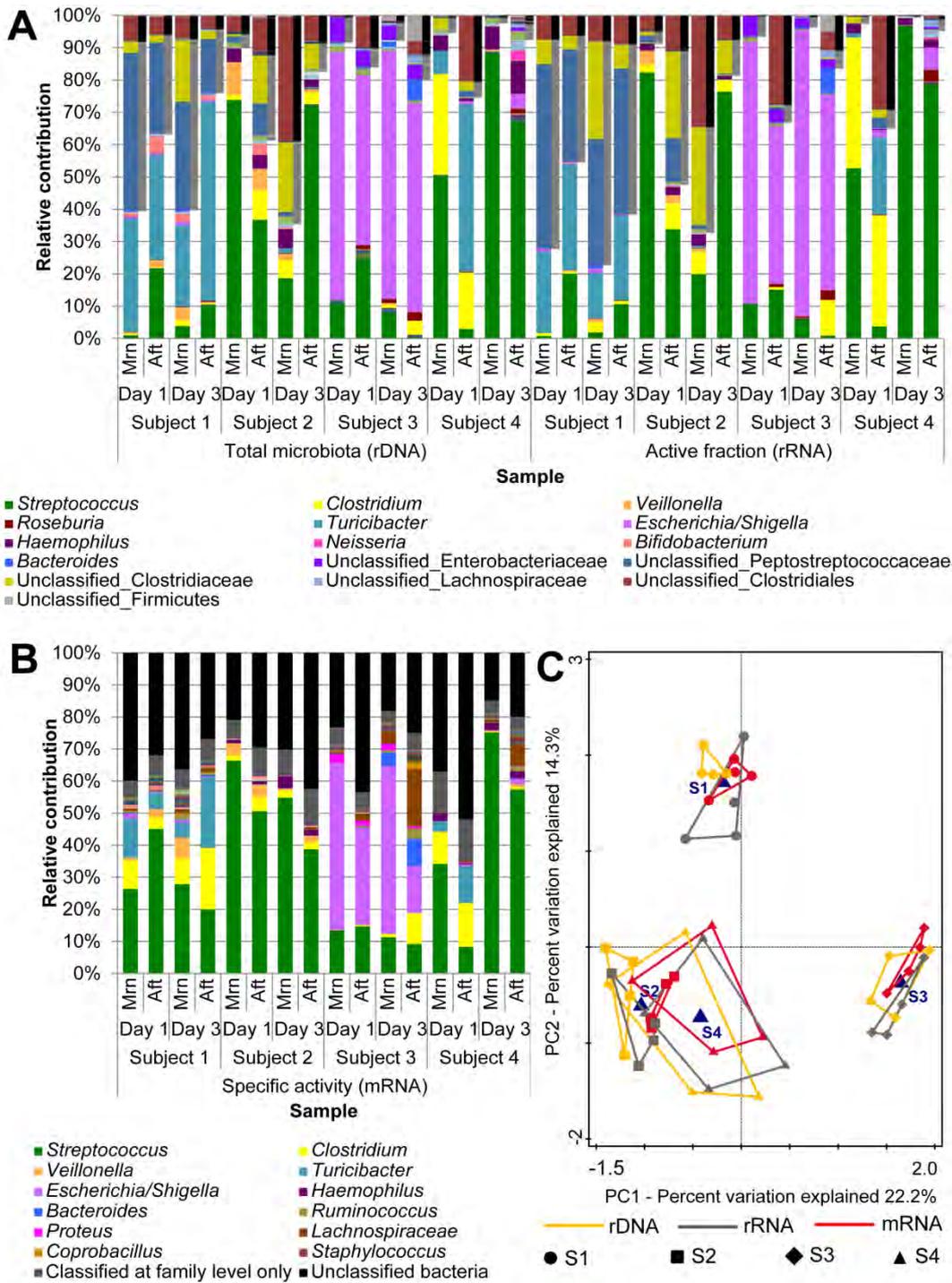


Figure 3.2. Relative contributions of detected bacterial taxa at genus level in afternoon (Mrn) and evening (Aft) ileostoma samples from four subjects with pyrosequencing (of the total (rDNA) and active fraction (rRNA) of the microbial community (A) and with sequencing of mRNA reads (B) as well as the first two principal components (PC 1 and PC 2) from RDA analysis based on genus level community data (C). Phylogenetic groups that contribute at least 2.5% to one of the profiles are indicated in the color key. Pyrosequences that could not be classified above the confidence threshold of 80% are grouped to “Unclassified_” at the specific rank per taxon, which is indicated in the microbial profiles with shadowing grey bars for groups classified no further than family level and black bars for groups with classifications that did not reach family level. mRNA reads with bit scores above 148 were used for taxonomic classification at genus level, scores between 110 and 148 are grouped to Family, while the scores lower than 110 were grouped as “Unclassified Bacteria”.

Weighted UniFrac-based hierarchical clustering (Figure 3.1) and principal coordinate analysis (PCoA) analysis (Figure 3.3) revealed a grouping by subject on basis of community composition, although the bacterial communities detected in samples from subject 1, 2, and 4 displayed considerable overlap. This is likely due to the relatively short phylogenetic distance between the microbial communities in samples from these subjects that were dominated by bacterial species belonging to Firmicutes, while *Escherichia* belonging to Proteobacteria predominated in the samples from subject 3. Unweighted UniFrac based PCoA analysis resulted in a more prominent separation of the microbial communities by subject (Figure 3.3), implying that the divergence in microbial composition between individuals is likely due to variations in the less abundant phylogenetic groups. These findings are in line with previous results (31), showing that each individual harbors a distinct small intestinal microbiota. This notion is clearly supported by the pyrosequencing based microbial profiles generated on basis of genus taxonomic assignments (Figure 3.2A) that indicate that samples from subject 1 were dominated by *Turicibacter* and members of the family Peptostreptococcaceae, while samples from subject 3 were abundant in *Escherichia/Shigella*. Although, the microbial communities in samples from subject 2 and 4 were enriched with *Streptococcus*, the divergence between these subjects could predominantly be attributed to less abundant genera such as *Bifidobacterium* (Figure 3.2A).

Despite the subject-specific phylogenetic composition of the ileostoma microbiota, the genera *Streptococcus*, *Veillonella*, and *Actinomyces*, were represented in the total microbial communities of all samples (Figure S3.5). Nevertheless, *Streptococcus* abundance estimates varied substantially between samples, ranging from 0.4% to 88.3% (Figure 3.2A). *Veillonella* and *Actinomyces* were generally present at lower relative abundances and ranged from <0.1% to 10.1% and <0.1% to 1.3% of the microbial community, respectively (Figure 3.2A). Bacteria belonging to the genera *Clostridium* and *Haemophilus* were detected in the total microbial communities of all but one sample (afternoon samples of day 1 from subject 2 and day 3 from subject 4, respectively) with variable relative abundances ranging from <0.1% to 31.1% and <0.1% to 10.2%, respectively. Furthermore, the genus *Escherichia* was detected in most samples collected from subject 1, 2, and 4, with relative abundance estimates up to 4.4% of the total microbial community. Notably, *Escherichia* assigned relative abundance in samples from subject 3 were considerably higher and ranged from 52.3% to 76.8% of the total ileostoma microbiota (Figure 3.2A). Analogously, the relative abundance estimates for *Turicibacter* (genus) and Peptostreptococcaceae (family) were generally low, although these bacterial groups were highly abundant in samples from subject 1 (25.2 – 61.5% and 16.6 – 48.7%, respectively; Figure 3.2A).

Streptococcus spp. were detected in the active fraction of each sample with abundances varying from 0.7% to 96.8%, while *Veillonella* and *Actinomyces* were detected in the active fraction of most, but not all ileostoma samples with relative contributions as high as 4.4% and 0.9%, respectively (Figure 3.2A and figure S3.5). *Clostridium* abundance estimates in the active fraction of the ileostoma microbiota

varied between subjects and could account for 40.4% of the active microbial fraction within a single sample. Relative abundances of the genera *Haemophilus* and *Escherichia* in the active fraction commonly amounted up to 3.5% and 6.8%, respectively, although samples obtained from subject 3 had *Escherichia* abundance levels of up to 88.7%. (Figure 3.2A). The high relative abundance of *Turicibacter* observed in samples obtained from subject 1 (see above), was also reflected by the high activity of this genus in the same samples (Figure 3.2A).

Considering the abundance and prevalence of specific genera, we conclude that *Streptococcus*, *Veillonella*, and *Clostridium* can be considered as (active) signature members of the small intestinal microbiota, while *Actinomyces* and *Haemophilus* represent groups that are also common but at a lower abundance. Notably, although the abundance estimates for the genera *Escherichia*, *Turicibacter* and the family of Peptostreptococcaceae was commonly low, specific subjects may have large communities of these bacterial groups.

Comparison of the total and active fraction of the small intestinal microbiota

The congruency between the total and active ileostoma microbial communities was evaluated. Based on observations in faecal microbial communities where microbial presence was not very well reflected by microbial activity (257, 316, 376), differences between the total and active fractions of the ileostoma microbial composition might also be anticipated. However, weighted UniFrac-based hierarchical clustering and principal component analysis revealed that the ileostoma effluent microbiota composition and activity profiles displayed close pairing per sample (Figure 3.1 and 3.3). This is illustrated by the highly similar composition of the total and active ileostoma communities (Figure 3.2A and S3.3) and suggests that the vast majority of the bacterial groups present in the small intestine are active players in shaping small intestinal physiology. Nonetheless, in-depth analysis revealed that species belonging to the genera *Grunulicatella* and *Actinomyces* were significantly less abundant in the active fraction of the ileostoma microbiota compared to the total ($p < 0.01$). This suggests that although bacterial groups (e.g. *Actinomyces*) may be present in all ileostoma samples (see above), their activity at the time of sampling may be limited.

The number of different taxonomic families and genera identified in the active fraction of each sample was consistently lower than that in the microbiota composition analysis (Table S3.1). This is also reflected in the lower number of OTUs, Chao1 richness estimations and Shannon diversity indices calculated for the active fraction relative to the corresponding values for the whole microbiota composition, suggesting that the active community is less rich and diverse compared to the total community residing in the ecosystem (Figure S3.6; Table S3.2). These metrics were not skewed due to differences in sequencing depth per sample between the active fraction and the total community (Table S3.2). The observations that the phylogenetic make-up of the active and total community profiles are highly similar (see above) and unweighted UniFrac distances calculated for total and active community pairs showed a high degree of variation (Figure S3.3), indicate that only low abundance phylogenetic groups of the total ileostoma microbiota were not detected in the active fraction of the

ileostoma microbial community. This notion was supported by the fact that out of the total of 163 genus-level assignments, 46 were not detected in the active fractions, representing an average relative abundance of maximally 0.065% of the total microbial community per sample (Data not shown).

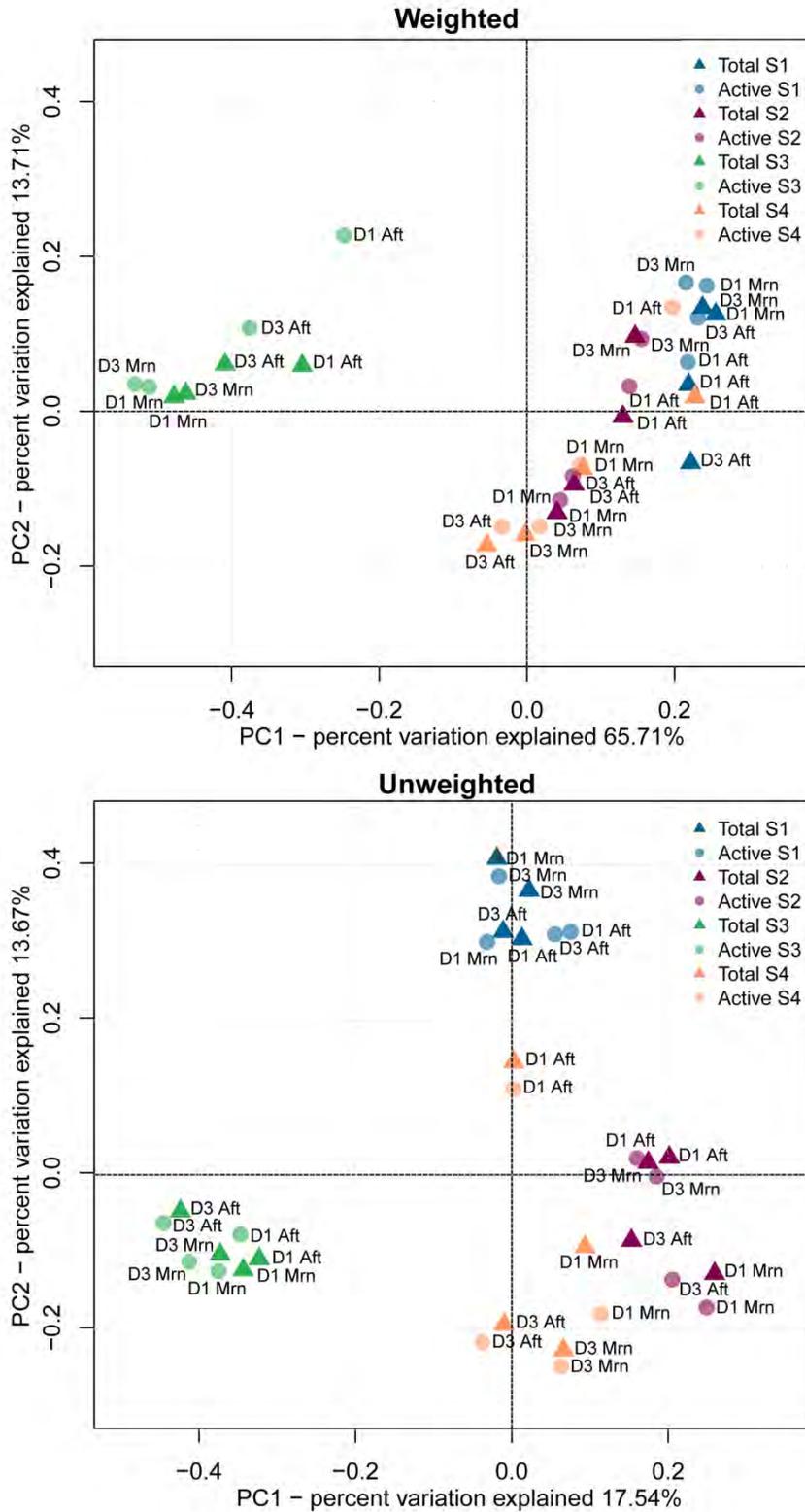


Figure 3.3. First two principal components (PC 1 and PC 2) from weighted and unweighted UniFrac-based PCoA analysis of the total microbial community (16S rDNA) and the active fraction (16S rRNA) in ileostoma effluent samples obtained from four subjects (S1-S4).

mRNA based phylogenetic profiling of the small intestinal microbiota

The active fraction of the human small intestinal microbiota was further investigated by metatranscriptome analysis through sequencing of Illumina cDNA libraries derived from enriched mRNA samples from all 16 samples used in this study. A total of 519,198,745 quality filtered sequences with an average of 32,449,922 (\pm standard deviation of 14,704,832) sequences per sample was obtained (read-length of 101nt). Due to incomplete rRNA removal during the mRNA enrichment using MICROBExpress (Ambion) (104, 198), an average of 93% (\pm standard deviation of 4%) metatranscriptome sequences were removed from the datasets, leaving an average of 2,364,118 (\pm standard deviation of 1,678,586) putative mRNA sequence reads that were subjected for further analysis (Table S3.3).

Between 35% to 78% of the total mRNA reads were classified at genus level while between 7% to 30% could be assigned no further than family level (Figure 3.2B). Analogous to the 16S rRNA (gene) based community profiles, phylogenetic distribution of mRNA reads at genus level revealed that streptococci were detected in datasets from all samples with relative abundances ranging from 7% and 70%. *Streptococcus* derived mRNA sequences dominated the metatranscriptome datasets from subject 2 and most samples from subject 1 and 4, while datasets from subject 3 were dominated by *Escherichia* with relative abundances ranging from 14% to 52%. Notably, datasets from subject 1 and the afternoon sample collected on day 1 from subject 4, were enriched with mRNA assigned to the genera *Clostridium* and *Turicibacter*, which were also abundant in the 16S rRNA (gene) based profiles (Figure 3.2A). Though mRNA based phylogenetic profiles contained a larger percentage of reads that could not be assigned at genus level compared to those from 16S rRNA (gene) based profiling, the same dominant genera were detected at all three levels of analysis performed here (rDNA, rRNA, and mRNA). Furthermore, RDA analysis of the genus level community data from pyrosequencing (16S rDNA and 16S rRNA) and from RNA-sequencing (mRNA) revealed a significant separation of microbial profiles (p-value <0.05) between different subjects (Figure 3.2C), implying that individual-specific variation of microbial composition consistently are the predominant discriminator between the datasets, irrespective of the level of analysis (rDNA, rRNA, or mRNA).

Functional analysis of ileostoma microbial communities

To obtain insight in the functional properties of the human small intestinal microbiota, genes in the prokaryote genome database that were significantly aligned with metatranscriptome mRNA sequence reads were assigned to COG (321) for functional analysis and KEGG (164) for metabolic mapping purposes. Of the number of mRNA sequence reads assigned to the prokaryote genome database, between 55% to 87% were assigned to genes that have COG annotation, whereas between 44% to 75% reads were assigned to genes that have KEGG annotation. In total, 3220 expressed COGs were identified in the metatranscriptome datasets, of which 1022 COGs were shared between all datasets. Notably, these shared COGs captured between 64% to 95% of the total COG-assigned mRNA reads, indicating

that the metatranscriptomes of all samples are dominated by common functions. These common functions were frequently assigned to different phylogenetic origins, implying prominent functional redundancy of the community members of the small intestine microbiota. Analogously, many of the shared COGs belonged to the functional category of “information storage and processing”, and were predominated by functions associated to translation, ribosomal structure, and biogenesis, which represent more or less universally conserved microbial functions. Another dominant COG category among the shared functions was associated with “metabolism” (between 32 to 48% of the total COG-assigned mRNA reads), encompassing functions related to carbohydrate transport and metabolism (between 22 to 48% of the metabolism category), which indicates that the shared metabolic activity profile is executed by variable members of the microbial ecosystem. Moreover, this finding underpins that fermentation of diet-derived carbohydrates is an important functional driver of the small intestinal microbiota (381).

PCA analysis of the COG distribution derived from all metatranscriptome datasets revealed grouping by subject (Figure S3.7) with subjects 1, 2, and 4 clustering more closely together and clearly distinct from subject 3. This function-expression based clustering is remarkably similar to that from weighted UniFrac-based PCoA analysis using community data from 16S rDNA and rRNA pyrosequencing (Figure 3.3). In fact, RDA analysis on the COG distributions and OTU distributions (derived from 16S rDNA and 16S rRNA sequences) from all samples sustained the grouping by subject (Figure 3.4). However, COG distributions per sample displayed less variation between samples from different subjects (inter) and the same subject (intra), compared to the variation observed by composition (rDNA) and global-activity (rRNA) profiles. This observation is in good agreement with the observation that the COGs that were shared by all samples, captured the majority of the metatranscriptome mRNA reads (see above).

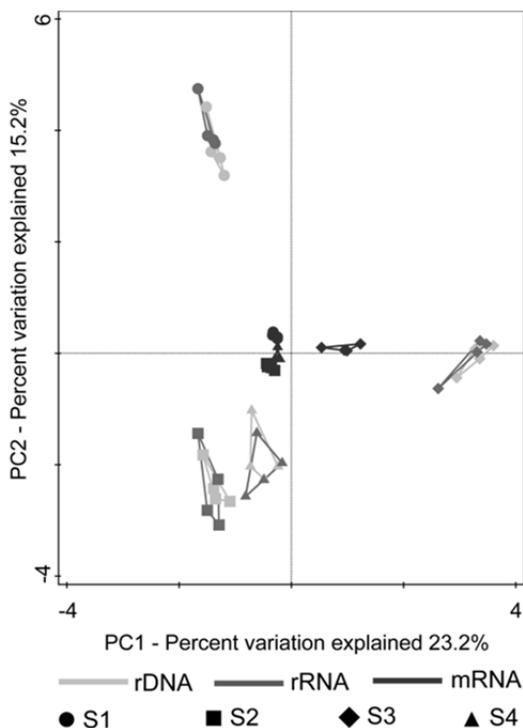


Figure 3.4. First two principal components (PC 1 and PC 2) from RDA analysis based on COG relative abundances obtained from metatranscriptome dataset and OTU relative abundance obtained from pyrosequencing from ileostoma effluent samples obtained from four subjects (S1-S4).

KEGG annotations were mapped onto metabolic pathways using the iPath mapping module (368). Exemplary metabolic pathway maps were generated for the datasets from the transcriptome of the morning sample of day 1 obtained from subject 2 and 4 (Figure 3.5). A high degree of similarity was observed for the metatranscriptomic landscapes of those two datasets for pathways related to nucleotide, carbohydrate, amino acid, energy, and lipid metabolism, as well as cofactor and vitamin synthesis. Nevertheless, distinct patterns of expression were observed for pathways related to oxidative phosphorylation and propanoate metabolism which were more prominently expressed in the sample from subject 2, whereas pathways related to metabolism of specific amino acid appeared to be higher expressed in the sample from subject 4 (Figure 3.5). These differences may reflect ecosystem adaptations to environmental differences such as variation in the dietary composition of subject 2 and 4.

Further analysis of the transcripts assigned to the genes encoding metabolic functions was performed to decipher genus specific metabolic activity in the small intestine ecosystem. Metabolic pathway-mapping specific for the genera *Streptococcus*, *Clostridium*, and *Veillonella* for the metatranscriptome dataset derived from the morning sample on day one from subject 2, revealed a strong domination of *Streptococcus* in pathways related to the primary metabolism of carbohydrate substrates. In contrast, metabolic pathways related to butanoate and propanoate metabolism were exclusively observed for *Clostridium* and *Veillonella*, respectively (Figure 3.6). These findings support the hypothesis that members of the small intestine microbiota intensely interact at a metabolic level, and illustrate the metabolic focus of the streptococci on the conversion of simple carbohydrate substrates to lactate and acetate, where the former metabolite serves as a substrate for subsequent conversion to acetate and propionate by *Veillonella* (381), while the latter metabolite (and in some cases also lactate) may serve as a substrate for butyrate production by the *Clostridium* members of the ecosystem (209). In fact, phosphotransbutyrylase and butyrate kinase involved in the butyrate formation pathway were found to be expressed by members of *Clostridium* cluster I.

The individual activities and proposed food-chain relationships between individual groups within the ecosystem are supported by high expression levels of sugar PTS transporters and glycolytic enzymes by *Streptococcus*, in combination with high level expression of the lactate import permease and membrane associated lactate conversion machinery of *Veillonella* as well as the butyrate formation pathway in *Clostridium*.

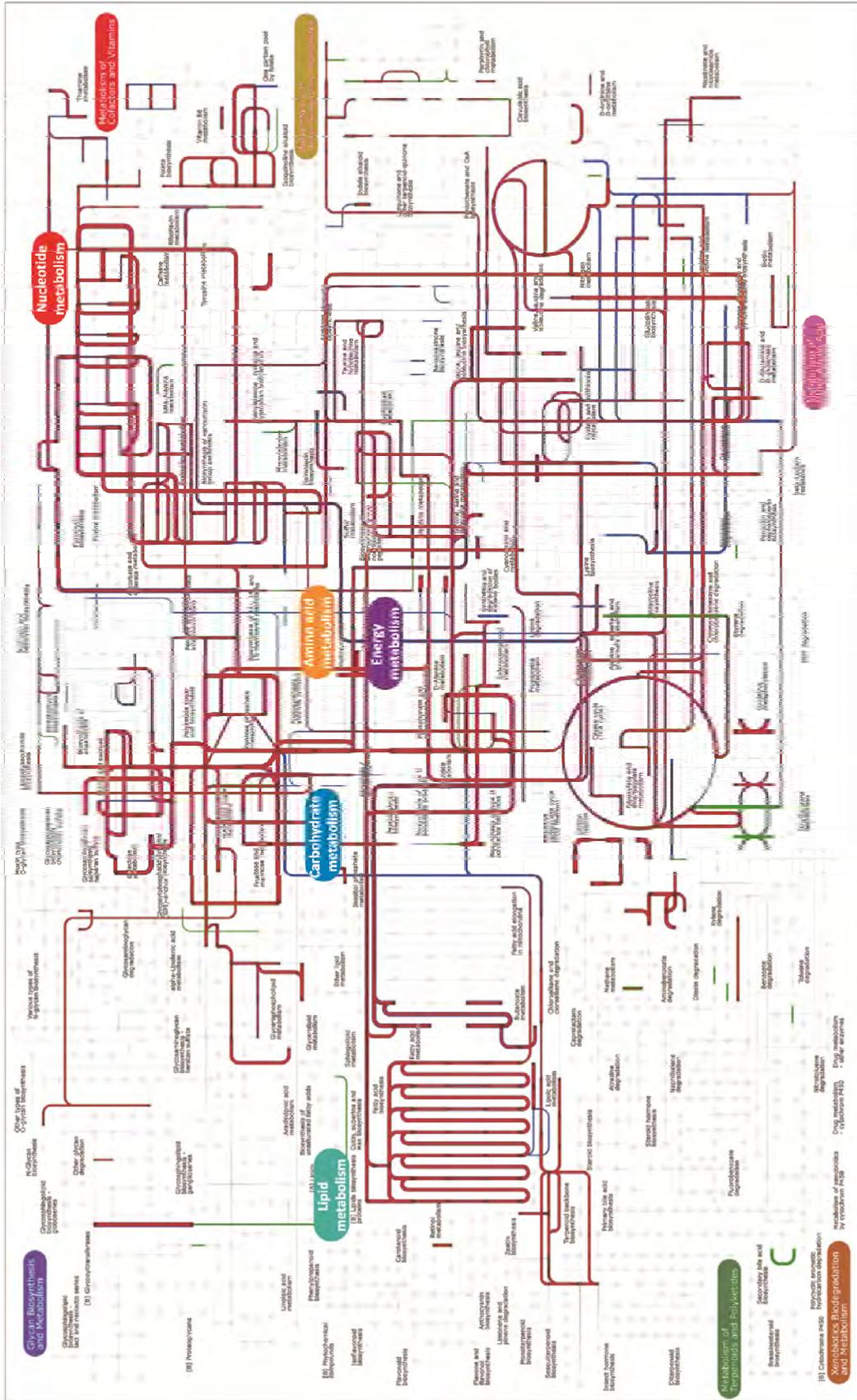


Figure 3.5. Metabolic pathways mapping of metatranscriptome datasets derived from the day one morning samples of subject 2 and 4. The majority of the metabolic pathways were overlapping between both datasets (red lines), while unique pathways for dataset of subject 2 or subject four were indicated as green and blue lines, respectively. The line width indicates gene expression level. Metabolic pathways were generated using iPath v2 based on KEGG annotations of expressed genes. A high resolution image is available upon request.

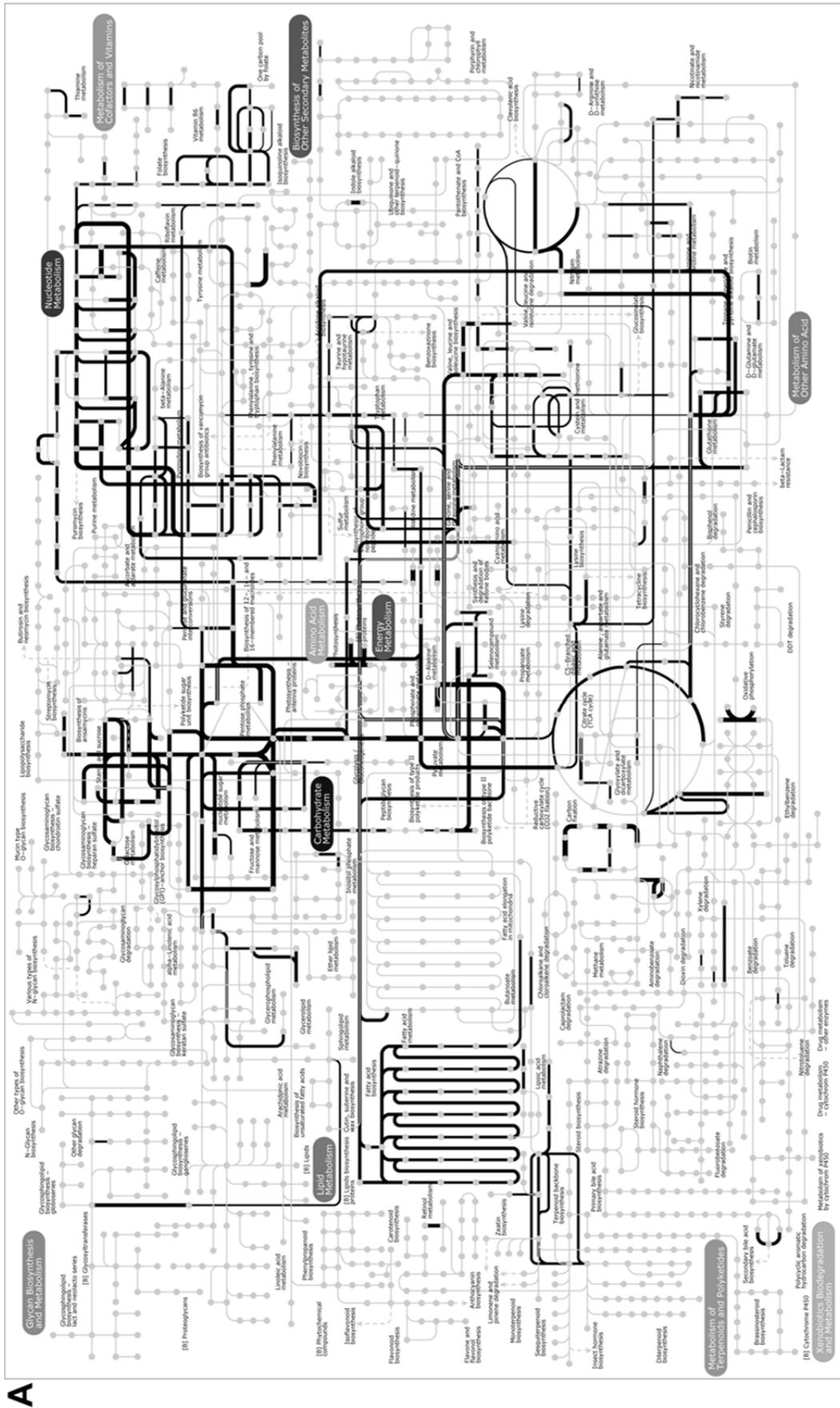


Figure 3.6. Example of genus specific metabolic pathways generated from metatranscriptome dataset of the morning day one sample of subject 2. Metabolic pathways of *Streptococcus (A)*, *Veillonella (B)*, and *Clostridium (C)*. The line width indicates gene expression level. Metabolic pathways were generated using iPath v2 based on KEGG annotations of expressed genes. A high resolution image is available upon request.

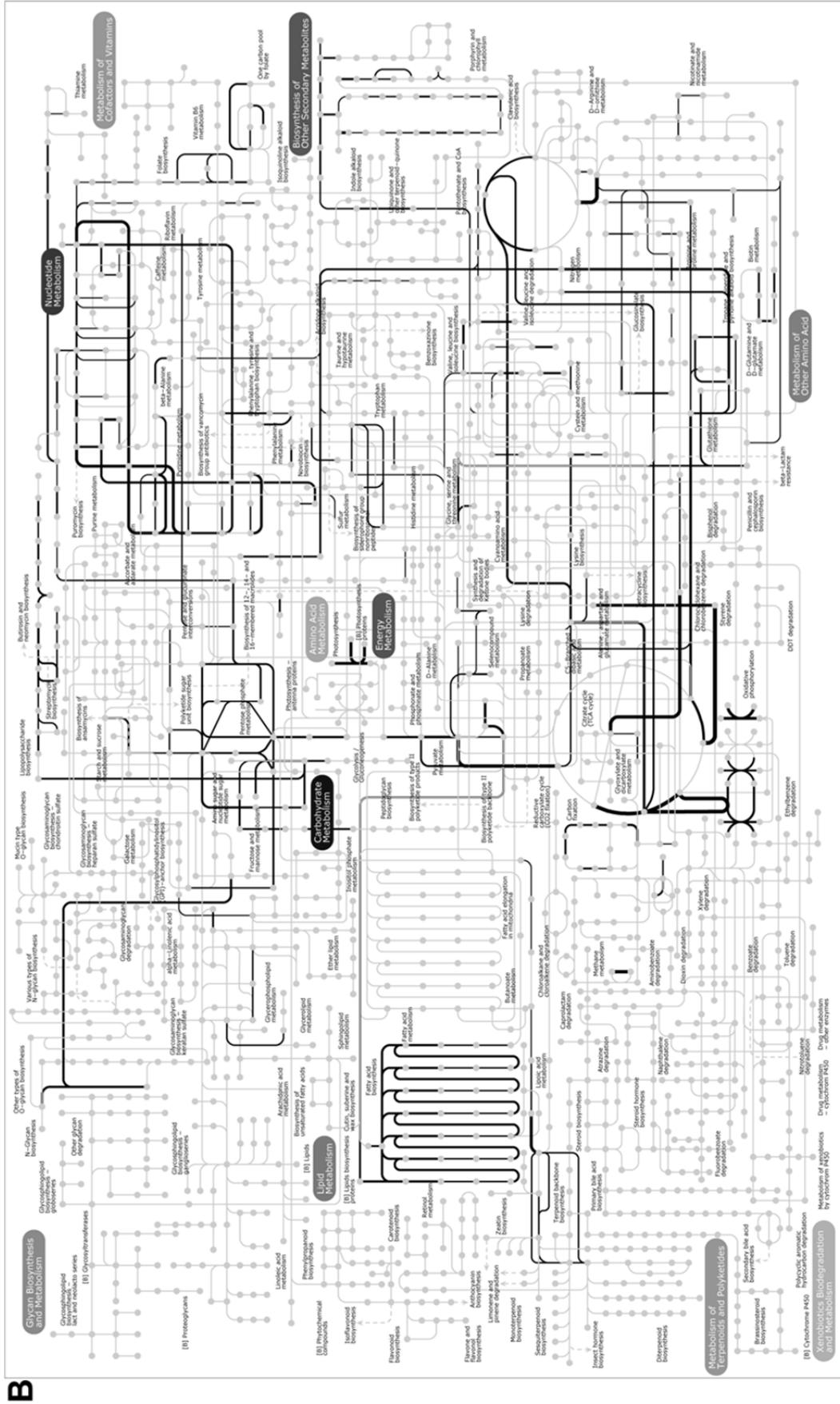


Figure 3.6 (Continued). Example of genus specific metabolic pathways generated from metatranscriptome dataset of the morning day one sample of subject 2. Metabolic pathways of *Streptococcus* (A), *Veillonella* (B), and *Clostridium* (C). The line width indicates gene expression level. Metabolic pathways were generated using iPath v2 based on KEGG annotations of expressed genes. A high resolution image is available upon request.

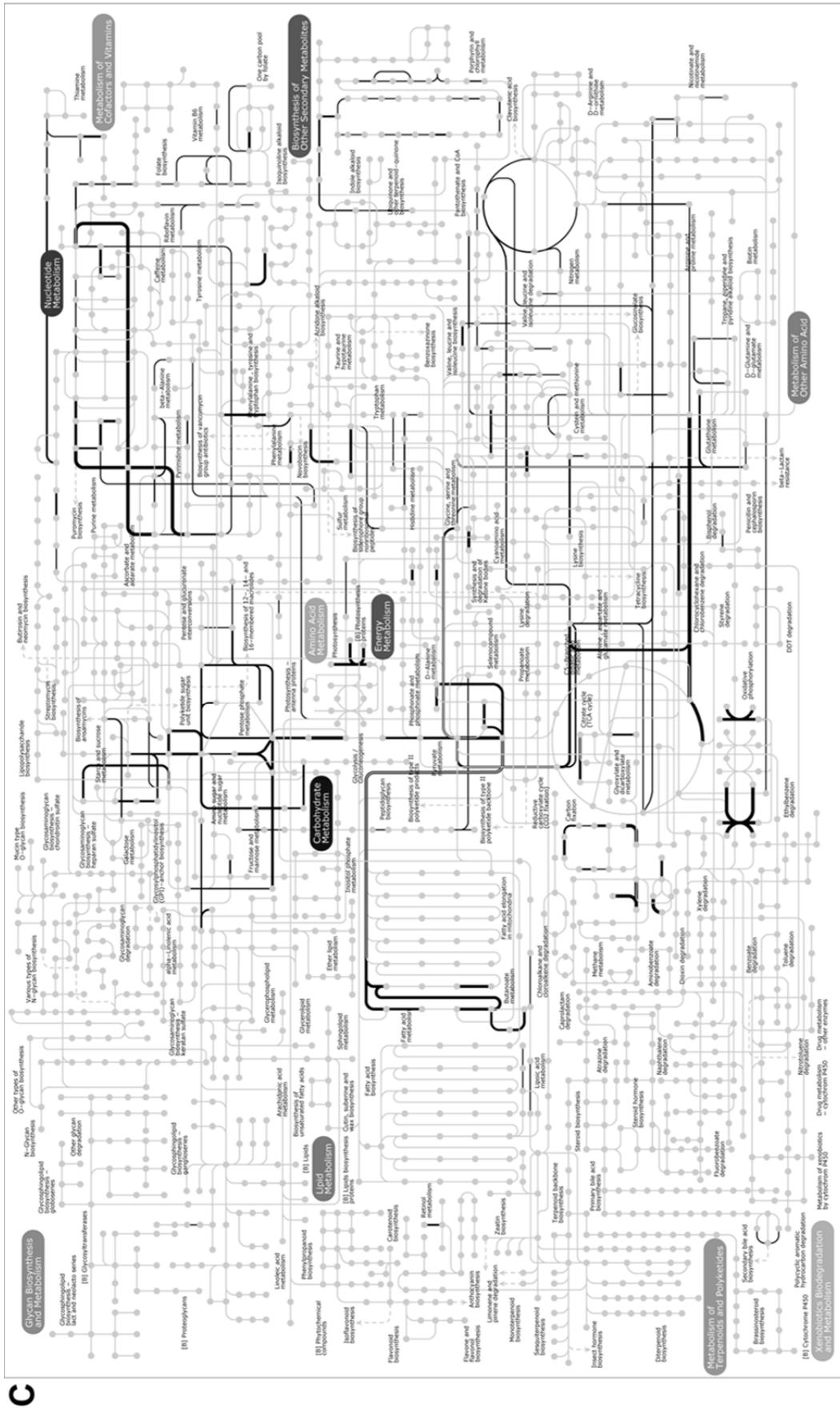


Figure 3.6 (Continued). Example of genus specific metabolic pathways generated from metatranscriptome dataset of the morning day one sample of subject 2. Metabolic pathways of *Streptococcus* (A), *Veillonella* (B), and *Clostridium* (C). The line width indicates gene expression level. Metabolic pathways were generated using iPath v2 based on KEGG annotations of expressed genes. A high resolution image is available upon request.

Discussion

This study determines the phylogenetic composition (rDNA) of the small intestine microbiota obtained in ileostoma samples, in combination with the determination of the overall activity of each of the bacterial ecosystem members (rRNA), as well as the specific activity of the overall ecosystem and its microbial members (metatranscriptomics). Moreover, by performing this in multiple individuals, and at multiple timepoints, both inter- and intra-personal dynamics of each of these ecosystem characteristics could be assessed. Last but not least, lysis and 16S rRNA (gene) PCR procedures were harmonized allowing accurate one to one comparisons at the different profiling levels.

Sequence analysis of the 16S rDNA and rRNA content, used as proxies for assessing total and active community composition, revealed a subject specific phylogenetic composition. Notably, relative abundances of specific bacterial groups of the ileostoma microbiota of a subject could fluctuate considerably over time, which is different from what was concluded with respect to the 'stable' microbiota in fecal samples (40, 46, 66). Moreover, temporal stability of the ileostoma ecosystem displayed substantial inter-individual variation, which was apparent from the distinct clustering of ileostoma microbial communities in morning and afternoon samples in some subjects but not in others.

The genera *Streptococcus*, *Veillonella*, and *Actinomyces*, co-occurred in the total community fractions of all samples and were detected in the majority of the active ileostoma community, albeit that *Actinomyces* abundance estimates were significantly lower in the active fraction compared to the total microbial communities. These findings support the notion that *Streptococcus* and *Veillonella* are typical and active inhabitants of the human small intestine (31, 340, 381). This observation as well as the notion that these microbial groups are also abundant in the upper GI tract (1, 3, 7, 43) supports their 'autochthonous' status in the small intestine microbiota (353). Moreover, the high expression of genes involved in fermentation of carbohydrates and short chain fatty acid metabolism (381) observed for these bacterial groups supports their prominent role in the overall functioning of the ecosystem.

The results described here are in good agreement with findings presented in a previous study by Booiijink and colleagues (31), that employed a phylogenetic microarray (the human intestinal tract chip (HITChip)) to investigate the composition and dynamics of the ileostoma effluent microbiota. Despite the principal differences between the HITChip and pyrosequencing methodologies (49, 339), the biological conclusions they generate appear to be similar with respect to the subject-specific nature, the temporal stability, and the predominant bacterial members encountered within the small intestinal microbial communities. Furthermore, taxonomic assignments of mRNA sequences at genus level comprised of the same dominant genera as were detected by 16S rRNA (gene) targeted pyrosequencing, demonstrating that independent of technology and targeted biomolecules the dominant community members are identified. However, the mRNA derived

phylogenetic profiles contained a considerable raised abundance of sequences that could not be classified at genus level, compared to the 16S rRNA (gene) based profiles. This confirms the high resolution taxonomic information encompassed in the 16S rRNA gene, especially when compared to sequences of protein coding genes that are much less discriminative. Furthermore, as the majority of the primary RNA sequence data contained rRNA reads, the depth of functional analysis can be improved by increasing the efficiency of rRNA removal.

Previously it has been postulated that ileostomy enhances oxygen penetration into the small intestine lumen, leading to disruption of the normal anaerobic environment (125). However, our analyses detected substantial abundances of typically anaerobic genera *Clostridium*, *Veillonella*, *Turicibacter*, and members of the Peptostreptococcaceae family, demonstrating that the small intestine of ileostomy subjects supports the growth and activity of strict anaerobes.

Although the phylogenetic composition of the total and active fraction of the small intestinal microbiota were highly similar, species richness and diversity were lower in the active population compared to the total microbiota. This finding implies that most phylogenetic groups represented in each ileostoma sample contribute to the overall function and that especially low abundance populations may not participate prominently in the active microbial community. However, this conclusion may be skewed by differences in ribosomes per cell that may differ between fast-growing and slow growing microorganisms (93, 113, 181), while also reduced amplification efficiency of lowly abundant template sequences may have contributed (109). Our observations contrasts with compositional differences between the active and total (257) as well as viable (active) and injured or dead (non-active) fractions (18) of fecal microbial communities, where the total microbiota did not represent the active players of the fecal ecosystems. Both microbial ecosystems in the small and the large intestine are subjected to environmental conditions (e.g. antimicrobial peptides, host immune responses, and pH (32, 75)) that influence bacterial survival and thus the outcome of compositional analysis of the total and active microbial fractions. Although these factors differ between the distinct regions of the intestinal tract, they do not explain the divergence between total and active fractions in the fecal microbiota versus the congruent total and active fraction composition in the ileostoma microbiota. Fast transit, high nutrient availability, and low bacterial load in the small intestine as compared to further down the GI tract, may be significant determinants in the observed differences. We anticipate that there is severe microbial competition for the remaining (complex) carbohydrates in the colon (351). Consequently, colonic and fecal bacteria may suffer more significant nutrient starvation stress, leading to cell death as compared to bacteria in the small intestine.

The microbial composition in ileostoma samples from the different subjects did not cluster consistently by sampling day and/or by morning versus afternoon, which indicates that factors other than day-night rhythms may determine fluctuations in this microbial ecosystem. Factors that can be considered here include variation in diet intake (amount and/or composition), which could have profound effects on the physico-chemical properties of the small intestinal habitat and in turn may modulate

temporal microbiota composition and activity. These considerations imply substantial opportunities to modulate the small intestinal microbiota by dietary interventions. Eventually this may enable the rational design of diets that aim to predictably modulate the relative abundance of particular phylogenetic groups of the endogenous microbiota. In this respect, it is of particular interest that individual *Streptococcus* lineages isolated from the small intestine display variable carbohydrate utilization capacity and have distinct immunomodulatory characteristics (Chapter 4, 5, and 6; (340)), which suggests that nutritional modulation of streptococcal population densities in the small intestinal lumen may affect local immune functions in the mucosa.

Subject-specific phylogenetic composition is also reflected in subject-specific activity patterns revealed by metatranscriptome analyses. However, activity profile variations were relatively small, which is in agreement with a high degree of functional redundancy among the microbiota members in different individuals. Metabolism was among the dominantly expressed functional domains, which corroborates earlier findings from our laboratory and underpins the role of the small intestinal microbiota plays in metabolic conversion (fermentation) of diet-derived carbohydrates (381). Though, these functional categories were also found to dominate the metatranscriptome of the fecal microbiota (30, 112), the metatranscriptome of the small intestinal microbiota is particularly enriched with PTS and other carbohydrate systems involved in metabolism of simple carbohydrates (381). Metatranscriptomics also enabled the (partial) assignment of specific metabolic functions to distinct phylogenetic groups (e.g. the genera *Streptococcus*, *Clostridium*, and *Veillonella*), illustrating that these data can enable the reconstruction of interactive metabolic relationships (e.g., syntrophy, cross-feeding, competition, etc.) between microbial groups that are residing in the ecosystem. This is clearly exemplified by the metatranscriptome support for the previously postulated food-chain relationship between *Streptococcus* and *Veillonella* (381) as well as members of the genus *Clostridium* that utilize acetate for formation of butyrate. Interestingly, small intestinal bacteria, mostly belonging to *Clostridium* cluster I, appeared to use phosphotransbutyrylase and butyrate kinase for the final step in butyrate formation, rather than butyryl-CoA transferase utilized by butyrate producers from the lower GI tract (77, 209).

In conclusion, this study demonstrates that 16S rRNA (gene) and metatranscriptome community profiling are complementary, and in the small intestinal ecosystem reveal that the phylogenetic composition of the microbiota as such and its activity distribution are highly congruent. Functional and phylogenetic interpretation of the metatranscriptome also allows the reconstruction of ecosystem interactions especially for dominating signature genera of the small intestine community, such as *Streptococcus* and *Veillonella*. The relatively low degree of temporal stability of the small intestine microbiota implies that external influences like dietary intake strongly impacts the composition and activity of the small intestinal microbiota. This notion, in combination with the prominent role of the small intestine in immunological perception of the environment, may offer opportunities to modulate local host-

immune-microbe interactions by specific dietary regimes, thereby aiming to improve health of the host organism.

Acknowledgements

This project was partially supported by the Netherlands Bioinformatics Centre (NBIC). The authors would like to thank Christopher Bauser and Julia Löcherbach of GATC-Biotech (Konstanz, Germany) for assistance in the set-up of the pyrosequencing experiments.

Supplementary information

Supplementary figures

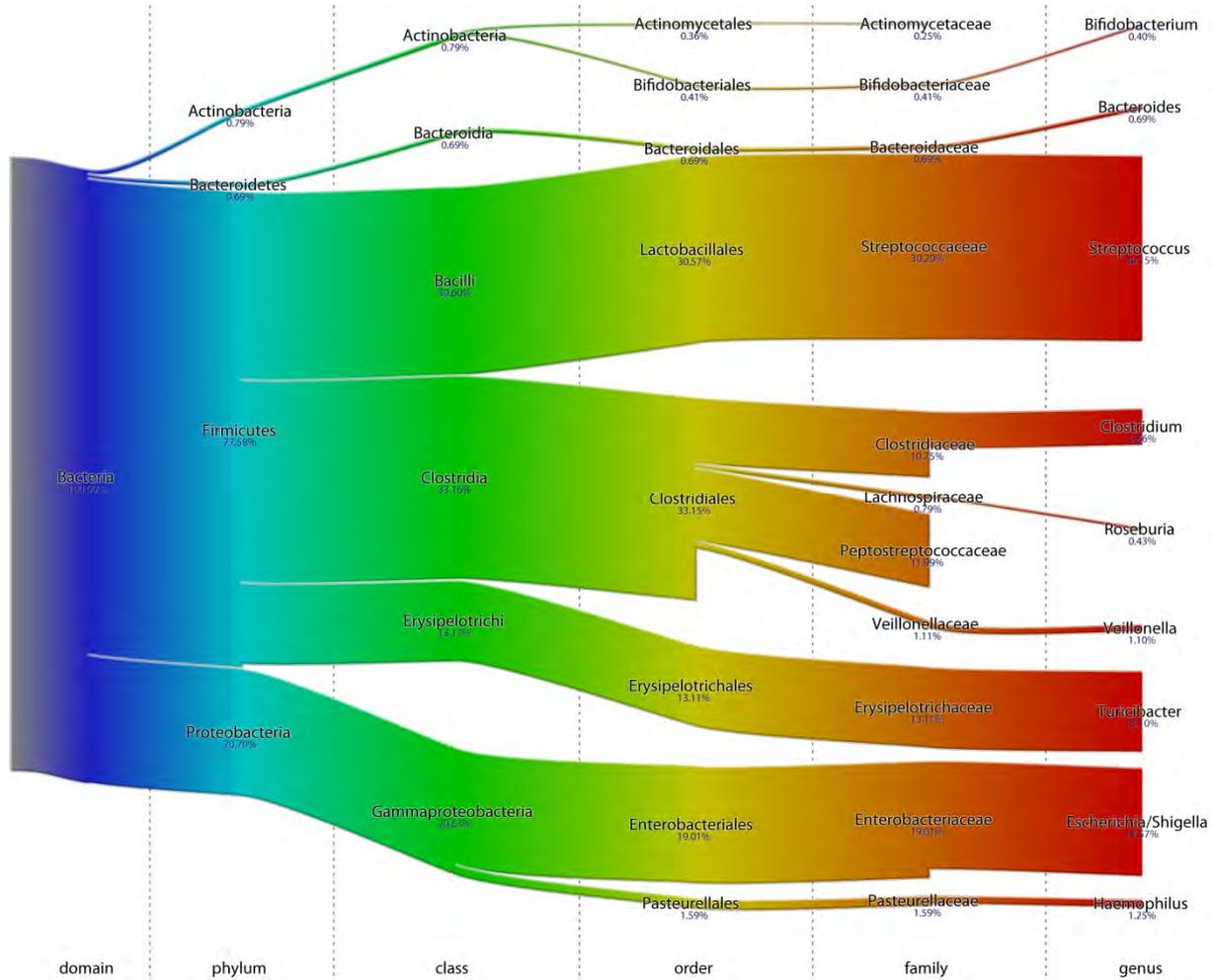


Figure S3.1. Relative contributions of detected bacterial taxa from domain to genus level through pyrosequencing of 16S ribosomal RNA gene (rDNA) and ribosomal RNA (rRNA) content from ileostoma effluent sample collected from four subjects at four time points. Figure generated using software described in Sundquist, et al (311), also providing a more detailed description of the visualization.

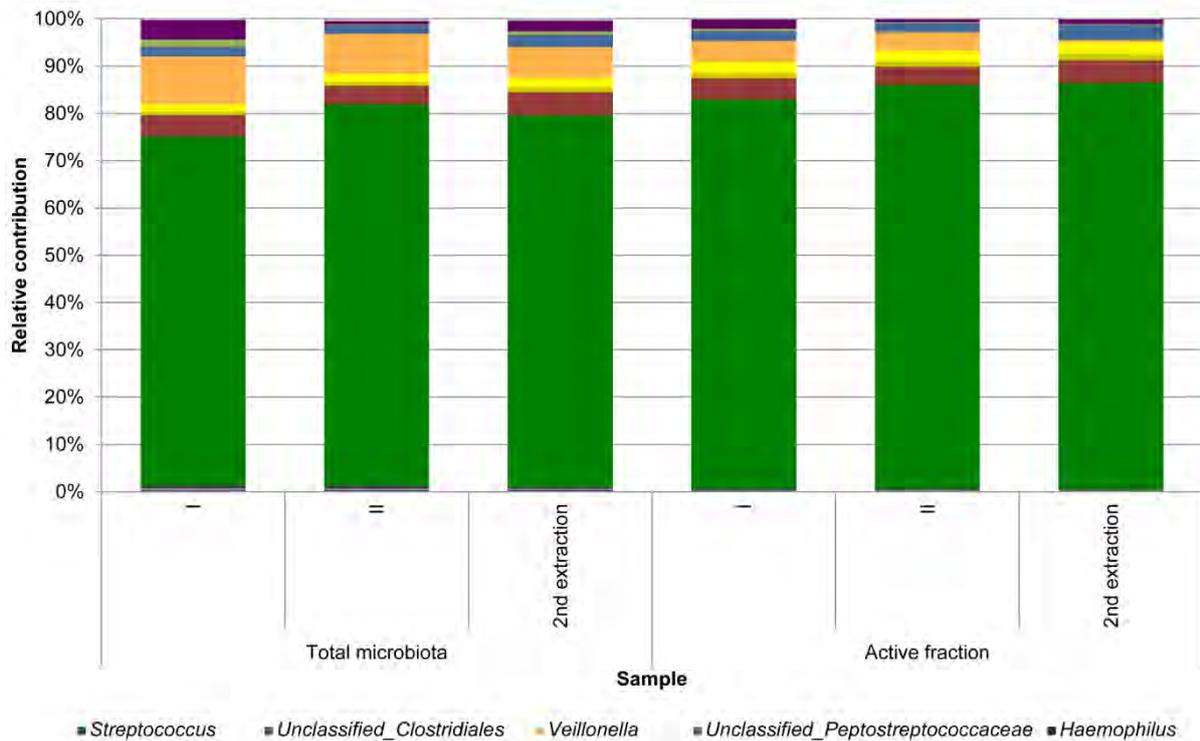


Figure S3.2. Relative contributions of detected bacterial taxa at genus level with pyrosequencing of the whole microbial community (16S rDNA) and the active fraction (16S rRNA) in the morning ileostoma sample collected on day 1 of ileostomy subject 2 for which the nucleic acids were amplified twice in separate reactions (I and II) and were extracted twice (2nd extraction). Phylogenetic groups that contribute at least 2.5% to one of the profiles are indicated in the color key. Pyrosequences that could not be classified above the confidence threshold of 80% are grouped to “Unclassified_” at the specific rank per taxon.

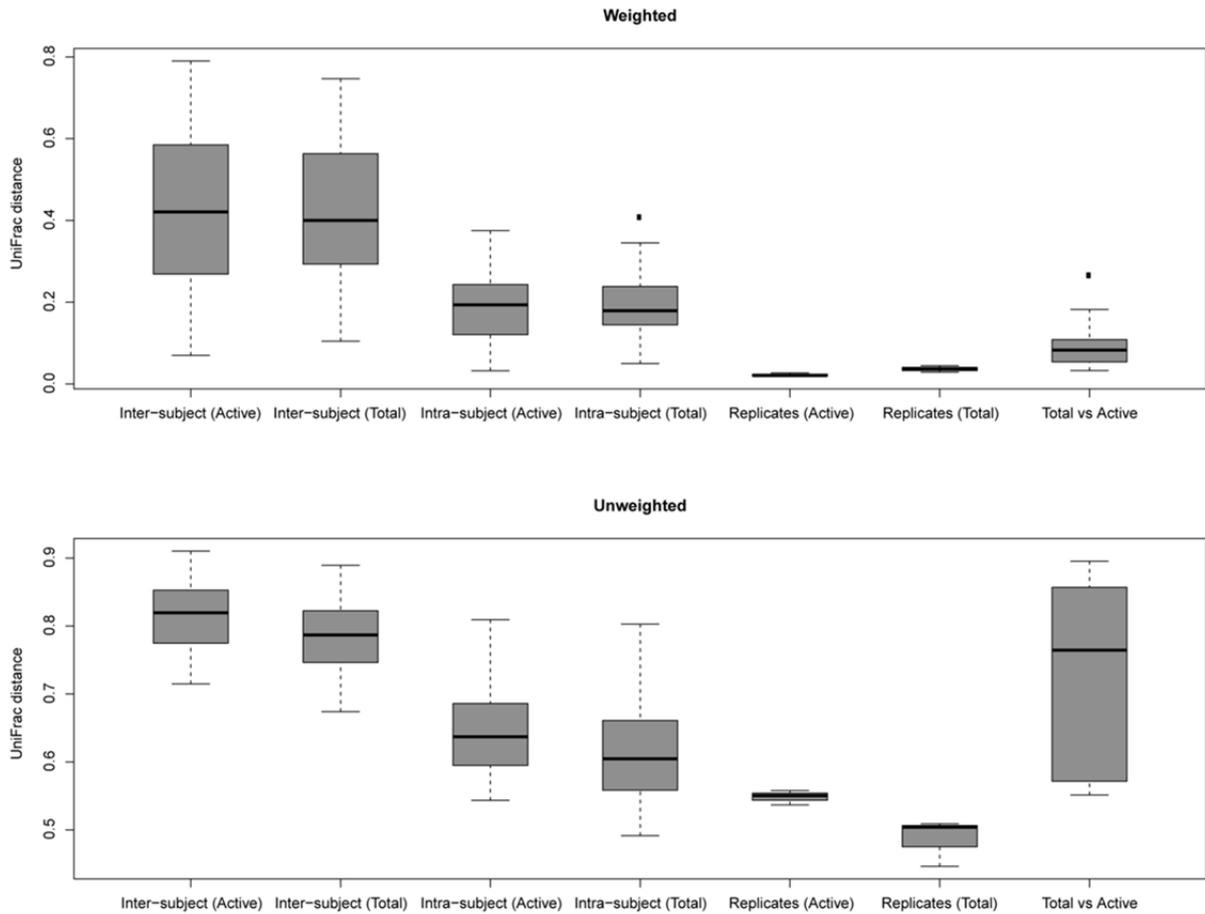


Figure S3.3. Boxplots based on weighted (upper panel) and unweighted (lower panel) UniFrac distance calculated between subjects, within subjects, between different sample preparations (replicates), and within pair-distance between total and active ileostoma microbiota as assessed by pyrosequencing of 16S rDNA and 16S rRNA, respectively. Lower UniFrac distances represent more similar microbial communities.

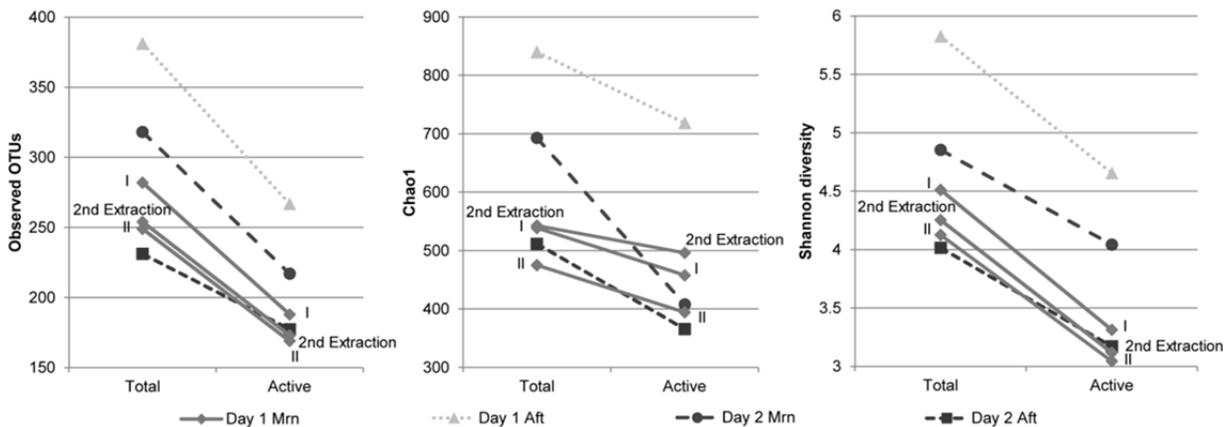


Figure S3.4. Number of OTUs, Chao1 richness estimations, and Shannon diversity from sequence analysis of the total and the active microbial fraction in the collected ileostoma samples from subject 2. Data analysis for sample preparations from the morning sample on day 1 for which the nucleic acids were amplified twice in separate reactions (I and II) and were extracted twice (I and 2nd extraction). Values in represented ecological metrics are based on random sub-samplings of 2836 reads per sample, average of four trials.

3

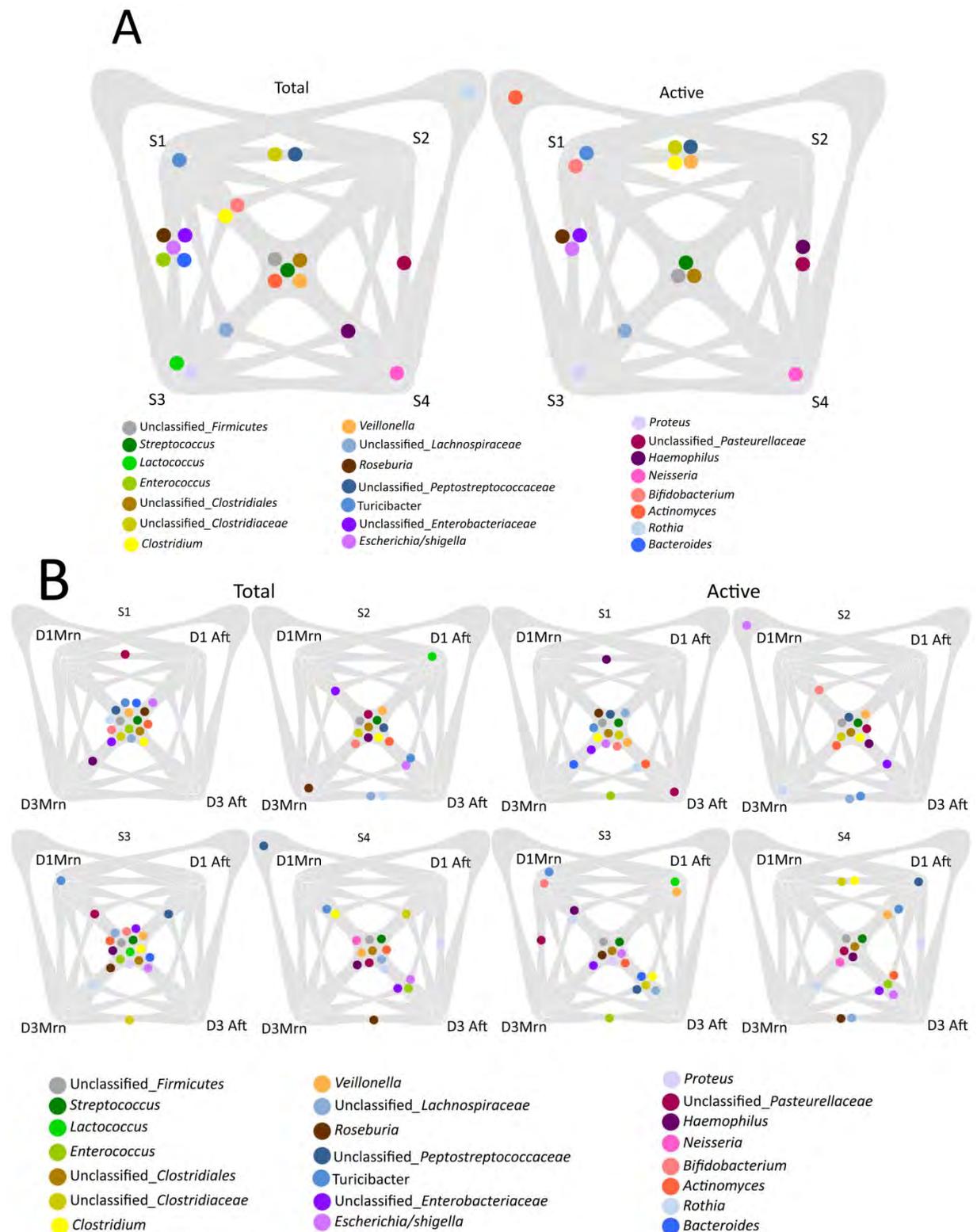


Figure S3.5. Venn diagram with distribution of genus-level taxa detected through pyrosequencing of the total microbial community (16S rDNA) and the active fraction (16S rRNA) that contribute at least 1% to one of the profiles (Figure 2) from ileostoma effluent samples collected from four subjects (A; each sphere indicates that the specific phylogenetic group was detected in all four samples from one or more subjects) and samples per individual (B). Pyrosequences that could not be classified above the confidence threshold of 80% are grouped to “Unclassified_” at the specific rank per taxon

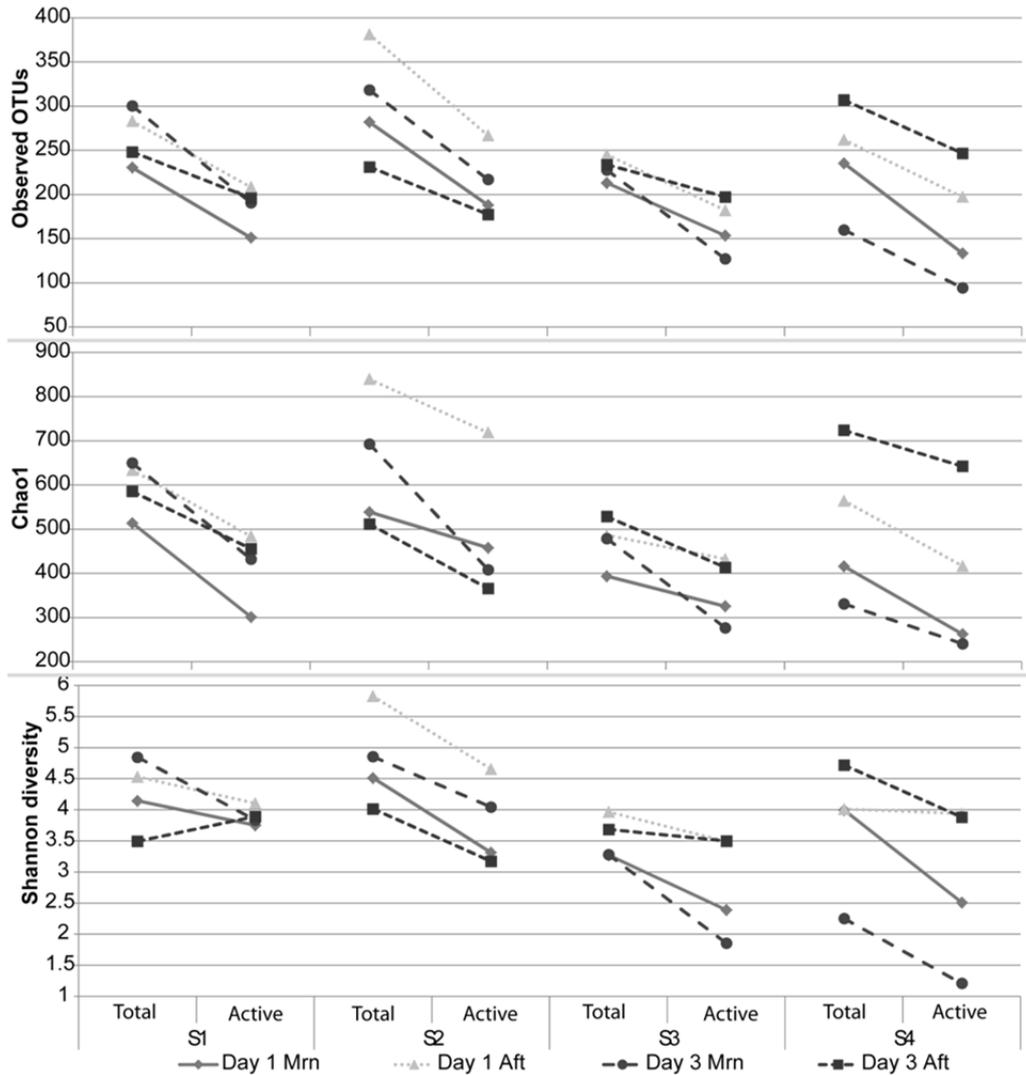


Figure S3.6. Number of OTUs, Chao1 richness estimations, and Shannon diversity from sequence analysis of the total and the active microbial fraction in afternoon (Mrn) and evening (Aft) ileostoma samples. Values in represented ecological metrics are based on random subsamplings of 2836 reads per sample, average of four trials.

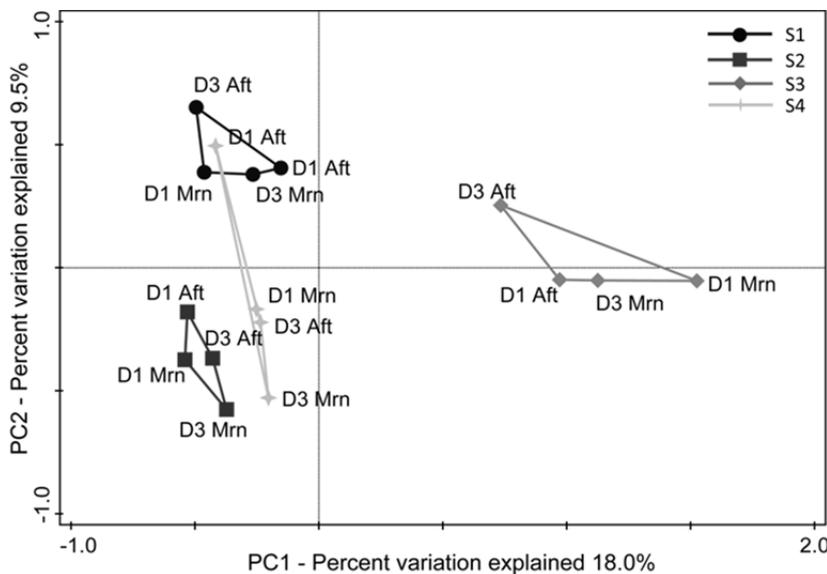


Figure S3.7. First two principal components (PC 1 and PC 2) from PCA analysis based on the COG distribution derived from all metatranscriptome datasets obtained from effluent samples obtained from four subjects (S1-S4).

Supplementary tables

Table S3.1. Number of detected taxa in the total and active fraction* of ileostoma effluent samples

16S rDNA/rRNA	Sample			Domain	Phylum	Class	Order	Family	Genus		
	Subject	Day	Timepoint								
16S rDNA (total)*	1	1	Mrn	1	4	6	10	21	21		
			Aft	1	5	8	12	20	20		
		2	1	Mrn	1	5	7	11	21	25	
				Aft	1	5	8	12	26	27	
	2	1	Mrn I**	1	4	6	8	12	13		
			Mrn II**	1	4	7	9	14	15		
			Mrn (2 nd extraction)**	1	5	6	9	14	17		
			Aft	1	4	6	10	14	18		
		2	1	Mrn	1	4	6	10	17	22	
				Aft	1	8	13	18	27	27	
			3	1	Mrn	1	4	6	10	18	26
					Aft	1	4	6	9	16	21
	4	1	Mrn	1	6	9	12	19	20		
			Aft	1	5	8	11	19	21		
		2	1	Mrn	1	5	9	13	24	27	
				Aft	1	7	10	13	24	26	
	16S rRNA (active)*	1	1	Mrn	1	4	6	8	11	10	
				Aft	1	4	6	9	15	15	
			2	1	Mrn	1	5	7	9	18	19
					Aft	1	5	7	10	19	19
2		1	Mrn I**	1	4	6	9	11	11		
			Mrn II**	1	4	6	7	10	12		
			Mrn (2 nd extraction)**	1	4	5	6	8	9		
			Aft	1	3	4	7	9	9		
		2	1	Mrn	1	4	6	9	14	13	
				Aft	1	3	5	6	10	9	
			3	1	Mrn	1	3	5	8	9	12
					Aft	1	4	5	6	11	15
4		1	Mrn	1	4	5	8	15	17		
			Aft	1	4	6	7	12	16		
		2	1	Mrn	1	4	7	7	7	7	
				Aft	1	4	7	10	13	14	
2		1	Mrn	1	4	6	8	13	15		
			Aft	1	4	7	9	17	21		

Mrn: Morning; Aft: Afternoon

*: The composition of the whole microbial community and the active fraction was analysed through pyrosequencing of 16S rDNA and rRNA, respectively.

** : DNA and RNA from the morning ileostoma sample collected on day 1 of ileostomy subject 2 were extracted twice (I and 2nd extraction) of which the nucleic acids from the first extraction were amplified twice in separate reactions (I and II).

Table S3.2. Diversity statistics for amplicon sequencing analysis in the total and active fraction of ileostoma effluent samples

		Sample		Timepoint		Reads	OTUs	Observed*	PD_whole_tree*	Fisher_alpha*	Equitability*	Singles*
16S rDNA/ rRNA	Subject	Day	Day	Timepoint	Timepoint							
16S rDNA (Total)	1	1	Mirn	17956	603	230.6	10.6147	59.3322	0.5281	137.6		
			Aft	9491	552	283.0	12.9697	78.2280	0.5562	172.2		
		3	Mirn	8118	537	300.4	13.5902	84.9290	0.5889	174.0		
	2	1	Aft	14397	601	248.2	11.7880	65.4961	0.4392	151.6		
			Mirn I**	5966	425	282.0	11.5804	77.8590	0.5544	157.2		
		Mirn II**	7178	417	249.0	10.5463	65.7519	0.5184	136.2			
	3	1	Mirn (2 nd extraction)**	7230	428	254.0	10.7002	67.5744	0.5325	145.0		
			Aft	4915	511	381.2	14.0151	118.5760	0.6796	225.0		
		3	Mirn	8274	555	318.2	12.8403	91.9596	0.5840	186.8		
	16S rRNA (Active)	1	1	Aft	5783	346	231.2	13.6660	59.5243	0.5113	133.4	
				Mirn	11267	432	213.2	10.3499	53.4334	0.4232	111.4	
			3	Aft	7009	403	245.0	12.0680	64.3530	0.4997	135.4	
2		1	Mirn	13361	557	227.8	12.1389	58.4813	0.4191	129.8		
			Aft	13493	553	233.8	11.9381	60.4377	0.4685	139.6		
		3	Mirn	11625	445	235.4	10.4003	61.0021	0.5073	118.8		
3		1	Aft	14333	597	262.0	11.6612	70.4368	0.4992	148.0		
			Mirn	10995	334	160.0	8.0293	36.7143	0.3079	88.4		
		3	Aft	5802	466	307.0	14.0234	87.5094	0.5712	180.2		
4		1	Mirn	6266	228	151.2	7.0318	34.1192	0.5185	81.2		
			Aft	16106	541	208.4	8.7081	51.8632	0.5328	123.0		
		3	Mirn	12452	442	190.6	9.2017	46.0861	0.5061	113.4		
16S rRNA (Active)	1	1	Aft	22922	599	196.0	8.5427	47.8174	0.5114	115.6		
			Mirn I**	5128	266	188.0	8.4035	45.2930	0.4388	114.0		
		Mirn II**	9378	353	169.0	7.0167	39.4096	0.4115	100.8			
	2	1	Mirn (2 nd extraction)**	10227	359	173.0	6.9787	40.6274	0.4197	109.2		
			Aft	2902	271	266.8	10.7964	72.1864	0.5776	169.4		
		3	Mirn	7120	367	217.0	9.3852	54.7114	0.5211	123.6		
	3	1	Aft	7899	322	177.4	6.8189	41.9665	0.4246	100.4		
			Mirn	8810	300	153.6	7.5296	34.8229	0.3291	89.0		
		3	Aft	11334	403	182.2	8.7049	43.4603	0.4650	107.2		
	4	1	Mirn	11057	277	127.2	7.3217	27.3522	0.2657	73.2		
			Aft	15010	464	197.2	9.2343	48.2071	0.4590	115.6		
		3	Mirn	4192	166	133.6	7.1723	29.1193	0.3556	77.2		
4	1	Aft	21040	560	197.4	8.6538	48.2708	0.5167	107.6			
		Mirn	13500	229	94.4	5.4861	18.8060	0.1844	60.8			
	3	Aft	6649	408	246.6	10.6320	64.8973	0.4887	150.6			

Mirn: Morning; Aft: Afternoon

*: Values in represented diversity metrics are based on random sub-samplings of 2836 reads per sample, average of four trials

**: DNA and RNA from the morning ileostoma sample collected on day 1 of ileostomy subject 2 were extracted twice (I and 2nd extraction) of which the nucleic acids from the first were amplified twice in separate reactions (I and II).

Table S3.2 (Continued). Diversity statistics for amplicon sequencing analysis in the total and active fraction of ileostoma effluent samples

		Sample		Timepoint		Diversity Statistics						
16S rDNA/ rRNA	Subject	Day	Timepoint	Chao1*	Simpson_e*	observed_species*	simpson*	doubles*	dominance*	shannon*		
16S rDNA (Total)	1	1	Mrn	513.7268	0.0052	230.6	0.8388	32.8	0.1612	4.1447		
		3	Aft	634.0824	0.0041	283.0	0.8718	41.0	0.1282	4.5303		
	2	1	Mrn	649.7462	0.0037	300.4	0.8926	45.6	0.1074	4.8464		
		3	Aft	585.5620	0.0059	248.2	0.6825	33.2	0.3175	3.4940		
	16S rRNA (Active)	1	1	Mrn I**	539.0734	0.0042	282.0	0.8390	47.2	0.1610	4.5126	
			3	Aft	475.2512	0.0050	249.0	0.7991	40.8	0.2009	4.1266	
2		1	Mrn (2 nd extraction)**	542.9420	0.0048	254.0	0.8150	36.4	0.1850	4.2536		
		3	Aft	839.8798	0.0028	381.2	0.9535	54.2	0.0465	5.8265		
3		1	Mrn	693.0098	0.0036	318.2	0.8852	47.6	0.1148	4.8553		
		3	Aft	511.7129	0.0057	231.2	0.7651	30.6	0.2349	4.0148		
16S rDNA (Total)	1	1	Mrn	393.8063	0.0078	213.2	0.6042	33.8	0.3958	3.2740		
		3	Aft	486.5281	0.0052	245.0	0.7818	37.6	0.2182	3.9654		
	2	1	Mrn	478.7322	0.0074	227.8	0.5971	34.0	0.4029	3.2802		
		3	Aft	528.7053	0.0060	233.8	0.7139	32.6	0.2861	3.6865		
	3	1	Mrn	416.0137	0.0053	235.4	0.7978	38.6	0.2022	3.9960		
		3	Aft	564.5927	0.0050	262.0	0.7693	36.2	0.2307	4.0099		
16S rRNA (Active)	1	1	Mrn	330.9900	0.0140	160.0	0.4489	22.0	0.5511	2.2545		
		3	Aft	723.9957	0.0040	307.0	0.8239	37.8	0.1761	4.7189		
	2	1	Mrn	301.4391	0.0080	151.2	0.8312	20.8	0.1688	3.7535		
		3	Aft	484.0682	0.0056	208.4	0.8556	26.8	0.1444	4.1039		
	3	1	Mrn	432.2259	0.0063	190.6	0.8300	25.8	0.1700	3.8333		
		3	Aft	455.8430	0.0061	196.0	0.8335	25.4	0.1665	3.8938		
16S rRNA (Active)	1	1	Mrn I**	457.9945	0.0076	188.0	0.7012	24.2	0.2988	3.3139		
		3	Aft	394.5232	0.0090	169.0	0.6568	22.0	0.3432	3.0448		
	2	1	Mrn (2 nd extraction)**	496.1779	0.0087	173.0	0.6686	19.0	0.3314	3.1195		
		3	Aft	718.7281	0.0042	266.8	0.8925	30.6	0.1075	4.6553		
	3	1	Mrn	408.2433	0.0054	217.0	0.8476	39.0	0.1524	4.0443		
		3	Aft	365.8168	0.0085	177.4	0.6609	26.2	0.3391	3.1721		
16S rDNA (Total)	1	1	Mrn	325.6055	0.0130	153.6	0.5023	24.4	0.4977	2.3899		
		3	Aft	432.3766	0.0070	182.2	0.7826	23.6	0.2174	3.4909		
	2	1	Mrn	276.6518	0.0200	127.2	0.3932	18.2	0.6068	1.8575		
		3	Aft	413.5286	0.0070	197.2	0.7211	30.0	0.2789	3.4990		
	3	1	Mrn	262.8225	0.0114	133.6	0.6596	22.4	0.3404	2.5106		
		3	Aft	416.6695	0.0060	197.4	0.8423	26.4	0.1577	3.9396		
16S rRNA (Active)	1	1	Mrn	240.7913	0.0432	94.4	0.2466	12.0	0.7534	1.2097		
		3	Aft	642.5723	0.0055	246.6	0.7329	27.6	0.2671	3.8831		

Mrn: Morning; Aft: Afternoon

*: Values in represented diversity metrics are based on random sub-samplings of 2836 reads per sample, average of four trials

**: DNA and RNA from the morning ileostoma sample collected on day 1 of ileostomy subject 2 were extracted twice (I and 2nd extraction) of which the nucleic acids from the first extraction were amplified twice in separate reactions (I and II).

Table S3.3. Number of sequence reads before and after rRNA removal

Subject	Sample		Reads entering the pipeline	rRNA reads	Putative mRNA reads
	Day	Moment			
1	1	Mrn	30,550,945	30,214,438	336,507
		Aft	19,387,391	18,886,114	501,277
	3	Mrn	25,470,937	25,163,975	306,962
		Aft	52,975,043	50,836,679	2,138,364
2	1	Mrn	29,208,566	24,912,430	4,296,136
		Aft	36,593,420	34,322,727	2,270,693
	3	Mrn	45,673,033	42,620,544	3,052,489
		Aft	23,663,706	22,015,045	1,648,661
3	1	Mrn	20,982,982	19,478,754	1,504,228
		Aft	28,135,249	26,571,132	1,564,117
	3	Mrn	71,783,677	66,981,885	4,801,792
		Aft	14,927,709	14,013,161	914,548
4	1	Mrn	42,006,183	36,075,240	5,930,943
		Aft	18,635,381	16,570,840	2,064,541
	3	Mrn	26,697,579	24,419,378	2,278,201
		Aft	32,506,944	28,290,511	4,216,433

Mrn: Morning; Aft: Afternoon



Chapter 4

Diversity of human small intestinal
Streptococcus and *Veillonella* populations

Bartholomeus van den Bogert, Oylum Erkus, Jos Boekhorst, Marcus de Goffau, Eddy J. Smid, Erwin G. Zoetendal, and Michiel Kleerebezem

FEMS Microbiol Ecol, 2013. 85: 376-388.

Abstract

Molecular and cultivation approaches were employed to study the phylogenetic richness and temporal dynamics of *Streptococcus* and *Veillonella* populations in the small intestine. Microbial profiling of human small intestinal samples collected from four ileostomy subjects at four time points displayed abundant populations of *Streptococcus* spp. most affiliated with *S. salivarius*, *S. thermophilus* and *S. parasanguinis*, as well as *Veillonella* spp. affiliated with *V. atypica*, *V. parvula*, *V. dispar*, and *V. rogosae*. Relative abundances varied per subject and time of sampling. *Streptococcus* and *Veillonella* isolates were cultured using selective media from ileostoma effluent samples collected at two time points from a single subject. The richness of the *Streptococcus* and *Veillonella* isolates were assessed at species and strain level by 16S rRNA gene sequencing and genetic fingerprinting, respectively. A total of 160 *Streptococcus* and 37 *Veillonella* isolates were obtained. Genetic fingerprinting differentiated 7 *Streptococcus* lineages from ileostoma effluent, illustrating the strain richness within this ecosystem. The *Veillonella* isolates were represented by a single phylotype. Our study demonstrated that the small intestinal *Streptococcus* populations displayed considerable changes over time at the genetic lineage level since only representative strains of a single *Streptococcus* lineage could be cultivated from ileostoma effluent at both time points.

Introduction

The human body is populated with complex microbial communities, which vary in composition between body sites (57). The microbiota in the human gastrointestinal (GI) tract, for example, has adapted to the different conditions in the specific GI habitats (for a recent review see (353)), and is impressive not only because of its very high population density, but also because of its high phylogenetic diversity (268, 341), and its extensive functional capabilities that complement the human genetic potential (262). The composition and dynamics of the bacterial community in the lower GI tract have been well described (31, 266, 317, 322, 329), whereas the upper GI tract microbiota in healthy humans is less well characterized as a consequence of sampling difficulties (32, 58, 203). Nevertheless, the different sections of the upper GI tract encompass distinct bacterial groups that interact with the host (195). Furthermore, the small intestine represents the first region where food components and the intestinal bacteria encounter each other, and it is also the region of the intestine that is predominantly involved in primary nutrient digestion and absorption (32, 203). Therefore, the small intestinal microbiota are expected to be of great importance to the host by playing a prominent role in the primary carbohydrate metabolism (381) and have an important influence on host physiology and health status (for a recent review see (58)) by, for example, immune-system modulation through luminal sampling and handling of bacteria (see (75) for a review).

Recently, high-throughput 16S ribosomal RNA (rRNA) gene profiling was employed to characterize the human small intestinal microbiota in samples obtained from healthy individuals using an orally introduced catheter as well as samples collected from ileostomy subjects (31, 125, 381). The latter group of individuals underwent surgical removal of their colon, and as a consequence luminal content of their terminal ileum is excreted from an abdominal stoma, and can repetitively be collected in a non-invasive manner. These studies revealed that the bacterial community in ileostoma effluent is also encountered in the small intestine of healthy subjects with *Streptococcus* and *Veillonella* spp. as predominant components. Interestingly, both bacterial populations are not only abundant in the small intestine but also in the microbiota of the stomach (23), esophagus (256), throat (9), and oral cavity (169).

The fact that members of *Streptococcus* and *Veillonella* are frequently co-occurring at these body sites may be partially attributable to their potential for metabolic interaction that has been shown to occur in the oral cavity (84) and postulated for the small intestine (381). *Streptococcus* spp. are involved in the fermentation of sugars, yielding lactic acid as their predominant fermentation end-product. In turn, *Veillonella* are renowned for their capacity to use lactic acid as a carbon and energy source (243).

Since no small intestinal *Streptococcus* and *Veillonella* isolates have to our knowledge been described, this study focuses on in-depth assessment of their phylogenetic richness and population dynamics as well as the streptococcal carbohydrate metabolic capacities through a combination of cultivation and molecular typing methodologies.

Materials and methods

Ethics statement

The study was approved by the University Hospital Maastricht Ethical Committee, and was conducted in full accordance with the principles of the 'Declaration of Helsinki' (52nd WMA General Assembly, Edinburgh, Scotland, October 2000). Volunteers were informed about the study orally and in writing. Moreover, the volunteers signed a written informed consent before participation.

Profiling small intestinal populations

In total, 16 ileostoma effluent samples were collected from four ileostomy subjects (Table 4.1) who were colectomized at least 5 years prior to testing, are clinically considered to be healthy, and have a normally functioning small intestine: patients did not report any complaints related to GI functioning for at least three years prior to testing and were not following any treatment for GI-related symptoms. The subjects donated four samples each, collected on two distinct time points of the day (morning and afternoon), on two separate days (at least two days apart). The subjects collected the ileostoma effluent in a clean, empty ileostoma bag, which was emptied in centrifuge bottles (Nalgene, Rochester, NY, USA) containing 100 ml RNA^{later}® (Ambion, Austin, TX, USA), immediately after the bulk of the effluent flowed into the bag. Samples were stored for 4-10 hours at room temperature, after which the samples were frozen by transferring the tubes to dry ice. Frozen samples were transported to the laboratory, where they were kept at -80°C until further analysis.

Table 4.1. Characteristics of subjects included in this study

Subject	Gender	Age
S1	Male	79
S2	Female	65
S3	Male	60
S4	Female	60

DNA was extracted using a method as described previously (377), with minor modifications. In short, 1 ml ileostoma effluent suspension was mixed with 4 ml PBS followed by centrifugation at 4600g at 4°C for 10 minutes. The cell pellet was re-suspended in 0.5 ml ice-cold TE buffer (Tris-HCl pH 7.6, EDTA pH 8.0), after which the mixture was transferred to a microfuge tube containing 0.18 gram macaloid suspension (377), 0.1 mm zirconium beads, and 50 µl 10% SDS (Invitrogen, Carlsbad, CA, USA). The solution was mixed with 500 µl acid phenol (Invitrogen), followed by three Fastprep (Bertin Technologies, Montigny le Bretonneux, France) treatments at 5.5 m/s for 45 s with cooling on ice for 90 s between treatments. The sample was centrifuged at 13,400g at 4°C for 15 minutes, after which the nucleic acids in the aqueous phase were purified by consecutive extraction with phenol:chloroform:isoamylalcohol (25:24:1) and chloroform:isoamylalcohol (24:1). Phases were separated by centrifugation (13,400g at 4°C, 5 minutes) using Phase

Lock Gel tubes (5 Prime, Hamburg, Germany). Subsequently, 300 µl of the aqueous phase was treated with 3 µl RNase A (10 mg/ml; Qiagen GmbH, Hilden, Germany) and incubated at 37°C for 15 minutes. Subsequent steps employed a modified version of the QIAamp DNA Stool Mini Kit protocol: 22,5 µl proteinase K (20 mg/ml; Ambion) and 300 µl buffer AL were added to the sample followed by incubation at 70°C for 10 minutes. After addition of 300 µl ethanol (VWR, Amsterdam, The Netherlands), the sample was transferred to a QIamp column and centrifuged (13,000g, 1 minute). DNA pellets were washed with AW1 and AW2 buffer according to manufacturer's instructions. Finally, the DNA was eluted with 30 µl Nuclease Free Water (Promega, Leiden, Netherlands).

For 16S rRNA gene based microbial composition profiling, barcoded amplicons from the V1-V2 region of 16S rRNA genes were generated by PCR using the 27F-DegS primer ((339); Chapter 2) that was appended with the titanium sequencing adaptor A and a 8 nt sample specific barcode (121) at the 5'-end, and an equimolar mix of two reverse primers (338R I and II (60); Table 4.2), that carried the titanium adaptor B at the 5'-end.

Table 4.2. Adaptors and primers used in this study

Primer ^a	Primer sequence (5'-3') ^b	Reference
Adaptor A	CCATCTCATCCCTGCGTGTCTCCGACTCAG	Provided by GATC-Biotech
Adaptor B	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG	
27F-DegS	GTTYGATYMTGGCTCAG	(339); Chapter 2
338R-I	GCWGCCTCCCGTAGGAGT	(60)
338R-II	GCWGCCACCCGTAGGTGT	
27F	GTTTGATCCTGGCTCAG	(192)
1492R-rev	CGGCTACCTTGTTACGAC	
357F	CTCCTACGGGAGGCAGCAG	
(GTG ₅)	GTGGTGGTGGTGGTG	

^a: Primer names may not correspond to original publication

^b: M = A or C; R = A or G; W = A or T; Y = C or T

PCRs were performed using a thermocycler GS0001 (Gene Technologies, Braintree, U.K.) in a total volume of 100 µl containing 1× HF buffer (Finnzymes, Vantaa, Finland), 2 µl PCR Grade Nucleotide Mix (Roche, Diagnostics GmbH, Mannheim, Germany), 2 U of Phusion® Hot Start II High-Fidelity DNA polymerase, 500 nM of a forward and the reverse primer mix (Biologio BV, Nijmegen, The Netherlands), and 0.2-0.4 ng/µl of template DNA. The amplification program consisted of an initial denaturation at 98°C for 30 s, 30 cycles of: denaturation at 98°C for 10 s, annealing at 56°C for 20 s and elongation at 72°C for 20 s, and a final extension at 72°C for 10 minutes. The size of the PCR products (~375 bp) was confirmed by gel electrophoresis using 5 µl of the amplification-reaction mixture on a 1% (w/v) agarose gel containing 1× SYBR® Safe (Invitrogen, Carlsbad, CA, USA). PCR products were purified with the High Pure Cleanup Micro Kit (Roche) using 10 µl Nuclease Free Water for elution, and quantified using a NanoDrop ND-1000 spectrophotometer

(Nano-Drop Technologies, Wilmington, DE). Purified PCR products were mixed in approximately equimolar amounts and electrophoresed on an agarose gel, followed by excision and purification using the DNA gel extraction kit (Millipore, Billerica, MA, USA). Purified amplicon pools were pyrosequenced using a Genome Sequencer FLX in combination with titanium chemistry (GATC-Biotech, Konstanz, Germany).

The pyrosequencing data analysis was carried out with a workflow employing the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (38) using settings as recommended in the QIIME 1.2 tutorial with the following exceptions: reads were filtered for chimeric sequences using Chimera Slayer (118); and OTU clustering was performed with an identity threshold of 97%, using parameters as recommended in the QIIME newsletter of December 17th 2010 (<http://qiime.wordpress.com/2010/12/17/new-default-parameters-for-uclust-otu-pickers/>). Additional data handling was done using in-house developed Python and Perl scripts. The Ribosomal Database Project (RDP) classifier version 2.2 (355) was used for taxonomic classifications up to the genus level. The most-likely species was determined by comparing sequences against the RDP reference set using NCBI BLAST (7), tentatively classifying an OTU as a specific species when the BLAST score of the OTU-reference sequence pair was higher than the lowest-scoring reference sequence-reference sequence pair for that species in the RDP set. Results of the tentative assignments were evaluated by generating maximum likelihood phylogenetic trees containing a representative sequence from each OTU and reference sequences from the RDP database (53). Multiple sequence alignments were done with Muscle (81), the part of the alignment matching the amplicon was retrieved using JalView (357) and maximum-likelihood phylogenetic trees were generated using Phylml (116).

Sample collection for cultivation

An ileostoma effluent sample was obtained in the evening from subject 1 (Table 4.1; t = 0) by transferring approximately 20 ml ileostoma effluent from the ileostoma bag to a 50 ml tube containing 20 ml Phosphate Buffered Saline (PBS)-cysteine solution (8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l Na₂HPO₄, 0.24 g/l KH₂PO₄, [Sigma, St. Louis, MO, USA] and 1 g/l cysteine-HCl [Sigma], pH 6.8). The PBS-cysteine solution was flushed with N₂ to enhance survival of anaerobic bacteria such as *Veillonella*. Since large fluctuations on phylotype and function level were observed over a time span of one year (381), a second ileostoma effluent sample was collected in the morning from the same ileostomy subject (t = 1) one year after the first sampling to determine the population dynamics at the genetic lineage level.

All samples were placed in sealed plastic bags with an anoxic atmosphere, generated by an anaerocult® A mini sachet (Merck, Darmstadt, Germany) and stored at home in the refrigerator. Samples were transported to the laboratory and processed within 24 hours after collection.

Cultivation

Of all samples serial dilutions were prepared in PBS-cysteine solution (10^1 - 10^5), and plated on *Mitis Salivarius* (MS) agar (Becton Dickinson, Breda, the Netherlands) supplemented with Tellurite solution 1% (Becton Dickinson) according to manufacturer's instructions, and *Veillonella* selective agar (VSA) (274, 275) to facilitate selective isolation of *Streptococcus* and *Veillonella*, respectively. MS agar plates were incubated aerobically, whereas VSA plates were incubated in anaerobic jars with an anaerobic atmosphere generated by an anaerocult® A sachet (Merck). All plates were incubated at 37°C for 18-48 hours. Emerging colonies were randomly picked and grown in liquid media at 37°C for 18-48 hours.

The MS isolates were grown in MS medium, based on the MS agar, but lacking trypan blue, crystal violet, and agar. The VSA isolates were cultivated under anoxic atmosphere, generated by the anaerocult® A mini system in *Veillonella* medium described in the DSMZ catalogue (Medium 136), which contains lactate as its main carbon source for growth. Bacterial isolates were stored at -80°C in these same media to which 15% glycerol was added.

16S rRNA gene sequencing and analysis

Near-full length 16S rRNA gene fragments from the bacterial isolates were PCR amplified using a PCR protocol described previously ((339); Chapter 2) with the 27F and Uni-1492-rev primers (Table 4.2) and either a single colony or 2.5 µl of bacterial suspension/glycerol stock as a template source. Amplicon size was verified by electrophoresis on a 1% (w/v) agarose gel. PCR products were purified and subsequently sequenced from the 27F, 357F, and Uni-1492-rev (192) priming sites (GATC-Biotech, Konstanz, Germany; Table 4.2). The obtained sequence reads per amplicon were assembled using Clone Manager 9 Professional Edition (Scientific & Educational Software, Cary, NC, USA), yielding near full-length 16S rRNA gene sequences, which were taxonomically classified using a locally installed version of the RDP classifier version 2.2 (355) with a default confidence threshold of 80%.

The 16S rRNA gene sequences were aligned using the SILVA Webaligner (261), subsequently imported into ARB (215), and merged with the SILVA reference database release 106. A neighbor joining distance matrix employing no correction was calculated using ARB to group sequences into distinct phylotypes based on a threshold of 97% sequence identity.

The near full-length 16S rRNA gene sequences were deposited in the Genbank database and are available under accession numbers JQ680047 to JQ680145 and JQ680199 to JQ680348.

Typing of bacterial isolates

Bacterial isolates were classified into genetic lineages using amplified fragment length polymorphism (AFLP) and Rep-PCR genetic fingerprinting. AFLP was performed as described previously (189). Based on analysis of replicates with the AFLP protocol, 90% of similarity was used as threshold for the separation of individual genetic lineages (data not shown).

Rep-PCR fingerprinting analyses were performed using a thermocycler GS0001 with an amplification program described by Matsheka, et al. (227). Each reaction was performed in a total volume of 25 μ l composed of 1 \times PCR buffer (Promega), 1 μ M of the (GTG)₅ primer (Biogio BV, Nijmegen, The Netherlands; Table 4.2), 200 μ M of each deoxyribonucleotide triphosphate (Roche, Diagnostics GmbH, Mannheim, Germany), 1.25 U GoTaq[®] DNA polymerase (Promega) and 2.5 μ l of glycerol stock. Rep-PCR products were separated by electrophoresis on a 1.5% (w/v) agarose gel and stained by in gel 1 \times SYBR[®] Safe DNA-stain. After standardized electrophoresis (75 minutes, 100 V, 1 \times TAE), banding patterns were visualized under UV light and digitally captured using the Gel Doc XR System (Biorad) with Quantity One software version 4.6.6 build 102. Comparative analysis of the resulting fingerprints was performed using the BioNumerics suite (version 4.6.1; Applied Maths, St Martens Latem, Belgium). Similarities among profiles were calculated using the Pearson correlation coefficient, and cluster analyses were performed applying the unweighted-pair group method with arithmetic averages algorithm (UPGMA) with an optimization of 0.69%. Based on analysis of replicates with the Rep-PCR protocol, 84% of similarity was used as threshold for the separation of individual genetic lineages (data not shown). Rep-PCR based groupings for which no near full-length 16S rRNA gene sequences were obtained, or that were inconsistent with 16S rRNA gene classification results (see above), were excluded from further analysis.

The validity of using glycerol stocks as template source for Rep-PCRs was verified by comparing the resulting profiles with those obtained from Rep-PCRs using isolated genomic DNA of 3 randomly selected *Streptococcus* isolates, yielding identical profiles (data not shown).

Substrate conversion capacity of individual *Streptococcus* isolates was evaluated using API 50 CH strips in combination with API CHL medium (Biomérieux, Marcy l'Etoile, France). To this end, isolates were grown overnight in MS medium and washed twice with 0.9% NaCl prior to inoculation of the strips, incubation at 37°C, and assessment of the reactions after 24 and 48 hours according to manufacturer's instructions. To determine if fermentation patterns were consistent for isolates belonging to the same genetic lineage, two isolates from each genetic lineage (if available) were tested and compared.

Growth of the isolates on different carbon sources was assessed using carbohydrate-free MS-medium (MS^{Basal}), supplemented with L-arabinose, D-(+)-glucose monohydrate, D-mannitol, N-acetylglucosamine, sucrose, D-trehalose, D-raffinose, soluble starch, or glycogen [Sigma] at a standard concentration of 1% (w/v). Bacteria suspended in 0.9% NaCl (see above) were diluted to an OD₆₀₀ of 0.002 in the different MS-derived media, and incubated at 37°C for 18 hours. Growth was assessed by OD₆₀₀ determination.

MALDI-TOF MS analysis was performed with a Microflex mass spectrometer (Bruker Daltonics, Bremen, Germany) using FlexControl software (version 3.0). Spectra were recorded in the positive linear mode (laser frequency, 20 Hz; ion source 1, voltage at 20 kV; ion source 2, voltage at 18.4 kV; lens voltage, 9.1 kV; mass range, 2000–20 000 Da). Spectra were internally calibrated using *Escherichia coli* ribosomal

proteins. The spectra were imported into the integrated Biotyper software (version 2.0) and analyzed by standard pattern matching with default settings. The spectrum of each isolate was compared with those in the database containing 9 *Veillonella* and 61 *Streptococcus* spp. (see supplementary Materials and methods). Identification was provided with an accompanying score (log score 0–3) of reliability. This score is based on (i) matching of the spectrum in general, (ii) matching of the locus of the peaks and (iii) matching of the height of the peaks. Scores <1.7 represent no reliable identification. A score ≥ 1.7 and <2.0 is considered identification at the genus level, scores ≥ 2.0 identification at the species level (347).

Colonies of each isolate were directly spotted on the MALDI-plate and were overlaid with 1 μ l of matrix solution (α -cyano-4-hydroxy-cinnamic acid in 50% acetonitrile and 2.5% trifluoroacetic acid) and air-dried. Measurements were performed as described previously (345). If an identification score was below a 1.7 cutoff, the isolate was again spotted on a MALDI-plate and pre-treated with 1 μ l of 70% formic acid before being overlaid with the matrix solution. The highest of all the scores per isolate was considered the final result.

Results

Multiple *Streptococcus* and *Veillonella* spp. consistently co-occur in ileostoma effluent

To establish the *Streptococcus* and *Veillonella* spp. relative abundances in ileostoma effluent, we characterized the microbial communities collected from four ileostomy subjects at four different time points through 16S rRNA gene pyrosequencing. In total 162,785 quality filtered sequences with 10,174 (\pm standard deviation of 3,855) sequences per sample were obtained. From all sequences, 39,812 and 1,986 were assigned to the genera *Streptococcus* and *Veillonella*, respectively. Both genera were detected in all samples, with *Streptococcus* relative abundances ranging from 0.4% to 88.3% (Figure 4.1A). *Veillonella* was generally present at lower relative abundance with variable abundances per sample ranging from <0.1% to 10.1% (Figure 4.1A). Interestingly, out of the different *Streptococcus* phylotypes detected in the whole dataset, phylotypes most closely related to the species *S. salivarius*, *S. thermophilus* (*S. salivarius* species-group), and *S. parasanguinis* (*S. mitis* species-group) were of high relative abundance in all samples (Figure 4.1B and S4.1A). The *Veillonella* population in ileostoma effluent was represented by phylotypes, that most closely resembled the species *V. dispar*, *V. parvula*, *V. rogosae*, and *V. atypica*. Since discrimination of these species is not possible based on partial 16S rRNA gene sequences generated by pyrosequencing (Figure S4.1B), relative abundances for all phylotypes were combined (Figure 4.1A). The clustering in Figure S4.1B does show that we can specify that most of the OTUs detected in the small intestine are closely related to these four species and not to other *Veillonella* spp. (e.g. *V. ratti* and *V. criceti*). These findings demonstrate that the same *Streptococcus* and *Veillonella* spp. are detected in ileostoma effluent samples collected at different time points as well as from different subjects, indicating that these species are typical small intestinal

commensals. To further characterize the small intestinal *Streptococcus* and *Veillonella* population richness as well as its temporal dynamics at the strain-level, we applied an integrated approach that combined cultivation and physiological characterization (substrate utilization assays) with molecular typing, including genetic fingerprinting and MALDI-TOF MS-typing of bacterial isolates of these bacterial genera from ileostoma effluent of one of the subjects.

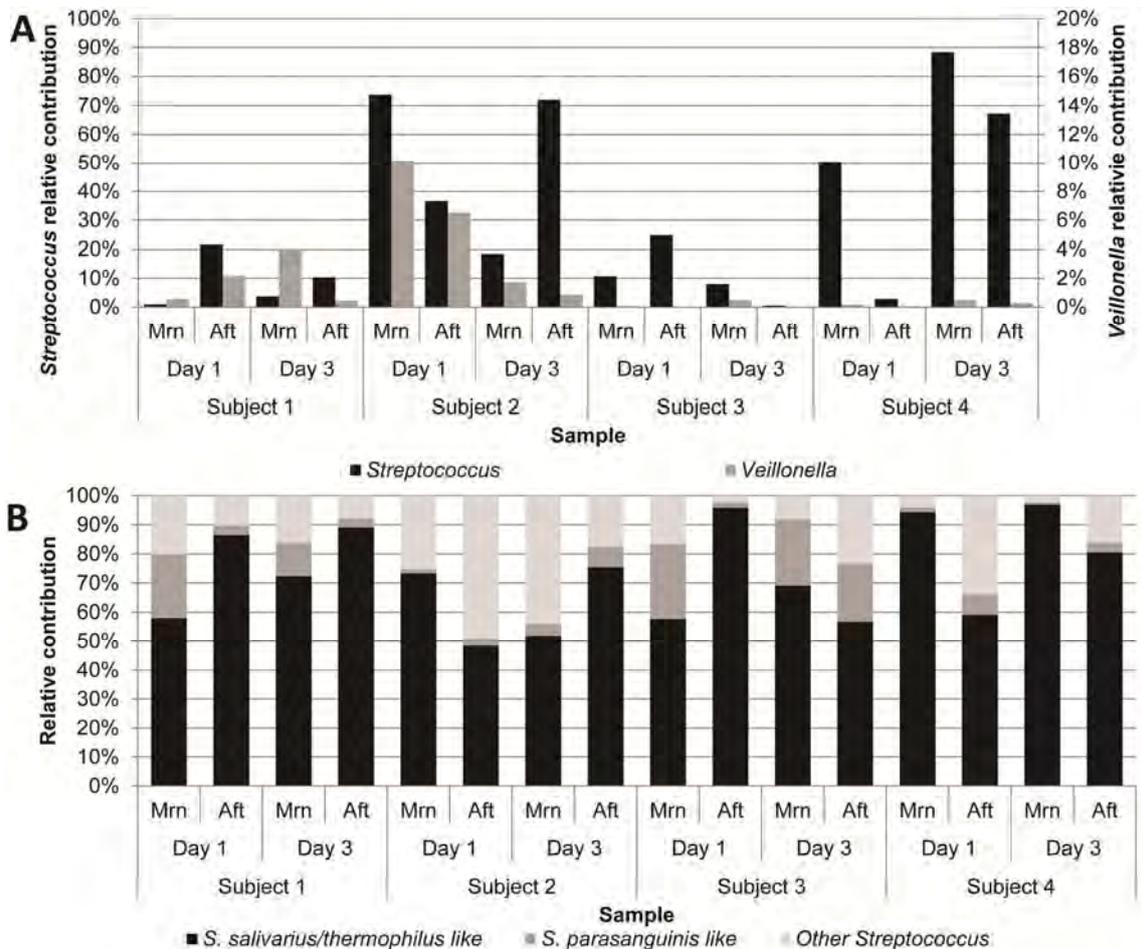


Figure 4.1. Relative contribution of *Streptococcus* (primary axis) and *Veillonella* (secondary axis; A) and *Streptococcus* spp. (B) as detected with pyrosequencing in morning (Mrn) and afternoon (Aft) ileostoma samples. Since discrimination of *S. salivarius* and *S. thermophilus* was not possible based on partial 16S rRNA pyrosequencing, relative abundances for phylotypes assigned to these species were combined to “*S. salivarius/thermophilus* like”.

Phylotype diversity of ileostoma effluent-derived bacterial isolates

A total of 272 bacterial isolates collected from two ileostoma effluent samples at 2 time points ($t = 0$ and, one year later, at $t = 1$; Table 4.3) were classified on basis of a combination of 16S rRNA gene sequencing and genetic fingerprinting by AFLP and Rep-PCR. These isolates were classified as the genera *Streptococcus* (160), *Enterococcus* (66), *Veillonella* (37), *Bacteroides* (5), and *Lactobacillus* (4) (Table 4.3). While *Streptococcus* was exclusively isolated using MS agar and *Veillonella*, *Lactobacillus*, and *Bacteroides* were only isolated from VSA, *Enterococcus* isolates were recovered from both media (Table 4.3). In the remainder of the paper we focus

on analysis of the *Streptococcus* and *Veillonella* isolates as typical commensal inhabitants of the small intestine (31, 381). Characteristics of the isolates are included in Table 4.3 and S4.2.

Table 4.3. Number of isolates per phylotype, lineage, and sample origin based on identification employing 16S rRNA gene classification, AFLP analysis and (GTG)5-PCR fingerprinting^a

16S rRNA gene classification (genus)	Phylotype	Lineage	Species group identification	Total isolates ^b	Isolates per ileostoma effluent sample	
					t = 0	t = 1
<i>Streptococcus</i>	1 ^c	1	<i>S. mitis</i>	3	3	
	2 ^c	2	<i>S. bovis</i>	17	13	4
		3		64		64
	3c	4	<i>S. salivarius</i>	1 ^f	1	
		5		1	1	
		6		8	8	
		7		66	66	
<i>Veillonella</i>	4 ^d	N.D. ^e	<i>V. parvula</i>	37	21	16
		8	<i>E. avium</i>	1	1	
<i>Enterococcus</i>		9	<i>E. faecium</i>	1		1
	5 ^c	10	<i>E. gallinarum</i>	5	4	1
		11	<i>E. avium</i>	6		6
		12	<i>E. avium</i>	10	1	9
		13	<i>E. faecium</i>	11	2	9
	5 ^d	N.D. ^e	<i>E. gallinarum</i>	11	11	
	6 ^c	14	<i>E. faecalis</i>	21	20	1
<i>Bacteroides</i>	7 ^d	N.D. ^e	<i>B. fragilis</i>	5		5
<i>Lactobacillus</i>	8 ^d	N.D. ^e	<i>L. fermentum</i>	4		4

^a: Characteristics of the isolates are included in table S2.; A graphic representation of the data included in this table and table S2 is added as figure S4.3

^b: All isolates were obtained from a 79 year old male ileostomist (subject 1; Table 1)

^c: Isolated obtained from MS agar

^d: Isolates obtained from VSA

^e: N.D. Not determined because AFLP analysis and/or Rep-PCR genomic fingerprinting did not reveal discriminative lineages

^f: AFLP analysis unlike Rep-PCR genomic fingerprinting identified a single *Streptococcus* isolate as a separate genetic lineage

Based on the near full-length 16S rRNA gene sequences, isolates identified as *Streptococcus* were divided in 3 phylotypes and *Veillonella* was represented by a single phylotype at a sequence identity threshold of 97%. The 16S rRNA gene sequences within each *Streptococcus* phylotype showed >99% sequence identity. The least abundant phylotype, consisting of three isolates, showed highest similarity to species in the SILVA database belonging to *S. parasanguinis* (>99%; *S. mitis* species group), while the second phylotype, consisting of 81 isolates showed highest similarity to *S. equinus* (>98.5%) and *S. lutetiensis* (>99.7%; *S. bovis* species group). The last phylotype, represented by 76 isolates showed highest similarity to *S. salivarius* subsp. *salivarius* (>98.7%) and *S. vestibularis* (>99.3%; *S. salivarius* group; Figure 4.2). These identifications were confirmed by MALDI-TOF MS analysis of

randomly picked isolates from each of the *Streptococcus* phylotypes (data not shown). Considering the high sequence similarity with multiple species for isolates from one phylotype, in the remainder of the paper the phylotypes are indicated with the *Streptococcus* species group names (see above) as they are also used in the Bergey's Manual of Systematic Bacteriology (65).

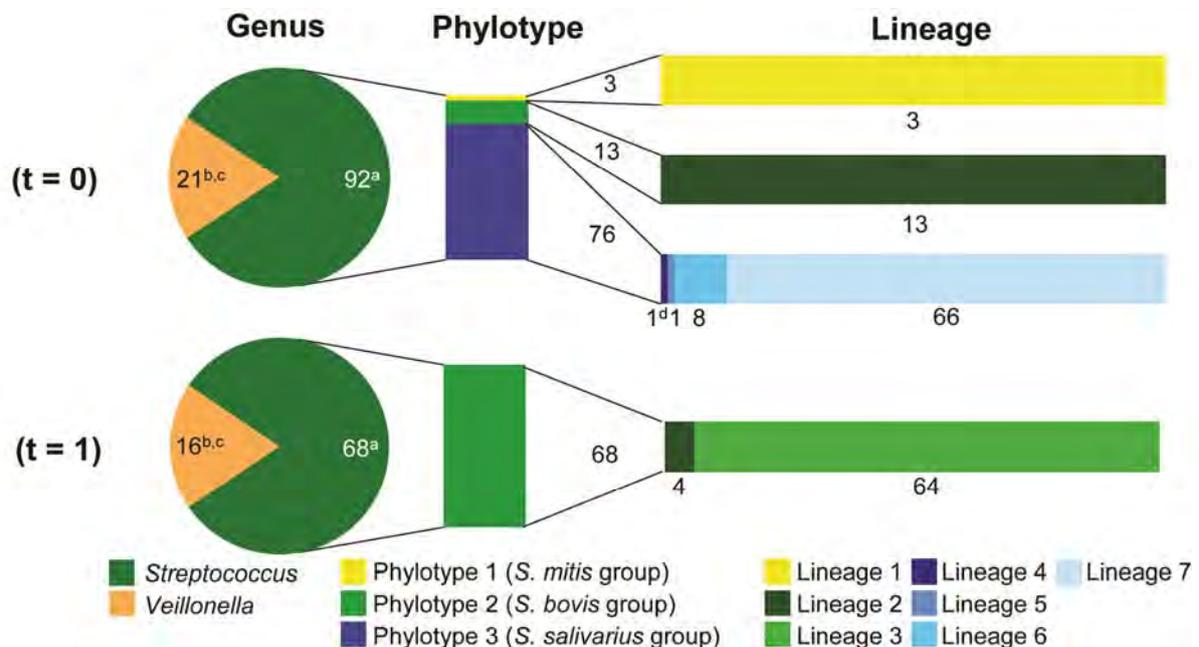


Figure 4.2. Relative contribution of *Streptococcus* and *Veillonella* isolates obtained from ileostoma effluent (t = 0 and 1) on genus, phylotype and lineage level. The bar plots next to the pie charts represent the division of the *Streptococcus* isolates in phylotypes and genetic lineages. The numbering of the *Streptococcus* phylotypes and lineages are based on the groupings in table 3. Characteristics of the isolates are included in table 3 and S2.; A graphic representation of the complete data is included as figure S4.3.

^a: Isolates obtained from MS agar

^b: Isolates obtained from VSA

^c: *Veillonella* lineage groupings are not determined because AFLP analysis and/or Rep-PCR genomic fingerprinting did not reveal discriminative lineages

^d: AFLP analysis unlike Rep-PCR genomic fingerprinting identified a single *Streptococcus* isolate as a separate genetic lineage. According to Rep-PCR genomic fingerprinting this isolate belongs to *Streptococcus* lineage 7.

Strain diversity of ileostoma *Streptococcus* populations exceeds the phylotype level

AFLP analysis and Rep-PCR genetic fingerprinting were employed to discriminate different bacterial lineages within a 16S rRNA phylotype group. Both techniques generated consistent results in terms of subtyping of the 152 bacterial isolates obtained from ileostoma effluent at t = 0 (Figure S4.2, Table S4.2). It should be noted, however, that AFLP fingerprinting discriminated isolates within the *S. salivarius* species group into two distinct lineages (lineage 4 and 7), while these isolates were grouped together by Rep-PCR genetic fingerprinting (Figure S4.2), illustrating the higher resolution of AFLP as a fingerprinting technique. Because of its

higher throughput, Rep-PCR fingerprinting was used to classify the 120 isolates from ileostoma effluent at $t = 1$, identifying 7 genetic lineages within the 3 *Streptococcus* phylotypes (Figure 4.2 and S4.3), whereas no separate genetic lineages were identified for the *Veillonella* phylotype as indicated above. Notably, 7 distinct genetic lineages could be identified among the 2 *Enterococcus* phylotypes (Table 4.3 and S4.2; Figure S4.3).

These findings illustrate the high degree of bacterial richness for the small intestinal ecosystem and confirm that 16S rRNA based approaches underestimate the true diversity of microbial ecosystems that extends to sub-phylotype levels.

Temporal dynamics of ileostoma *Streptococcus* populations

To determine if the occurrence of identified groupings on phylotype and sub-phylotype levels varies in time, the dynamics of the ileostoma effluent populations were assessed by comparing two samples collected one year apart from the same individual.

Isolates from the same *Veillonella* phylotype were obtained from ileostoma effluent at both time points. Intriguingly, only one of the three *Streptococcus* phylotypes that were cultivated from ileostoma effluent at $t = 0$ was recovered from the second ileostoma effluent sample obtained a year later (Figure 4.2). Detailed analysis revealed that of the 7 *Streptococcus* lineages that were cultivated from ileostoma effluent, 5 *Streptococcus* lineages were exclusively cultivated from ileostoma effluent at $t = 0$, while 1 *Streptococcus* lineage was only recovered from ileostoma effluent at $t = 1$. Additionally, 1 *Streptococcus* lineage was cultivated from ileostoma effluent at both time points (Figure 4.2). These findings suggest that while the same species are detected the small intestinal *Streptococcus* populations display population dynamics at the level of the genetic lineages within the phylotypes.

Carbohydrate fermentative capabilities differ between *Streptococcus* lineages

To assess whether genetic differences among the *Streptococcus* lineages are also reflected in their phenotypic characteristics, the carbohydrate-fermentation capabilities of one or two (if available) randomly picked bacterial isolates representing each of the 6 *Streptococcus* genetic lineages from ileostoma effluent at $t = 0$ (Figure 4.2) were tested using the API 50 CH system (Table S4.1). Fermentation profiles between duplicate tests for the same isolate were concordant, albeit that in some cases a clear positive or negative result was obtained for one duplicate while the second duplicate showed a weak reaction. Furthermore, fermentation patterns for isolates from the same lineage only showed minor differences (Table S4.1). All of the tested *Streptococcus* isolates were able to ferment the monosaccharides galactose, glucose, and fructose as well as the disaccharides maltose, lactose, and saccharose (sucrose). The three different *Streptococcus* phylotypes and most of the *Streptococcus* lineages could be discriminated based on their disparate capacity to ferment arabinose, N-acetylglucosamine, amygdalin, arbutin, esculin, salicin, cellobiose, melibiose, trehalose, raffinose, amidon (starch), glycogen, and gentiobiose (Table S4.1).

Table 4.4. OD600 measurements of *Streptococcus* cultures in MS^{Basal} medium supplemented with sugars after 18 hours incubation at 37°C

Lineage ^a (Total number of isolates)	MS ^{Basal} MS	MS ^{Basal} L-Arabinose	MS ^{Basal} D-Glucose	MS ^{Basal} D-Mannitol	MS ^{Basal} N-Acetyl glucosamine	MS ^{Basal} Sucrose	MS ^{Basal} Trehalose	MS ^{Basal} Raffinose	MS ^{Basal} Soluble starch	MS ^{Basal} glycogen
1 (3)	0.204	0.186	1.809	0.205	1.619	1.992	0.127	2.337	0.160	0.160
2 (17)	0.072	0.600	1.489	0.098	1.506	1.973	1.940	1.683	1.119	0.104
4 ^b (1)	0.011	0.104	2.099	0.031	0.035	0.977	2.258	0.029	0.039	0.030
5 (1)	0.097	0.133	1.652	0.096	0.036	1.375	0.015	0.027	0.121	0.087
6 (8)	0.075	0.113	1.755	0.035	0.108	1.348	0.055	0.269	0.079	0.076
7 (66)	0.116	0.202	2.204	0.098	0.19	1.459	2.383	0.175	0.121	0.099

Black: positive growth; White: regarded as no growth

^a The numbering of the phylotypes and lineages are based on the order of groupings in table 3 and (clockwise) groupings in figure S4.3; ^b AFLP analysis unlike Rep-PCR genomic fingerprinting identified a single *Streptococcus* isolate as a separate genetic lineage. According to Rep-PCR genomic fingerprinting this isolate belongs to *Streptococcus* lineage 7.

Members of *Streptococcus* lineage 4 and 7 as well as lineage 5 and 6 could, however, not be distinguished based on their fermentation profiles (Table S4.1). For the isolates belonging to lineage 4 and 7 this appears to reflect their close relatedness as was also concluded from the failure to distinguish these lineages by Rep-PCR genetic fingerprinting. However, the fermentation capabilities of the strains tested here did differ from phenotypic characteristics for closely related streptococci, including: *S. parasanguinis*, *S. equinus*, *S. lutetiensis*, *S. salivarius* and *S. vestibularis* (65). This illustrates the difference in metabolic capabilities exhibited by individual strains of a species, as is also apparent from the distinct fermentation profiles obtained for the 4 *S. salivarius* strains in this study (Table S4.1).

Isolates were further cultured in basal medium (MS^{Basal} medium) supplemented with different sugars as sole carbohydrate and energy source to test exemplarily if the isolates could also utilize the substrates as a carbon source for growth (Table 4.4). Growth was observed for the same substrates that were fermented in the API 50 CH assay, except for isolates from *Streptococcus* lineages, which were able to ferment N-acetyl-glucosamine (lineage 4 and 7), D-raffinose (lineage 5 and 6), and glycogen (lineage 2) according to the API50 assay, but were not able to use this substrate as a carbon source for growth (Table 4.4). Overall, these results demonstrate that the phenotypic diversity of the *Streptococcus* isolates is in good agreement with the determined phylotype-grouping and, to some extent is reflecting the grouping on the sub-phylotype level.

Discussion

Culture (in)dependent analysis of small intestinal *Streptococcus* and *Veillonella* populations

The current study employed cultivation and polyphasic molecular typing to provide increased insights into the richness and dynamics of the small intestinal *Streptococcus* and *Veillonella* populations. Ileostoma effluent samples were used as a representation of the luminal content of the human small intestinal ecosystem. A recent study, however, postulated that oxygen penetration disrupts the ileostoma microbiota and therefore do not represent that of the terminal ileum in healthy subjects (125). Although the influence of oxygen cannot be ruled out, investigations in our laboratory revealed that the ileostoma effluent microbiota contains a high relative abundance of strict anaerobes (31). Moreover, the ileostoma effluent microbiota resembles that encountered in the proximal part of the small intestine of individuals with a normal intestinal tract that includes a colon (339, 381). *Streptococcus* and *Veillonella* populations were detected in all ileostoma effluent samples and showed fluctuations in relative abundance in a 72 hour time frame, which are most likely due to the subject's diet composition. These results are in good agreement with a previous study performed in our laboratory (31). Although the relative abundances of *Streptococcus* and *Veillonella* populations varied per subject and time of sampling, two groups of *Streptococcus* (*S. salivarius* and *S. mitis* group), *Veillonella* spp. (affiliated with *V. atypica*, *V. rogosae*, *V. parvula*, and/or *V. dispar*) were consistently dominant in all ileostoma effluent samples. Therefore, we decided to proceed with deciphering the diversity of the *Streptococcus* and *Veillonella* populations at the genetic lineage level. To facilitate in depth analysis of isolates from ileostoma effluent, we focused on samples collected from a single male ileostomist (subject 1), rather than samples collected from several subjects. Samples were collected with a long time-interval (one year apart) to expand the collection of distinct bacterial genetic lineages relative to the diversity that may be expected from ileostoma effluent collected over a relative short time frame (381). Selective cultivation conditions enabled the targeted isolation of these *Streptococcus* and *Veillonella* spp., although isolates within the genera *Enterococcus*, *Lactobacillus* and *Bacteroides* were also obtained. Though the latter genera were also detected in other ileostoma samples from subject 1, their relative abundance was generally low (<0.7%; Van den Bogert, et al. Unpublished results). The number of obtained *Enterococcus* isolates might be explained by a cultivation bias resulting from the use of selective media that preferentially allow growth of specific microbial groups, and are known to provide a distorted view of bacterial abundances in the original samples (76, 239). Of each of the streptococcal species groups identified by pyrosequencing, representative bacterial isolates, including streptococci affiliated with *S. equinus* and *S. lutetiensis* belonging to the *S. bovis* species group, were cultured from ileostoma effluent. Unambiguous affiliation of the *Veillonella* isolates to specific species, i.e. *V. atypica*, *V. rogosae*, *V. parvula*, and/or *V. dispar*, was not possible based on 16S rRNA gene sequences, which is in agreement with what has previously been

described for species within this genus (65). Nonetheless, MALDI-TOF MS analysis identified the *Veillonella* phylotype as *V. parvula*. The fact that isolates from *Streptococcus*, *Veillonella* and three other genera were obtained shows that these populations were alive at the time of sampling and may be part of the active small intestinal microbiota.

The *Streptococcus* isolates obtained from ileostoma effluent collected at two time points could be clustered into multiple species-groups, which could be further subdivided into strains belonging to distinct genetic lineages on basis of genetic fingerprinting. Although not the prime subject of this study, also the *Enterococcus* isolates were found to display a substantial level of phylogenetic richness, indicating that these groups of small intestinal cocci encompass a high degree of genetic diversity. The two ileostoma samples that were taken one year apart from the same individual revealed distinct *Streptococcus* lineages, while only representative strains of one of the *Streptococcus* lineages were cultivated at both time points. This indicates that the fluctuations in relative abundance seen at the genus and species level are confirmed and expanded at the genetic lineage level. Moreover, while representative strains of the same species are frequently detected in multiple samples from one or several individuals, the occurrence of the corresponding genetic lineages may be quite different. Genetic fingerprinting methods have been widely applied to discriminate strains from *Streptococcus* spp. such as *S. pneumoniae* (79) and *S. pyogenes* (90), and were used to assess the oral *Streptococcus* diversity at the strain level, which showed that the oral cavity of most subjects harbored multiple genotypes of *S. mutans* (46, 375) and *S. oralis* (73). Similarly, *Enterococcus faecium* strains from different sources were differentiated (134). Furthermore, the temporal fluctuations of strain abundances of *Lactobacillus* and *Bifidobacterium* spp. in fecal samples were analyzed previously (230), revealing similar results as described here. However, to the best of our knowledge, the (small) intestinal *Streptococcus* richness has not yet been assessed to the sub-phylotype level. We hypothesize that this is important for the functioning of the ecosystem, since the substrate conversion capacities were different among the *Streptococcus* lineages. Furthermore, these observations underpin the limitation of species-level identifications of intestinal bacteria on basis of 16S rRNA gene sequences alone, when it comes to the prediction of function of the microbiota. This is especially true for the novel high throughput sequencing technologies that provide only partial 16S rRNA gene sequences.

Possible interactions of small intestinal bacterial populations with the human host

Notably, streptococci and *Veillonella* are also abundant in other sections of the upper GI tract (1, 9, 23, 256) and likely originate from their abundant populations in the oral cavity (353). Molecular typing of oral *Streptococcus* and *Veillonella* strains, isolated from the ileostomist studied here (subject 1), identified a *Veillonella* phylotype and three *S. salivarius* lineages that, remarkably, group together with those that were cultivated from ileostoma effluent (Van den Bogert, et al. Unpublished results;

Chapter 7). Although speculative, this first comparison between oral and small intestinal microbiotas suggests that the oral microbiota may serve as an inoculum for the upper GI tract. Though this suggests that these populations are allochthonous to the small intestine (353), the considerable common abundance of the streptococci and their high activity in efficient uptake and fermentation of the available (diet-derived) carbohydrates (381), suggest that *Streptococcus* populations play a prominent role in the primary carbohydrate metabolism occurring in the small intestinal ecosystem.

Since the number of isolates obtained from the ileostoma effluent samples is limited, it cannot be ruled out that lineages that were recovered only from one of the two ileostomy samples would in fact be shared lineages when larger numbers of isolates were to be analyzed. However, some of these lineages were represented by a relatively large number of isolates (e.g. *Streptococcus* lineage 3 and 7), indicating that even if these lineages were present at both time points, their relative abundance would have differed considerably.

Preliminary investigations in our laboratory identified unique genetic markers for several of the different *Streptococcus* lineages. Remarkably, these genetic markers were detected in small intestinal samples of other human individuals, suggesting that these lineages are common commensals of the small intestinal microbiota (Chapter 5). It is well-known that dietary changes lead to an alteration of intestinal microbial composition, which has a profound influence on responses of the host immune system (for a recent review see (226)). Furthermore, recent studies revealed that substantially different mucosal responses and immunoregulatory cascades can be modulated by closely related species (335), different bacterial strains (233), and even different preparations of the same strain (336). Based on these findings, in combination with the discriminating fermentation and growth patterns among the *Streptococcus* strains and lineages described here, it is tempting to speculate that there is a potential for directed modulation of mucosal immune responses by dietary modulation of the endogenous *Streptococcus* populations. To this end, elucidating the role of the small intestinal microbiota, especially of the abundant and diverse *Streptococcus* population, is a task for the future.

General conclusion

The work presented here demonstrates high intra-genus and intra-species genetic diversity of the small intestinal microbiota, focusing on populations of *Streptococcus* and *Veillonella*. It is of particular interest to assess whether the small intestinal *Streptococcus* and *Veillonella* isolates have the potential for metabolic interaction similar to what is observed in the oral cavity (84). In addition, our findings show that the *Streptococcus* population in the small intestine has a high phenotypic variability, which may be a dominant driver of the high population dynamics of the small intestinal streptococci in response to varying nutrient availability that is caused by variable food intake. These dynamic streptococcal populations may profoundly influence local host-microbe interactions, thereby modulating small intestinal physiology and immune system functions.

Acknowledgements

This project was partially supported by the Netherlands Bioinformatics Centre (NBIC). We appreciate the help of Freddy Troost in obtaining approval for this study by the Maastricht Medical Ethics Committee (METC). We thank Milkha Leimena for assistance with collecting ileostoma effluent samples and Kun Xie for assistance with cultivation of isolates and sequencing of 16S rRNA genes, as well as Hauke Smidt and Detmer Sipkema for critical reading of the manuscript. We are very grateful for the participation of all ileostomy subjects who provided samples. We thank Hermie Harmsen of the University Medical Centre Groningen (UMCG; Groningen, The Netherlands) for facilitating the MALDI-TOF MS analyses.

Supplementary information

Supplementary material and methods – MALDITOF MS database

Veillonella and *Streptococcus* spp. included in the MALDITOF MS database:

<i>Veillonella atypica</i>	2 strains	<i>Streptococcus infantis</i>	1 strain
<i>Veillonella caviae</i>	1 strain	<i>Streptococcus intermedius</i>	1 strain
<i>Veillonella criceti</i>	1 strain	<i>Streptococcus lutetiensis</i>	1 strain
<i>Veillonella denticariosi</i>	2 strains	<i>Streptococcus macacae</i>	1 strain
<i>Veillonella dispar</i>	1 strain	<i>Streptococcus marimammalium</i>	1 strain
<i>Veillonella montpellierensis</i>	1 strain	<i>Streptococcus massiliensis</i>	1 strain
<i>Veillonella parvula</i>	2 strains	<i>Streptococcus minor</i>	3 strains
<i>Veillonella ratti</i>	1 strain	<i>Streptococcus mitis</i>	1 strain
<i>Veillonella rogosae</i>	1 strain	<i>Streptococcus mutans</i>	1 strain
<i>Veillonella sp.</i>	1 strain	<i>Streptococcus oralis</i>	4 strains
<i>Streptococcus acidominimus</i>	1 strain	<i>Streptococcus orisratti</i>	1 strain
<i>Streptococcus agalactiae</i>	9 strains	<i>Streptococcus orisuis</i>	1 strain
<i>Streptococcus alactolyticus</i>	1 strain	<i>Streptococcus ovis</i>	1 strain
<i>Streptococcus anginosus</i>	2 strains	<i>Streptococcus parasanguinis</i>	2 strains
<i>Streptococcus australis</i>	1 strain	<i>Streptococcus parauberis</i>	2 strains
<i>Streptococcus caballi</i>	1 strain	<i>Streptococcus peroris</i>	1 strain
<i>Streptococcus canis</i>	2 strains	<i>Streptococcus phocae</i>	2 strains
<i>Streptococcus constellatus</i>	2 strains	<i>Streptococcus pleomorphus</i>	1 strain
<i>Streptococcus criceti</i>	1 strain	<i>Streptococcus pluranimalium</i>	1 strain
<i>Streptococcus cristatus</i>	2 strains	<i>Streptococcus pneumoniae</i>	9 strains
<i>Streptococcus dentirousetti</i>	1 strain	<i>Streptococcus porcinus</i>	1 strain
<i>Streptococcus devriesei</i>	2 strains	<i>Streptococcus pseudopneumoniae</i>	1 strain
<i>Streptococcus didelphis</i>	1 strain	<i>Streptococcus pseudoporcinus</i>	1 strain
<i>Streptococcus downei</i>	1 strain	<i>Streptococcus pyogenes</i>	8 strains
<i>Streptococcus dysgalactiae</i>	6 strains	<i>Streptococcus ratti</i>	1 strain
<i>Streptococcus entericus</i>	1 strain	<i>Streptococcus salivarius</i>	4 strains
<i>Streptococcus equi</i>	5 strains	<i>Streptococcus sanguinis</i>	4 strains
<i>Streptococcus equinus</i>	2 strains	<i>Streptococcus sinensis</i>	1 strain
<i>Streptococcus ferus</i>	1 strain	<i>Streptococcus sobrinus</i>	2 strains
<i>Streptococcus gallinaceus</i>	1 strain	<i>Streptococcus suis</i>	3 strains
<i>Streptococcus gallolyticus</i>	8 strains	<i>Streptococcus thermophilus</i>	6 strains
<i>Streptococcus gordonii</i>	3 strains	<i>Streptococcus thoralensis</i>	2 strains
<i>Streptococcus halichoeri</i>	1 strain	<i>Streptococcus uberis</i>	1 strain
<i>Streptococcus hyointestinalis</i>	1 strain	<i>Streptococcus urinalis</i>	1 strain
<i>Streptococcus hyovaginalis</i>	2 strains	<i>Streptococcus vestibularis</i>	1 strain
<i>Streptococcus infantarius</i>	1 strain	<i>Streptococcus sp.</i>	2 strains

Supplementary figures

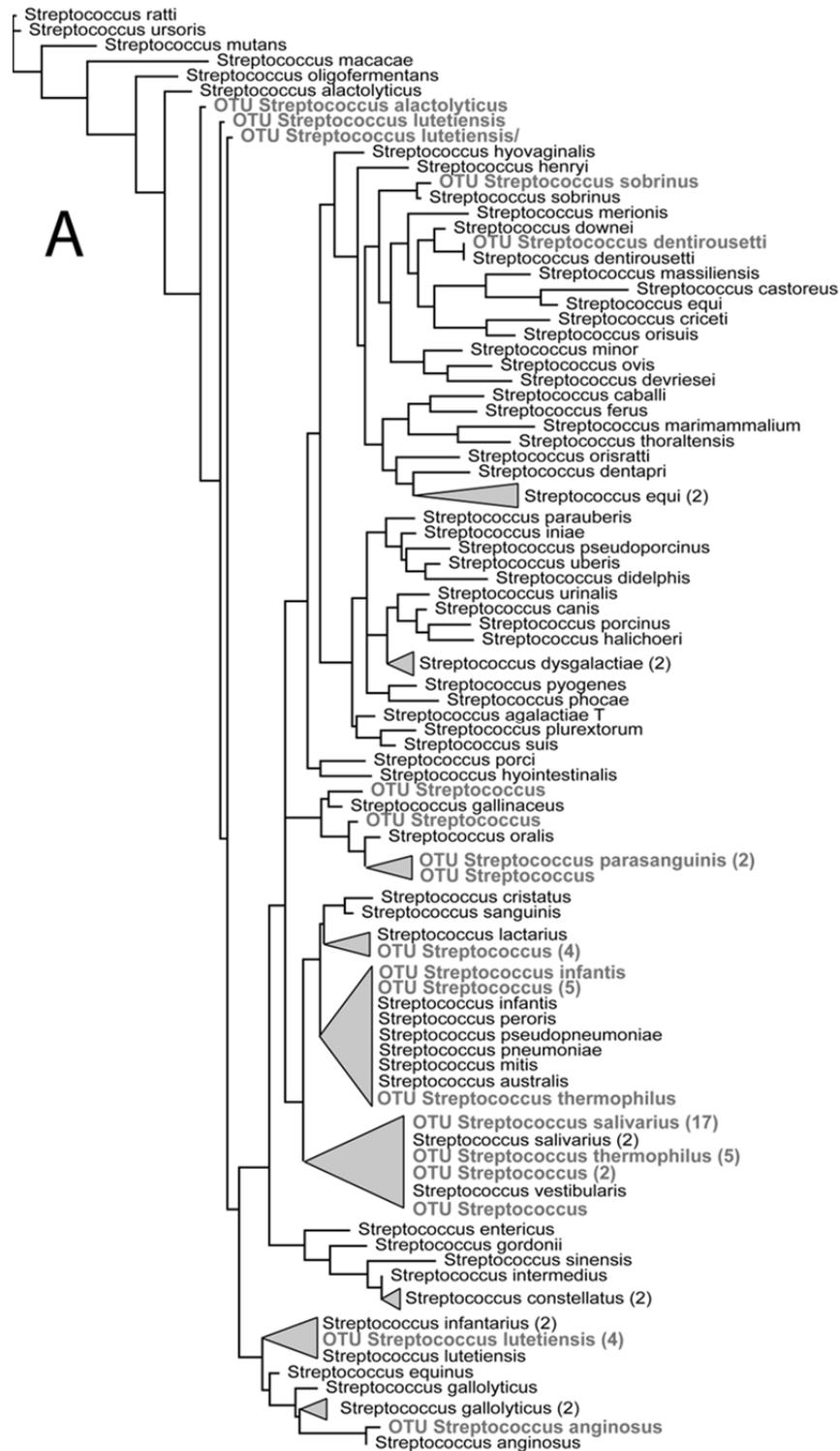


Figure S4.1. Maximum likelihood phylogenetic trees containing the reference sequences from the RDP database and a representative sequence from each OTU from pyrosequencing for *Streptococcus* (A) and *Veillonella* (B). Tree leaves beginning with 'OTU' represent sequences that were obtained from pyrosequencing. Numbers between brackets indicated the number of sequences for that tree leaf.

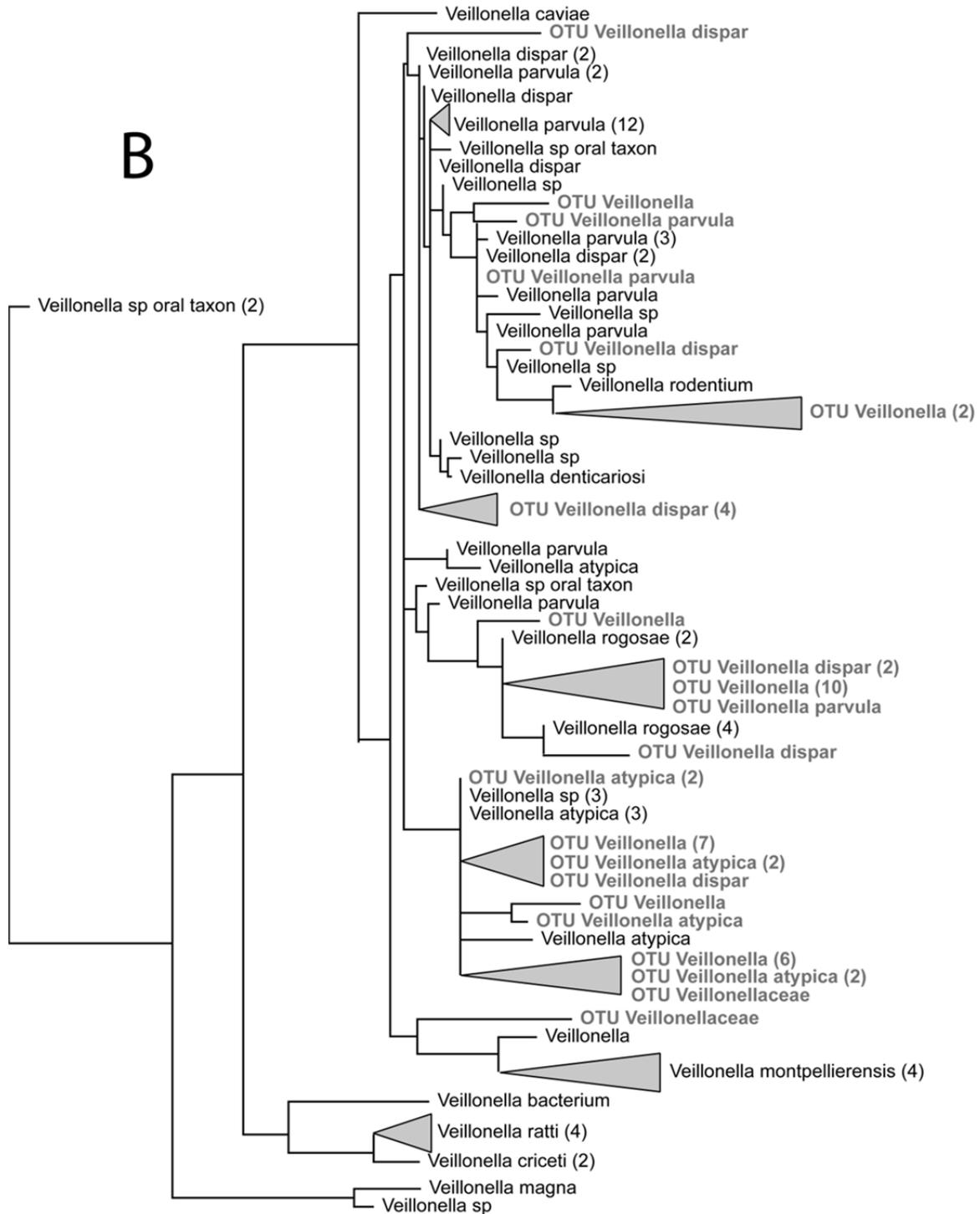


Figure S4.1 (Continued). Maximum likelihood phylogenetic trees containing the reference sequences from the RDP database and a representative sequence from each OTU from pyrosequencing for *Streptococcus* (A) and *Veillonella* (B). Tree leaves beginning with 'OTU' represent sequences that were obtained from pyrosequencing. Numbers between brackets indicated the number of sequences for that tree leaf.



Figure S4.2. UPGMA clustering of AFLP and Rep-PCR profiles with Pearson correlation coefficients from randomly selected *Streptococcus* and *Enterococcus* isolates obtained for genetic lineages cultured from ileostoma effluent at t = 0.

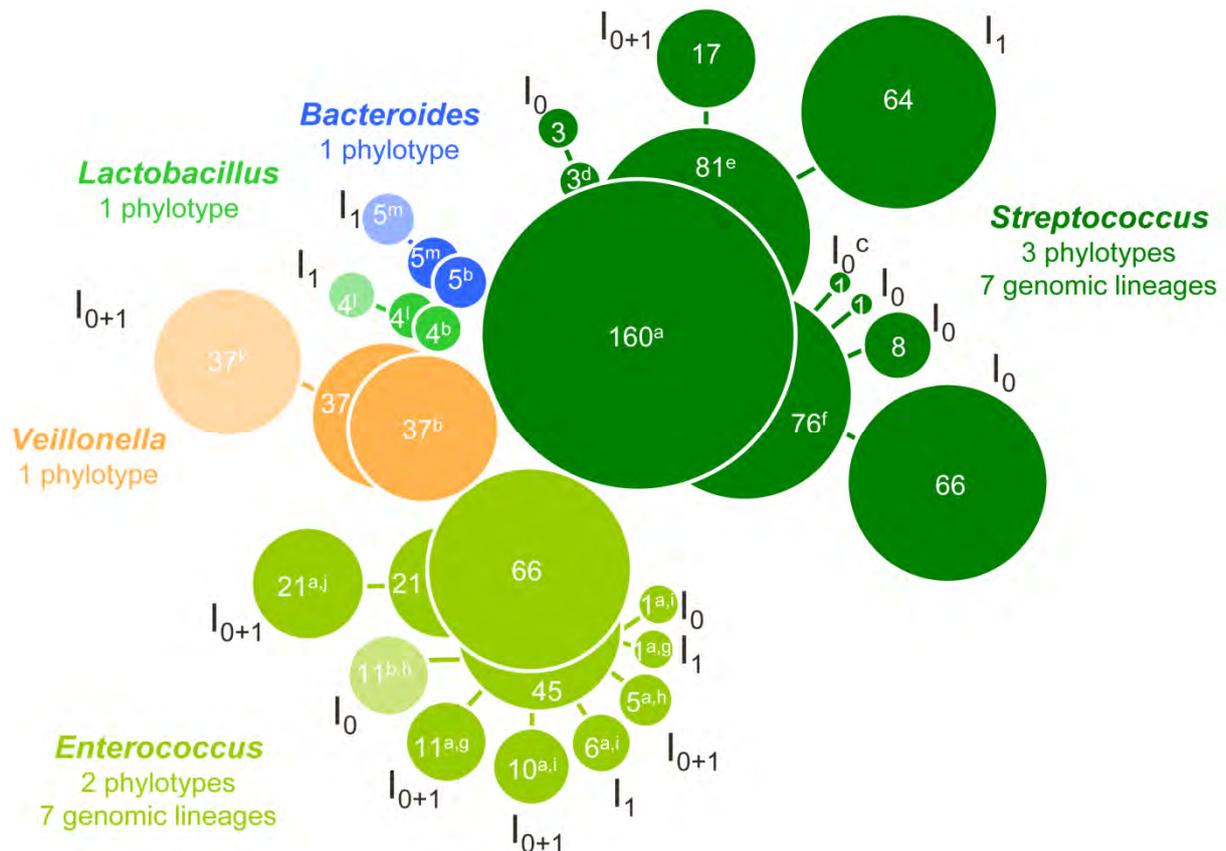


Figure S4.3. Groupings of *Streptococcus*, *Enterococcus*, *Veillonella*, *Lactobacillus*, and *Bacteroides* spp. (inner circles) isolates obtained from ileostoma effluent into phylotypes (middle circles) and genetic lineages (outer circles). Faded colored groupings represent the number of bacterial isolates AFLP analysis and/or Rep-PCR genomic fingerprinting did not reveal discriminative lineages. I₀/I₁: isolates obtained from ileostoma effluent at time point 0 and/or 1. a: isolates obtained from MS agar; b: isolates obtained from VSA; c: AFLP analysis unlike Rep-PCR genomic fingerprinting identified a single *Streptococcus* isolate as a separate genetic lineage; d: *S. mitis* group; e: *S. bovis* group; f: *S. salivarius* group; g: *E. faecium* group; h: *E. gallinarum* group; i: *E. avium* group; j: *E. faecalis*; k: *V. parvula*; l: *L. fermentum*; m: *B. fragilis*.

Supplementary tables

Table S4.1. API strip scores for MS bacterial isolates

Classification (genus)	Phylotype ^a	Lineage ^a	(Total number of isolates)	<i>Streptococcus</i>										
				1		2		3		7				
				1	2	1	2	1	1	1	2	1	2	
CONTROL	0													
Glycerol	1													
Erythritol	2													
D-Arabinose	3													
L-Arabinose**	4			w	w									
D-Ribose**	5													
D-Xylose**	6													
L-Xylose	7													
D-Adonitol	8													
Methyl-βD-Xylopyranoside	9													
D-Galactose**	10		w						w	w			w	
D-Glucose**	11		w						w	w				
D-Fructose**	12		w						w	w				
D-Mannose**	13		d	w				d	d	d	w			w
L-Sorbose	14													
L-Rhamnose	15													
Dulcitol	16													
Inositol	17													
D-Mannitol	18													
D-Sorbitol	19													
Methyl-αD-Mannopyranoside	20													
Methyl-αD-glucopyranoside	21													
N-AcetylGlucosamine	22		d					d					d	
Amygdalin	23													
Arbutin	24													
Esculin ferric citrate	25													
Salicin	26													w
D-Cellobiose**	27							w					w	
D-Maltose**	28		w								w			
D-Lactose (bovine origin)	29		w								w			w
D-Melibiose**	30		d						w					
D-Saccharose (sucrose) **	31		w								w			
D-Trehalose**	32													
Inulin	33							d						
D-MeLeZitose	34													
D-Raffinose	35								w	w				
AmiDon (starch)	36													
Glycogen	37													
Xylitol	38													
Gentiobiose	39													
D-Turanose	40													
D-Lyxose	41													
D-Tagatose**	42													
D-Fucose	43													
L-Fucose	44													
D-Arabitol	45													
L-arabitol	46													
Potassium GlucoNaTe	47													
Potassium 2-KetoGluconate	48													
Potassium 5-KetoGluconate	49													

Black: positive reaction; Grey
w: weak reaction, d: variable
reaction between duplicate
tests of the same isolate;
White: negative reaction;

^a: The numbering of the
phylotypes and lineages are
based on the order of
groupings in table 3 and
(clockwise) groupings in figure
S3;

^b: To determine if fermentation
patterns were consisted for
isolates belonging to the same
genomic lineage, two isolates
from most of genomic lineages
were tested, if possible; ^c:
Number 1 isolates were tested
in duplicate; ^d: AFLP analysis
unlike Rep-PCR genomic
fingerprinting identified a single
Streptococcus isolate as a
separate genomic lineage.
According to Rep-PCR
genomic fingerprinting this
isolate belongs to
Streptococcus lineage 7.

Table S4.2. Characteristics of isolates from ileostoma effluent

Available upon request or can be downloaded from the online version of this article
(<http://onlinelibrary.wiley.com/doi/10.1111/1574-6941.12127/supinfo>)

Chapter 5

Comparative genomics analysis of *Streptococcus* isolates from the human small intestine reveals their adaptation to a highly dynamic ecosystem

Bartholomeus van den Bogert, Jos Boekhorst, Ruth Herrmann, Eddy J. Smid, Erwin G. Zoetendal, and Michiel Kleerebezem

Manuscript submitted for publication

Abstract

The human small-intestinal microbiota is characterised by relatively large and dynamic *Streptococcus* populations. In this study, genome sequences of small-intestinal streptococci from *S. mitis*, *S. bovis*, and *S. salivarius* species groups were determined and compared with those from 58 *Streptococcus* strains in public databases. The *Streptococcus* pangenome consists of 12,403 orthologous groups of which 574 are shared among all sequenced streptococci and are defined as the *Streptococcus* core genome. Genome mining of the small-intestinal streptococci focused on functions playing an important role in the interaction of these streptococci in the small-intestinal ecosystem, including natural competence and nutrient-transport and metabolism. Analysis of the small-intestinal *Streptococcus* genomes predicts a high capacity to synthesize amino acids and various vitamins as well as substantial divergence in their carbohydrate transport and metabolic capacities, which is in agreement with observed physiological differences between these *Streptococcus* strains. Gene-specific PCR-strategies enabled evaluation of conservation of *Streptococcus* populations in intestinal samples from different human individuals, revealing that the *S. salivarius* strains were ubiquitously detected in the small-intestine microbiota, supporting the representative value of the genomes provided in this study. Finally, the *Streptococcus* genomes allow prediction of the effect of dietary substances on *Streptococcus* population dynamics in the human small-intestine.

Introduction

Streptococcus is a genus of Gram-positive, low GC-rich species belonging to the lactic acid bacteria (LAB) in the family Streptococcaceae (65). While several *Streptococcus* species, such as *S. pyogenes* (214) and *S. pneumonia* (162) are recognized as human pathogens, others like *S. salivarius*, *S. mitis*, *S. parasanguinis* are commonly detected as relatively dominant inhabitants in the upper respiratory tract (65), oral cavity (1, 169), throat (9), esophagus (256), stomach (23), and small intestine (31, 339, 381) of healthy individuals.

Studies carried out by our laboratory focused on elucidating the composition and function of the microbial community in the small intestine, using ileostoma effluent samples as a representation of the luminal content of the small intestine (31, 339, 340, 381). In ileostomy subjects the terminal ileum is connected to an abdominal stoma making this region of the intestinal tract accessible for non-invasive and repetitive sampling of the luminal fraction of the small-intestinal microbiota (31, 339, 381). The microbial composition in ileostoma effluent resembled the microbiota that resides in the proximal part of the small intestine from individuals with an intact intestinal tract (341, 381). Although *Streptococcus* spp. were detected in each of collected ileostoma effluent samples, their relative abundance fluctuated greatly between individuals and even between samples obtained from the same individual (31, 340). Furthermore, metatranscriptomic analysis of ileostoma effluent identified carbohydrate transport systems, including several phosphotransferase systems (PTS) among the highly represented expressed functions in the small-intestinal streptococci, suggesting that the activity of these bacteria is focused on efficient uptake and fermentation of the available (diet-derived) carbohydrates in the human small intestine (381).

Based on the above, the *Streptococcus* populations in the small intestine are predicted to play a prominent role in the metabolic conversion of primary carbohydrates that are present in this ecosystem, and may thereby effectively compete for dietary carbohydrate nutrients with the host mucosa. Therefore, a cultivation approach was employed to obtain representative *Streptococcus* isolates from the small-intestinal ecosystem ((340), Chapter 4). Classification of the isolates on the basis of molecular typing methodologies showed that from one ileostoma sample alone 3 different *Streptococcus* species were recovered belonging to the *S. mitis* group, *S. bovis* group, and *S. salivarius* group of which the latter could be further divided in four genetic lineages (strain level). Although considerable temporal fluctuations of distinguishable genetic lineages were observed when a second sample was collected and investigated one year later, isolates belonging to a single lineage were recovered from both ileostoma effluent samples. Moreover, the *Streptococcus* lineages displayed different carbohydrate conversion and growth patterns ((340); Chapter 4). However, the mechanisms underlying the dynamics at the genetic lineage level is unclear. Therefore, specific aspects of the environmental interaction-potential and the metabolic capacity of 6 small-intestinal *Streptococcus* strains were investigated through analysis of their genome sequences in this study.

Furthermore, the genomes enabled the comparison with other streptococci from other niches, allowing the identification of genetic targets for strain-specific PCR-based detection in intestinal samples from different individuals.

Materials and methods

***Streptococcus* isolates and chromosomal DNA extraction**

The isolation of the small-intestinal *Streptococcus* strains and their molecular typing was described previously ((340); Chapter 4). In short, isolates were obtained from ileostoma effluent plated on *Mitis Salivarius* (MS) agar (Becton Dickinson, Breda, the Netherlands) supplemented with Tellurite solution 1% (Becton Dickinson). The streptococcal isolates were classified by DNA fingerprinting into 6 genetic lineages that, in the remainder of the paper are indicated with their *Streptococcus* species group names: *S. mitis* (1 lineage), *S. bovis* (1 lineage) and *S. salivarius* (4 lineages). A randomly picked representative isolate of each lineage was selected for whole genome sequencing.

Genomic DNA of the isolates was extracted from bacterial cells that were grown overnight in 10 ml MS medium at 37°C. Cells were pelleted by centrifugation at 7250g at 4°C for 15 minutes and subsequently frozen at -20°C. Thawed cell-pellets were resuspended in 2 ml THMS (30 mM TRIS-HCl [Sigma, St. Louis, MO, USA; pH = 8.0], 25% (w/v) sucrose [Sigma], and 3 mM MgCl₂ [Riedel-de Haën, Seelze, Germany]) supplemented with 10 mg/ml lysozyme (Sigma) and 40 µl mutanolysin (Sigma; 5000 U/ml), aliquoted in equal amount into 2 eppendorf tubes, and incubated at 37°C for 30 minutes. After centrifugation at 14.000g for 5 minutes and discarding the supernatant, cells were resuspended in 100 µl THMS and mixed with 400 µl TES (50 mM TRIS-HCl [Sigma; pH = 8.0], 20 mM EDTA [Sigma; pH = 8.0], 50 mM NaCl [Merck], containing 0.5% (v/v) SDS [Ambion, Austin, TX, USA]) and 20 µl Proteinase K (20 mg/ml) followed by incubation at 56°C for 15 minutes. Nucleic acids were subsequently purified by sequential extraction with acid-phenol (Phenol:Water (3.75:1 v/v); pH = 4.45-5.68; Invitrogen, Carlsbad, CA, USA), acid-phenol:chloroform (1:1), and chloroform (Sigma-Aldrich, Zwijndrecht, Netherlands) using standard procedures as described by Sambrook, et al. (285). DNA was precipitated from the water-phase by standard ethanol precipitation (285). After drying, the DNA pellets were dissolved in 50 µl nuclease free water (Promega, Leiden, Netherlands). One µl RNase A (10 mg/ml; Qiagen GmbH, Hilden, Germany) was added to the solution followed by incubation at 37°C for 30 minutes. Samples were stored at 4°C. DNA quality and concentrations were determined by nanodrop and on a 1.0% (w/v) agarose gel containing 0.4 µg/ml ethidium bromide (Bio-rad).

Genome sequencing and annotation

DNA from bacterial isolates was sequenced using 454 GS FLX (Roche) technology in combination with titanium chemistry, producing 350-450 bp reads (234,320 ± 86,626 reads per genome), and by using Illumina HiSeq 2000 technology, producing 11,884,010 ± 1,026,060 paired reads of 50 bp per genome from 3 kb mate pair

libraries (Table S5.1; GATC-Biotech, Konstanz, Germany). Pyrosequence reads were assembled using the Celera Assembler v6.1 (http://sourceforge.net/apps/mediawiki/wgs-assembler/index.php?title=Main_Page), and the resulting contigs were subsequently combined with paired-read Illumina sequencing data to generate scaffolds using the SSPACE software v1.1 (26). Genome pseudo-assemblies were constructed by placing scaffolds in their likely order based on comparisons with the genomes from closely related bacteria: *Streptococcus parasanguinis* ATCC 15912 ([Genbank: NC_015678]; *S. mitis* species group), *Streptococcus gallolyticus* UCN34 ([Genbank: NC_013798], *S. bovis* species group), and *Streptococcus salivarius* CCHSS3 ([Genbank: NC_015760]; *S. salivarius* species group). These comparisons were manually screened for inconsistencies using the Artemis comparison tool (41). Genomes were annotated using the RAST server (14). The genes predicted in the genomes of the six small-intestinal isolates were assigned to Cluster of Orthologous groups (COG; (321)) categories, using blastp comparison with the COG database (NCBI, <ftp://ftp.ncbi.nih.gov/pub/COG/COG>) using an alignment E-value cut-off of 10^{-3} .

Strain identifiers and accession numbers

The Whole Genome Shotgun projects of the human small intestinal *Streptococcus* strains have been deposited at DDBJ/EMBL/GenBank under the following strain identifiers (and accession number): *S. mitis* species group strain: HSISM1 (ASKI00000000), *S. bovis* species group strain: HSISB1 (ASKA00000000), *S. salivarius* species group strain 1: HSISS1 (ASKB00000000), *S. salivarius* species group strain 2: HSISS2 (ASKC00000000), *S. salivarius* species group strain 3: HSISS3 (ASKH00000000), and *S. salivarius* species group strain 4: HSISS4 (ASKD00000000). The version described in this paper is version XXXX01000000.

Genome orthology

Orthology relationships were identified by comparing all predicted gene products from all 6 small-intestinal *Streptococcus* genomes with the genes predicted to be encoded by the 58 other *Streptococcus* genomes (see Table S5.2 for accession numbers) that were available within the NCBI database on February 22nd, 2012 using OrthoMCL v2.0.2 with default parameters. Genome metadata (e.g. isolation site) from the *Streptococcus* genomes was retrieved from the Genome OnLine database (GOLD; <http://genomesonline.org>) on February 27th, 2012 (Table S5.2).

Streptococcus phylogenetic tree reconstruction

Multiple protein sequence alignments of the 450 orthologous groups with exactly one member in each *Streptococcus* genome were generated using MUSCLE (81). The variable positions were concatenated into a single alignment (length 5605 residues) and a maximum-likelihood phylogenetic tree was generated using PhyML (116). The phylogenetic tree was visualized using the TREEVIEW program (250).

Genome mining and metabolic mapping

Bacterial genomes were mined for systems involved in responses to external stimuli, focusing on bacteriocins, identified using BAGEL2 employing no re-annotation (62), and two-component systems (TCS) consisting of sensor histidine kinase (HK) and response regulator (RR) pairs (307).

Moreover, genomes were screened for gene clusters involved in regulation of natural competence: comCDE, present in the *S. mitis* group species, or comRS in *S. bovis* and *S. salivarius* streptococci (126).

Genomes were further screened for sugar transport systems including constituents of the bacterial phosphotransferase system and ABC transporters. Metabolic and amino acid biosynthesis pathways were constructed for the newly sequenced genomes by mapping EC numbers from the genome annotations onto the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways (164). Pathways from individual KEGG maps that were represented in at least one of the *Streptococcus* genomes were included in combined metabolic visualizations for sugar metabolism and amino acid biosynthesis that were manually constructed. In cases where genes of key enzymes in specific pathways of interest were apparently absent from the genome-based predictions, a further effort was made to identify homologous gene candidates by dedicated BLAST searches (7).

Unique gene identification and PCR detection

Each of the newly sequenced genomes was screened for 'unique' genes that were not present in other small-intestinal *Streptococcus* genomes or other genomes in the NCBI database. Single copy unique genes with a sequence length of at least 750 nt were used for primer design employing the Primer-BLAST tool (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>), which uses the Primer3 program (278). Default parameters were used, except for the following changes: PCR product size: 150 to 300 bp; maximum primer size: 23 nt; minimum GC content: 40%; maximum poly-X (mononucleotide repeats): 3; maximum self-complementarity: 3.

Primer specificity was checked by submitting each primer to Primer-BLAST using genomes, "Genomes (chromosomes from all organisms)" from all Bacteria, as a reference database. An in-house perl script was used to determine if the primers designed had exact matches in small-intestinal *Streptococcus* genomes other than the intended *Streptococcus* strain target. This revealed that primers developed for *S. salivarius* lineage 4 were not exclusively specific for the intended target strain, but were predicted to be cross reactive with *S. salivarius* lineage 1. By decreasing the minimal gene sequence length to 500 nt, primers were developed that were specific for *S. salivarius* lineage 4.

Primers that passed each screening step, were specific for their target strain, and had a minimal tendency to form secondary structures, including hairpin loops, heterodimers, and homodimers (analysed by the IDTDNA Oligoanalyzer 3.1; Integrated DNA Technologies) were ordered (Biolegio BV, Nijmegen, Netherlands) and tested for their application in strain specific PCR detection assays (see below; Table S5.3).

All PCRs were performed on a C1000™ Thermal Cycler (Bio-rad) with a CFX96 optic module (Bio-rad) employing CFX Manager 2.1 (Bio-rad) software for analysis. Reactions were carried out in Hard-Shell semi skirted clear 96 well plates (Bio Rad) sealed with Microseal B film (Bio Rad) in 25 µl volumes using IQ SYBR green supermix (Bio-Rad) according to the manufacturer's instructions with 200 nM of forward and reverse primer and either 5 µl gDNA (10-20 ng/µl) or glycerol stock as a template source.

The optimal annealing temperature (60°C) for each primer pair was determined by an 8-degree temperature (53°C to 64°C) gradient PCR using gDNA from target strains as template (data not shown).

The PCR program started with a denaturation step at 95°C for 5 minutes, followed by 40 cycles consisting of denaturation at 95°C for 15 s, annealing for 60°C for 30 s and elongation at 72°C for 20 s with data collection, and a final elongation step at 72°C for 10 minutes. Ct values above 35 were considered negative. Melting curve analysis was carried out by incrementally increasing the temperature from 55°C to 95°C at 30 s per 0.5°C with continuous fluorescence collection. Control PCRs were performed alongside each separate amplification without addition of template and consistently yielded no product.

Small-intestinal and fecal sample collection

In total, 30 ileostoma effluent samples were collected in the morning or afternoon (at least 3 h apart) on separate days (at least two days apart) from 6 ileostomy subjects (4 male and 2 female; aged 55 to 79; A-F), as part of previous projects, results of which are reported elsewhere ((31, 340); Chapter 3 and 4). Small-intestinal fluid samples were obtained from 3 healthy individuals (3 males; 24 ± 4.5 years; G-I) and included a jejunal sample and an ileum sample from subject H and a single ileum sample from subjects G and I.

Fecal samples were collected from 10 individuals (4 male and 6 female; aged 19 to 33; J-S) as part of a previous project (322). DNA was extracted using the Repeated Bead Beating method described in (283) or using a method adapted from Zoetendal, et al. (379), depending on the study they originated from, and was used to screen for the unique targets of the *Streptococcus* genetic lineages.

Results

General features of small-intestinal streptococcal genome sequences

The entire genome set analysed in this study consisted of 64 genomes, encompassing 20 *Streptococcus* species. Six draft genome sequences were obtained from strains originating from the small intestine, which were determined in this study and ranged in genome size from 1.9 Mbp (*S. bovis*) to 2.4 Mbp (*S. salivarius* lineage 3; See table S5.1 for genome statistics). The full complement of genes (pangenome) of the *Streptococcus* genome set consisted of 12,403 orthologous groups (OG), of which 4,232 OG were represented in the genomes of at

least one of the six small-intestinal *Streptococcus* strains. The size of the *Streptococcus* pangenome estimated here is somewhat larger as has been suggested in previous studies (197, 216). However, these studies based their pangenome estimates on a smaller genome set comprising fewer species. Furthermore, the *Streptococcus* pangenome defined here does not seem to be exceptionally high compared to, for example, the *Lactobacillus* pangenome estimated to consist of over 13,000 protein-encoding genes (166) or gene families (216). Further analysis revealed that all 64 *Streptococcus* genomes shared 574 OG, defining the core *Streptococcus* genome. All OG belonging to the core *Streptococcus* genome could be classified to a COG, although 26% of these OGs was assigned to poorly characterized COG categories (Figure 5.1). Most OG in the core *Streptococcus* genome were predicted to be involved in information storage and processing (29.2%), with most genes belonging to typically conserved functions such as 'Translation, ribosomal structure and biogenesis' and 'Replication, recombination and repair'. Metabolic functions accounted for 28.4% of the core *Streptococcus* OG, followed by 15.7% of OG that were involved in cellular processes and signalling. Most OG belonging to 'metabolism' were assigned to functions in transport and metabolism of nucleotides and carbohydrates.

Most streptococcal genome sequences from the public databases that were included in the analysis here, were derived from clinical bacterial isolates cultivated from different human body sites (Table S5.2). Due to frequent obscurity concerning the source of isolation it is far from trivial to identify niche-specific OG. Nonetheless, 197 OG, mostly belonging to 'metabolism', were found to be present in at least two of the small-intestinal *Streptococcus* genomes, but not in any of the genomes from the public databases. Notably, a considerable fraction of these 197 small-intestine specific OG (65; 33.0%) could not be assigned to a COG, and were predominantly annotated as hypothetical proteins (44; Figure 5.2). This suggests that for a substantial amount of small-intestine niche-specific streptococcal OG, the function needs to be further elucidated (Figure 5.2).

Phylogenetic analysis of *Streptococcus* genomes

The subset of genes of the core *Streptococcus* genome (450) that were present in single copy in each genome was used to construct a phylogenetic tree (Figure 5.3). This core-genome-based phylogeny revealed a division of 5 distinct clusters of *Streptococcus* strains that belong to the *Streptococcus* species groups: Pyogenic (e.g. *S. pyogenes*), *S. bovis*, *S. mutans*, *S. salivarius*, and *S. mitis* groups. *S. suis* genomes could not be assigned to one of these *Streptococcus* species-groups (65), but represented an additional and separate phylogenetic branch (Figure 5.3). Notably, 5-140 orthologous groups were exclusively present in all genomes belonging to one of the clusters (cluster-specific genes) and might be used as cluster-markers for molecular detection and quantification (Table S5.4). The small-intestine *Streptococcus* genomes clustered within the *S. mitis*, *S. bovis*, and *S. salivarius* groups, corroborating previous classifications based on 16S rRNA gene analysis ((340); Chapter 4).

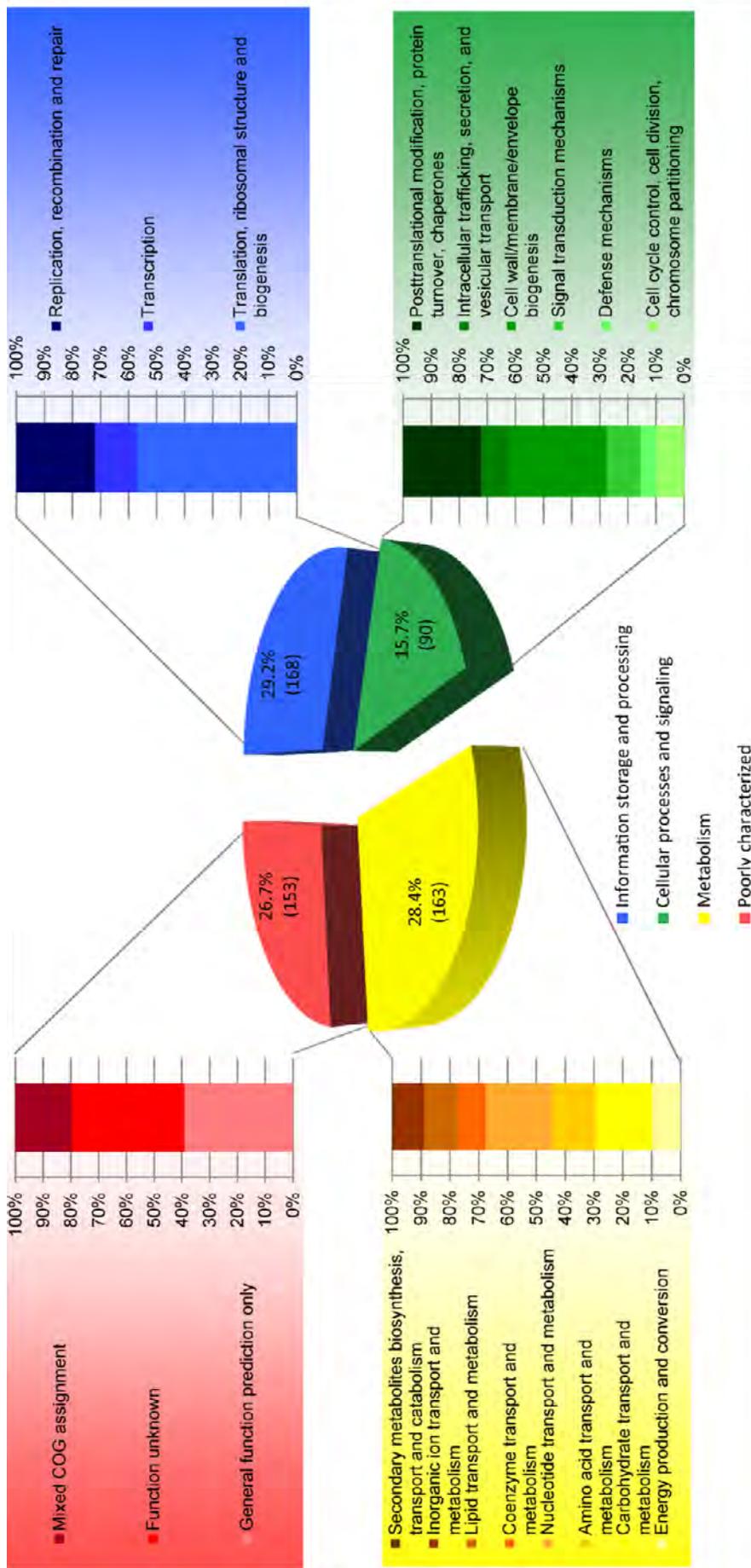


Figure 5.1. Distribution of COG classifications for 574 OG in the core *Streptococcus* genome

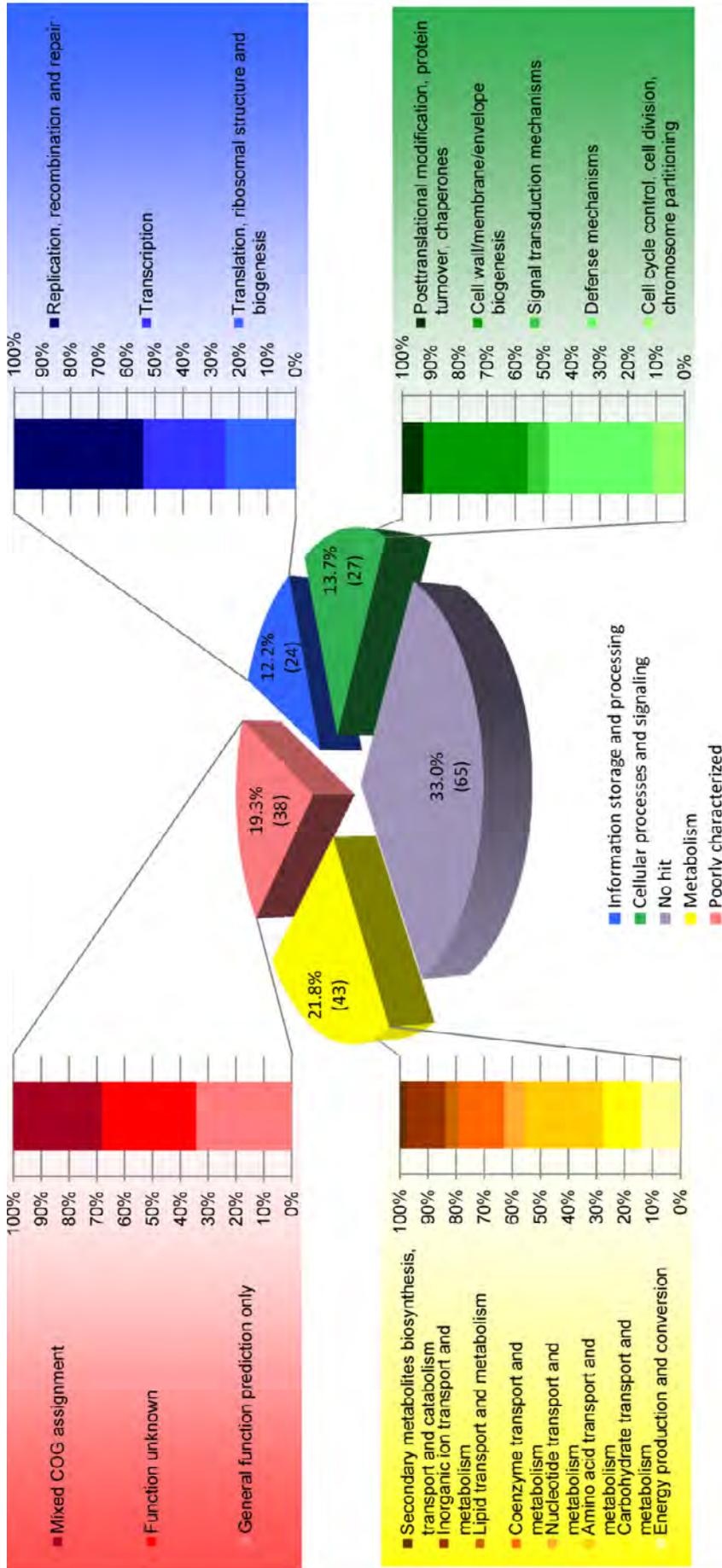


Figure 5.2. Distribution of COG classifications for 197 OG represented in 2-6 small-intestinal genomes

Moreover, the distance between the strains from *S. salivarius* lineage 1 and 4 was relatively small (Figure 5.3), which is in agreement with similarity of fermentation and growth ((340); Chapter 4) as well as the immunostimulatory profiles (Chapter 6) that were determined for these strains. To further assess the similarity between *S. salivarius* lineage 1 and 4, the number of shared genes was determined. This revealed that both lineages shared 1730 OG, which is high compared to the number of shared genes between the two other *S. salivarius* lineages (Table S5.5). Nonetheless, *S. salivarius* lineage 1 and 4 were still predicted to have 128 and 237 strain-specific OG, respectively. The set of 128 lineage 1 specific OGs were manually inspected for potential sequencing and / or gene-calling artifacts (see table S5.6). These genome sequence analyses confirmed that the representative isolates of *S. salivarius* lineages 1 and 4 are closely related, and confirmed and extended our previous observations based on AFLP and Rep-PCR fingerprinting ((340); Chapter 4).

Genomic mining to decipher environmental interaction potential

To obtain an impression how the analysed *Streptococcus* strains may react to external stimuli, we mined their genomes for the canonical two-component system (TCS) regulatory modules, consisting of a HK and a RR, that are known to play a prominent role in bacterial interaction with their environment (307). The strains appeared to encode 12-18 HPK/RR) pairs, which are predicted to respond to a wide variety of environmental responses (Table S5.7). TCS annotated as CiaRH, ComDE, VraSR, and CsrSR were identified in all strains.

The CiaRH system responds to environmental Ca^{2+} (103) and has been shown to be important for intracellular survival of group B *Streptococcus* (263). It has been shown that this system is involved in regulation of numerous functions in *S. pneumoniae*, including those associated with natural competence, which is a driver of evolution (126, 225). Analogously, the *comCDE* encoded TCS (ComDE), present in the *S. mitis* group species (126), has been shown to be the central regulatory module in the control of natural competence, involving a *comC* encoded extracellular competence stimulatory peptide (CSP) as its autoregulatory environmental cue (224). The small intestinal *S. mitis* strain appeared to encode two candidate *comDE* TCS, but a putative CSP encoding *comC* gene upstream of *comDE* could not be identified. The *S. bovis* and *S. salivarius* strains from the small-intestine were found to encode a distinct competence regulatory module consisting of a transcriptional regulator and a putative oligopeptide pheromone, that share similarity with ComR and ComS (97, 126) and are genetically linked to conserved *comX* promoter structures (225). The oligopeptide predicted for the *S. bovis* strain (MKVFSILLTGWWLG) contains the conserved double-tryptophan (WW) motif, which is a conserved feature of ComS from bovis streptococci (225).

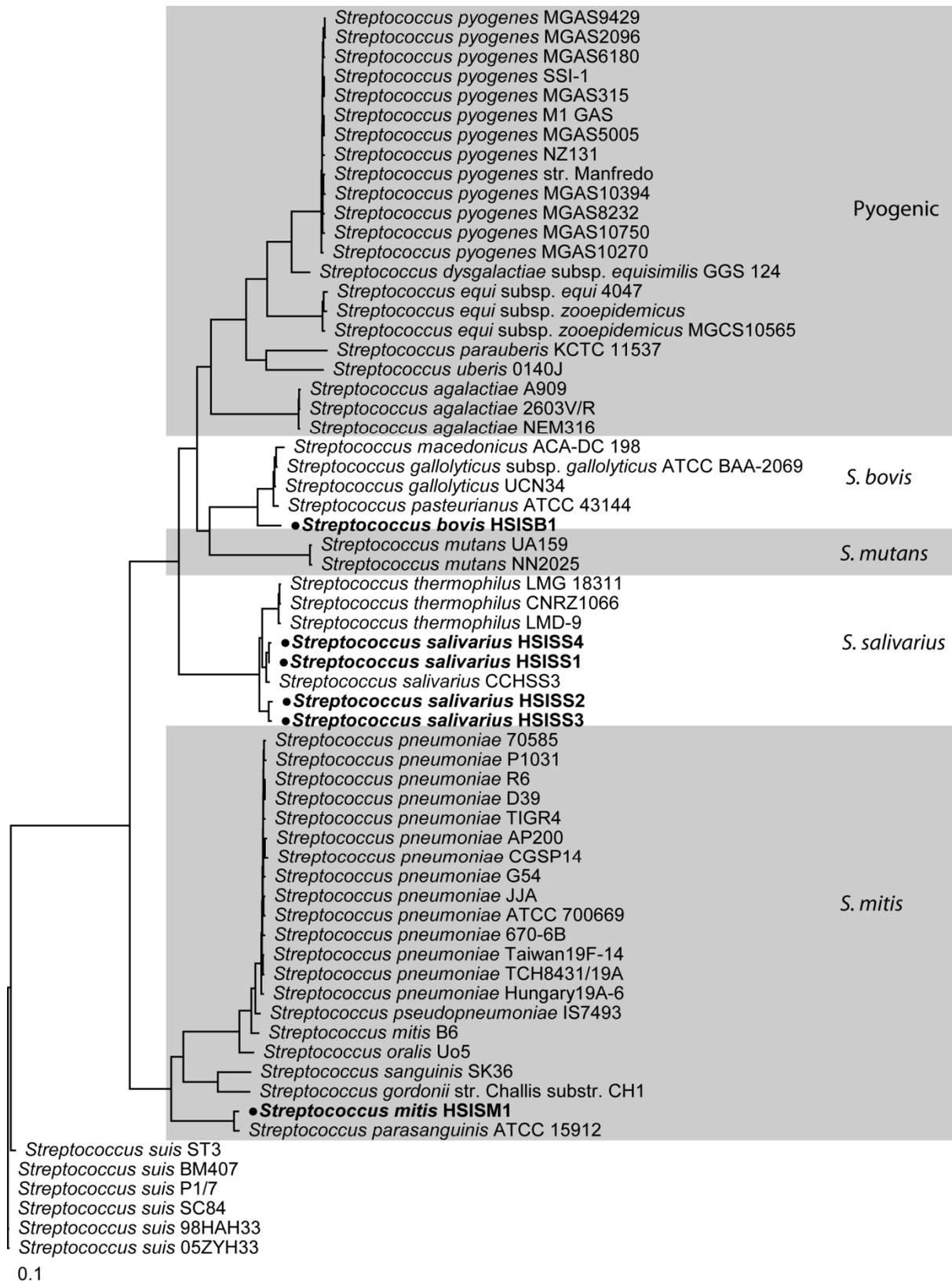


Figure 5.3. *Streptococcus* phylogenetic tree. Unrooted maximum-likelihood phylogenetic tree based on multiple protein sequence alignments (length 5605 residues) of the 450 orthologous groups with exactly one member in 64 *Streptococcus* genomes. Small-intestinal *Streptococcus* strains are highlighted and bulleted.

The oligopeptides predicted for the strains from *S. salivarius* lineage 1 and 4 are identical (MKKLLKFTLFSLLITILPYFAGCL) and resemble that of *S. salivarius* SK126 (97, 225), albeit that the lineage 1 representing strain appears to contain a frameshift in the region encoding the N-terminal end of the oligopeptide. The oligopeptides predicted for the salivarius strains from lineage 2 and 3 are also identical, but are distinct (MKNLRKFLVLLIAAAPFFIYY) from the sequence presented above. It is likely that competence could be induced in these strains via extracellular addition of the unmodified small peptides, especially since all genomes presented here appear to encode a complete competence regulon including genes encoding a competence specific sigma factor ComX, and late competence complexes (e.g. *comEA/C*, and *comGA/B/C/D/E/F/G*), which are involved in DNA uptake and DNA processing (see (126) for review).

The genomes of all the small-intestinal *Streptococcus* strains described here, appeared to encode a TCS that resembles the NisK-NisR and/or SpaK-SpaR TCS modules involved in quorum-sensing controlled autoregulation of nisin and subtilin biosynthesis in *Lactococcus lactis* and *Bacillus subtilis*, respectively (for a review see (174)). Both nisin and subtilin are antimicrobial peptides (bacteriocins) that contain extensive post-translational modification and belong to the class of the lantibiotics (for a review see (342)) and their biosynthesis depends on multi-gene clusters encoding modification, export, immunity, and the mentioned TCS functions (174). To investigate whether the identified streptococcal homologues of these lantibiotic TCS may be involved in regulation of lantibiotic production by these strains, the genetic context of the TCS encoding genes was investigated. However, this analysis failed to identify additional genes that were predicted to be involved in lantibiotic biosynthesis in these organisms. To perform a genome wide analysis of the capacity to produce antimicrobial peptides, we employed the BAGEL2 software module (62) that identified at least one putative bacteriocin encoding gene in the genomes of the *S. bovis* and *S. salivarius* strains. All candidate genes belonged to the non-lanthionine-containing bacteriocins of the pediocin-like (class IIA) and/or miscellaneous (class IID) class according to the scheme proposed by Cotter, et al. (59) (Table S5.8). This analysis indicates that despite their resemblance to NisRK-like TCS modules, these TCS systems are not involved in regulation of genetically linked or distantly located lantibiotic encoding gene clusters, and are thus most likely involved in regulation of other functions.

All *Streptococcus* strains analysed here appeared to encode the CsrSR system, although *S. mitis* appeared to lack a HK paired to the RR similar to CsrR. The CsrSR TCS module is known to play a major role in regulating the virulence of group A and B streptococci (191, 325). Group A streptococcal CsrSR regulates the expression of virulence factors (e.g. pyrogenic exotoxin A, DNase, streptolysin O, streptokinase, and hyaluronic acid capsule synthesis) depending on environmental Mg²⁺, as well as human antimicrobial peptide LL-37 concentrations (325). The CsrSR TCS in group B streptococci is known to repress the expression of certain genes (e.g. coding for β -haemolysin and secreted adhesins) while it stimulates expression of other genes (e.g. the *cps* operon coding for capsular polysaccharide (191)). All newly sequenced

Streptococcus genomes were predicted to encode genes with similarity to hemolysin III. However, only the *S. mitis* and *S. bovis* strains displayed partial (α) hemolysis and none displayed complete (β) hemolysis of blood cells when grown on blood agar (data not shown). Except for the *S. bovis* genome, the other streptococcal genomes were predicted to encode capsular polysaccharide biosynthesis. All *S. salivarius* strains appeared to encode a gene similar to exfoliative exotoxin B, and the strains representing *S. salivarius* lineage 1 and 4 also contained a gene with homology to the C5a peptidase precursor. The latter enzyme inactivates C5a, a chemotactic attractant of phagocytes to infection sites, and promotes streptococcal invasion (45, 360). Although the strains described here are not known to be virulent, they appear to encode at least remnants of the virulence genes known in related streptococci, which may be regulated by the conserved CsrSR TCS module, analogous to what is observed for group A and B streptococci. Remnants of virulence related genes were also encountered in the genomes of strains of *S. themophilus* (28), suggesting that these benign streptococci share specific functions with their known pathogenic relatives.

The VraSR TCS that appeared to be encoded by all streptococcal genomes reported here, has been extensively studied in *Staphylococcus aureus* where it belongs to the cell-wall-stress stimulon that is involved in maintenance of cell wall integrity under stress conditions (334). In *S. aureus* VraSR plays an important role in regulation of resistance to antibiotics that target the bacterial cell wall biosynthesis pathway. Whether the VraSR homologues in the small-intestinal streptococci play a similar role in cell-wall stress and possible antibiotic resistance control remains to be established.

Amino acid and vitamin requirements

The predicted enzyme functions of the newly sequenced *Streptococcus* genomes were mapped onto KEGG pathways to assess their predicted potential for amino acid biosynthesis. Each of the genomes was predicted to encode biosynthesis pathways for at least 18 amino acids (Table 5.1; Figure S5.1). However, none of the strains found to encode the enzymes required to synthesize lysine. Moreover, the biosynthesis of histidine from the pentose phosphate pathway intermediate phosphoribosyl pyrophosphate (PRPP) appears to be incomplete in the genomes of the *S. salivarius* strains representing lineage 2 and 3, as well as the *S. mitis* strain. Alanine biosynthesis appeared to depend on distinct enzymatic conversion of pyruvate to alanine, involving alanine transaminase dehydrogenase (EC 2.6.1.2) in the *S. bovis* strain and the *S. salivarius* strain representing lineage 3, while involving an alanine dehydrogenase (EC 1.4.1.1) in all other strains (Figure S5.1).

Although the small-intestinal streptococci encode the capacity for synthesis of the majority of the amino acids, they also were predicted to encode the oligopeptide import system, *oppABCDF* (70), but lacked a gene resembling an extracellular protease function (e.g. PrtP; (187)). These findings may reflect the adaptation to the peptide and exogenous protease-rich environment that is probably encountered in the human small intestine.

Table 5.1. Predicted amino acid requirements for growth of newly sequenced *Streptococcus* strains

Amino acid	<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>			
			1	2	3	4
Arginine	+	+	+	+	+	+
Histidine	-	+	+	-	-	+
Lysine	-	-	-	-	-	-
Aspartate	+	+	+	+	+	+
Glutamate	+	+	+	+	+	+
Serine	+	+	+	+	+	+
Threonine	+	+	+	+	+	+
Asparagine	+	+	+	+	+	+
Glutamine	+	+	+	+	+	+
Cysteine	+	+	+	+	+	+
Glycine	+	+	+	+	+	+
Proline	+	+	+	+	+	+
Alanine	+	+	+	+	+	+
Valine	+	+	+	+	+	+
Isoleucine	+	+	+	+	+	+
Leucine	+	+	+	+	+	+
Methionine	+	+	+	+	+	+
Phenylalanine	+	+	+	+	+	+
Tyrosine	+	+	+	+	+	+
Tryptophan	+	+	+	+	+	+

Black: required for growth; White: not required for growth

Next we investigated the predicted capacity to produce B-vitamins, which is known to be variable among streptococci (273). Genome analyses indicate that all small-intestine derived streptococci presented here encode the capacity to produce folate from phenylalanine. All strains, except *S. mitis*, also appeared to encode a complete pyridoxal-5-phosphate (B6) biosynthetic pathway. In addition, *S. bovis* was predicted to also encode the capacity to synthesize riboflavin (B2), nicotinate (B3), and pantothenate (B5), which appear to be lacking in *S. mitis* and *S. salivarius*. None of the strains is predicted to encode thiamine (B1), biotin (B8) and cobalamin (B12) biosynthesis pathways.

Primary carbon metabolism and pyruvate dissipation

As streptococci belong to the facultative heterofermentative LAB and generate energy through homolactic and mixed acid fermentation (163), we screened the genomes of the small-intestinal streptococci for genes involved in glycolysis and the pentose phosphate pathway. All strains encoded the required enzymes for glycolytic conversion of glucose to pyruvate (Table S5.9). Notably, only the *S. mitis* strain appeared to encode a complete and intact pentose phosphate pathway. However, the *S. bovis* and *S. salivarius* strains, appeared to code for a transketolase (EC 2.2.1.1) that interconnects the glycolysis and the pentose phosphate pathway, enabling the synthesis of the precursor required in *de novo* purine and pyrimidine synthesis, phosphoribosyl pyrophosphate (PRPP; Figure S5.2). In addition, the *S. bovis* strain, codes for a putative xylulose-5-phosphate phosphoketolase (EC 4.1.2.9;

Table S5.9 and figure S5.2), suggesting that this strain can ferment pentoses (e.g. arabinose ((340); Chapter 4); see below) that enter the pentose phosphate pathway as xylulose-5-phosphate. As expected, genomic analyses showed that none of the small-intestinal streptococci code for a complete tricarboxylic acid (TCA) cycle, albeit that several enzymes (e.g. EC 1.3.99.1, Fumarate reductase) from this pathway are predicted in the genome annotations. The presence of fumarate reductase in the genomes may indicate that these streptococci possess a rudimentary electron transport chain, similar to what is observed for other LAB, including *L. plantarum* WCFS1 (175).

As expected, all the small-intestinal *Streptococcus* genomes have the necessary enzymes to convert pyruvate to L-lactate (lactate dehydrogenase [EC 1.1.1.27]). Although the genomes appeared to lack the genes to produce a complete pyruvate dehydrogenase complex, they do encode the necessary enzymes for mixed acid fermentation via the formate lyase (EC 2.3.1.54), phosphate acetyltransferase (EC 2.3.1.8), and acetate kinase (EC 2.7.2.1) pathway. In addition, the genomes also encompass acetaldehyde dehydrogenase (EC 1.2.1.10) and alcohol dehydrogenase (EC 1.1.1.1) encoding genes, implying their capacity to produce ethanol (Figure S5.2). Finally, all streptococci appeared to encode both acetolactate synthase (EC 2.2.1.6) and acetolactate decarboxylase (EC 4.1.1.5) that could catalyze the conversion of pyruvate to acetoin.

Sugar metabolism

Streptococcus spp. have been proposed to contribute to microbial uptake and fermentation of the simple dietary carbohydrates in the small intestine (381). Therefore, we especially focused our genome annotation efforts on the strain-specific predictions of carbohydrate transport functions and metabolism that can be used as fuel for the downstream energy-generating pathways (e.g. glycolysis and pentose phosphate pathway). All small-intestinal strains encode the general cytoplasmic enzyme I (EI) and phosphor-carrier protein (HPr; Figure 5.4) involved in phospho-donation to several PTS transport systems. In total, 11 distinct PTS transporter functions were found to be encoded by the small-intestinal *Streptococcus* genomes. Those with predicted specificities for glucose/maltose, mannose, fructose, sucrose, β -glucosides, and trehalose were redundantly present in some of the genomes (Figure 5.4).

All sequenced genomes were predicted to encode complete enzyme II (EII) PTS complexes, consisting of IIa, IIb, and IIc (as well as IId in some cases) components, involved in import of glucose/maltose, mannose, fructose, and sucrose (Figure 5.4). Complete PTS with predicted specificity for cellobiose, β -glucosides, and trehalose appeared to be encoded in the *S. bovis* genome, while the PTS-mediated import capacity for the latter two substrates was also predicted for the lineage 1 and 4 representing *S. salivarius* isolates.

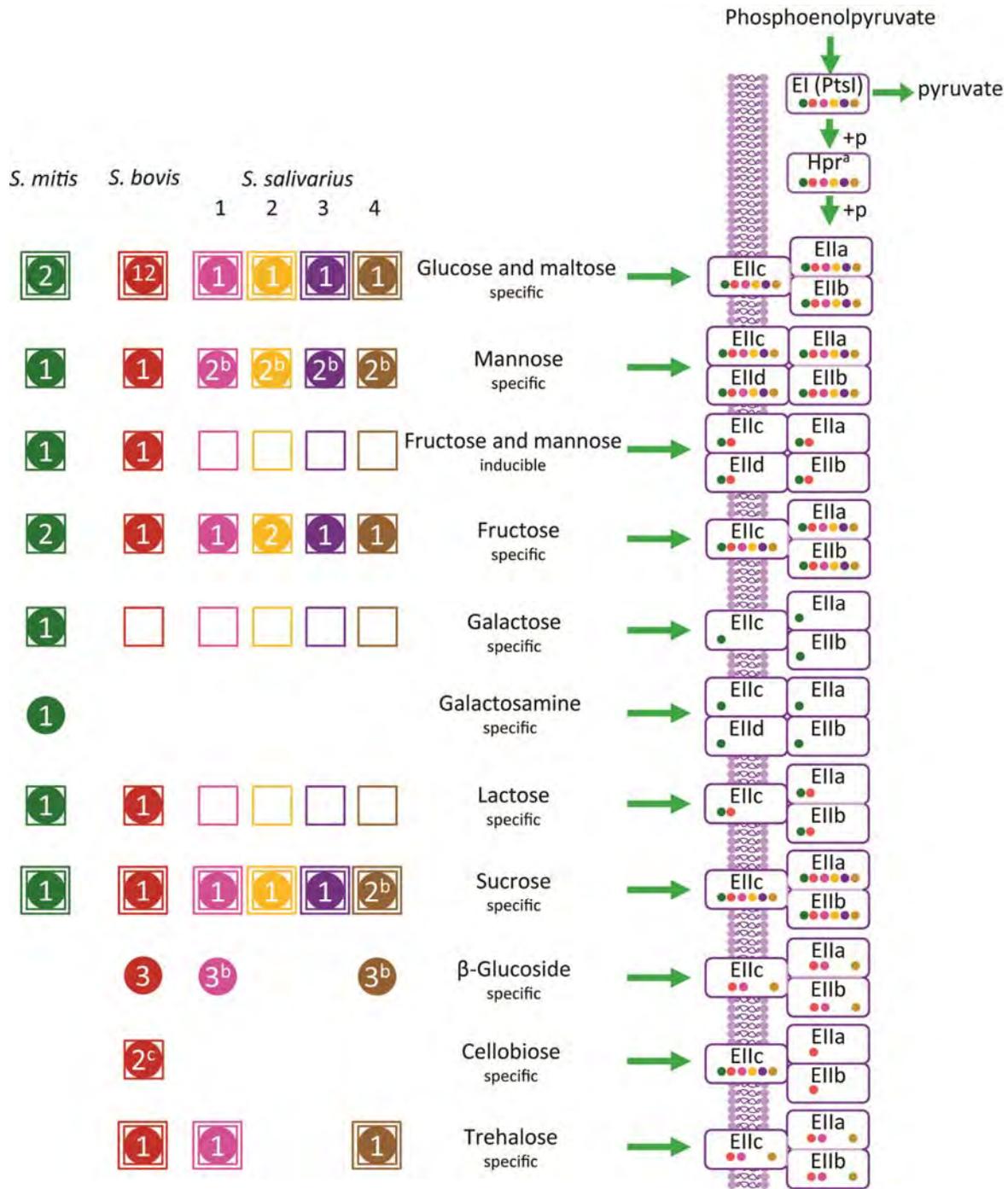


Figure 5.4. Overview of PTS in six sequenced genomes from small-intestinal *Streptococcus* strains. Dots indicate for which genome the corresponding PTS (component) was encoded. Numbers in dots represent the number of times a complete PTS complex was encoded in the genome. Squared dots indicate which isolates were able to ferment the corresponding substrate, double squared dots indicate which isolates were able to ferment and grow on the corresponding substrate ((340); Chapter 4). Components of PTS that are encoded by the same orthologous group are indicated with faded purple lines

^a: *S. mitis* and *S. salivarius* lineage 1 carried an additional phosphocarrier protein (Hpr), each belonging to different orthologous compared to the Hpr protein that belonged to the same OG that was shared between all *Streptococcus*

^b: complete PTS were encoded by genes belonging to different OG

^c: Component EIIa and EIIb from 2 complete cellobiose specific PTS were encoded by genes belonging to the same OG while component EIIc of the two systems belonged to different OG.

Next to these complete PTS EII complexes, all genomes also appeared to encode orphan PTS EIIc transport component(s), which lack the accompanying EIIb and EIIa encoding genes, and in all cases were predicted to have a cellobiose substrate-specificity. This is a feature that has been recognized in many other bacterial genomes, and has been proposed to play a role in environmental signalling (107, 175, 182, 297). The *S. mitis* and *S. bovis* genomes are predicted to encode β -glucosidases (EC 3.2.1.21) that are required for the conversion of cellobiose to β -D-glucose (figure S5.2). However, only *S. bovis* is able to ferment cellobiose ((340); Chapter 4), which appears to be in agreement with the presence of genes encoding a complete cellobiose PTS, which was exclusively encountered in the *S. bovis* genome (Figure 5.4).

These genome-based predictions are in good agreement with the differential carbohydrate-fermentation and growth patterns that were previously determined ((340); Chapter 4). However, all *S. salivarius* genomes lack the PTS for lactose, while all strains were able to ferment this substrate ((340); Chapter 4). It is known that *S. salivarius*, and its close relative *S. thermophilus*, can effectively ferment lactose and import this substrate by a dedicated lactose permease (LacS) that belongs to the galactoside-pentose-hexuronide translocator family (150, 204). Indeed, all *S. salivarius*, and *S. bovis* genomes presented here appeared to encode a *lacS* homologue. The *S. mitis* strain also encode a complete PTS EII complex predicted to be involved in import of galactose and galactosamine. To metabolize lactose and its galactose moiety, *S. mitis* and *S. bovis* encode the tagatose phosphate and/or Leloir pathways (350). The *S. salivarius* lineages, lacking the PTS for lactose, relied on the latter pathway for metabolism of these substrates (Figure S5.2).

All sequenced genomes encode maltose/maltodextrin ABC transporters, while *S. mitis* and *S. bovis* also appear to encode ABC transporters for multiple sugars (the so-called MSM system). These transporters have been previously described for *S. mutans* and can import multiple sugars, including raffinose and melibiose (281).

Notably, the *S. mitis* and *S. bovis* strains, also encode the downstream enzymes required for raffinose and melibiose metabolism. Finally, the *S. bovis* genome also contained genes encoding an “ α -arabinosides ABC transport permease (*araP*)” as well as the necessary enzymes to metabolize this sugar. These findings are in excellent agreement with the observation that only *S. bovis* was able to grow on arabinose, while only *S. mitis* and *S. bovis* could ferment melibiose and grow on raffinose ((340); Chapter 4).

Though all *Streptococcus* strains encoded at least one α -amylase, *S. bovis* could grow on media with starch as the sole carbohydrate source ((340); Chapter 4). However, this could be explained by the predicted subcellular location of the α -amylase enzymes. The enzymes encoded by *S. mitis* and *S. salivarius* genomes were all predicted to be cytoplasmic, while only the *S. bovis* strain appeared to encode an excreted α -amylase (as well as a cytoplasmic one). Thereby, it seems likely that only *S. bovis* can access extracellular starch as a substrate for growth, whereas *S. mitis* and the *S. salivarius* strains may use intracellular polysaccharides (IPS) for energy-storage. The latter is supported by the fact the *S. mitis* and *S.*

salivarius strains encoded three key enzymes required for IPS synthesis, namely glycogen synthase (EC 2.4.1.21), glucose-1-phosphate adenylyltransferase (EC 2.7.7.27), and branching enzyme (EC 2.4.1.18) (36). However, to the best of our knowledge there is no experimental evidence to support the capacity for IPS synthesis in any of the streptococcal species studied here, which may suggest that this only occurs under specific circumstances that were not studied to date.

Since the observed physiological characteristics of the *Streptococcus* strains are in excellent agreement with genome predictions ((340); Chapter 4), the small-intestinal *Streptococcus* genomes are helpful to predict the effect of dietary changes on the *Streptococcus* populations in the small intestine.

Detection of small-intestinal streptococcal strains and lineages

To monitor the dynamics of the small-intestinal *Streptococcus* populations during dietary intervention studies, fast and high throughput, PCR-based detection assays were developed to target small-intestinal *Streptococcus* strain-specific or lineage-specific genes.

To evaluate the specificity and conservation of the genes selected for strain-specific detection among the representative isolates that belong to the different streptococcal lineages, they were amplified from 92 *Streptococcus* isolates and 28 *Enterococcus* isolates as negative control ((340); Chapter 4). The PCR amplicons robustly discriminated the 92 *Streptococcus* isolates into 3 *S. mitis* isolates, 13 *S. bovis* isolates, 1 *S. salivarius* lineage 2 isolate, and 8 *S. salivarius* lineage 3 isolates, thereby perfectly matching with their grouping according to AFLP and Rep-PCR (Table 5.2 and S5.10; (340); Chapter 4).

Table 5.2. Number of positive PCR amplifications of bacterial isolates with *Streptococcus* lineage specific primers

Grouping	<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>				<i>Enterococcus</i>
			1	2	3	4	
AFLP and Rep-PCR analysis*	3	13	1	1	8	66	28
PCR assay	<i>S. mitis</i>	3	-	-	-	-	-
	<i>S. bovis</i>	-	13	-	-	-	-
	<i>S. salivarius</i> 1	-	-	1	-	-	56
	<i>S. salivarius</i> 2	-	-	-	1	-	-
	<i>S. salivarius</i> 3	-	-	-	-	8	-
	<i>S. salivarius</i> 4	-	-	1	-	-	66

*: Results from grouping according to AFLP and Rep-PCR ((340); Chapter 4)

These results showed that among representative strains of the identified lineages, the selected gene is conserved and specific for the isolates of that lineage. In contrast, the primers designed to selectively amplify *S. salivarius* strain 1 and 4, failed to consistently discriminate between isolates of these two lineages, albeit that not every bacterial isolate belonging to *S. salivarius* lineage 4 revealed a PCR product with primers for *S. salivarius* lineage 1. Nonetheless, these results confirm the close relatedness of these lineages, which was already apparent from the

identical Rep-PCR profiles they generated (see above; (340); Chapter 4). The *Streptococcus* PCR assays developed here provide a simple and effective means to detect the small-intestinal *S. mitis*, *S. bovis*, and *S. salivarius* lineages 2, 3 and the group of *S. salivarius* lineage 1 and 4.

All six *Streptococcus* strains from the small intestine were cultivated from a single ileostoma effluent sample. Therefore, the PCR assays were further evaluated with total DNA from 30 ileostoma effluent samples obtained from 6 ileostomists, 4 ileal fluid samples from 3 healthy individuals, and fecal samples from 10 healthy individuals to investigate the distribution of these genetic targets beyond the ileostomist they were derived from. As anticipated, the selected genes from all sequenced *Streptococcus* genomes were amplified in other ileostoma effluent samples collected from the ileostomist from which the strains were isolated (Subject A; Table 5.3). However, the *S. bovis* specific amplicon could only be detected in a single ileostoma effluent sample (Subject A) and in several fecal samples (Table 5.3). The *S. mitis* specific amplicon was exclusively detected in ileostoma samples, which were obtained from subject A and subject B. The latter samples were collected on two consecutive days, while additional samples that were collected from the same individual 5 years later did not allow the detection of this genetic marker. Considering that the *mitis* and, to a lesser extent, *bovis* group streptococci belong to the predominant streptococci in the small intestine in these samples ((340); Chapter 4), these findings imply that the small-intestinal microbiota in other individuals is encompassing other *S. mitis* and *S. bovis* lineages as compared to the strain targeted here. In contrast, the unique genes from at least 2 *S. salivarius* lineages (mostly lineage 2 and 4) were detected in all but one ileostoma effluent sample, all ileal fluid samples as well as several fecal samples obtained from other individuals (Table 5.3). These findings suggest that the *S. salivarius* lineages are highly conserved in the small intestine and, to a lesser extent in the terminal part of the gastrointestinal tract, among the different subjects.

Discussion

In this study, the genomes of six small-intestinal *Streptococcus* isolates were determined by next generation sequencing technologies and were compared with *Streptococcus* genomes from the public databases.

Phylogenetic analysis of the small-intestinal *Streptococcus* genomes placed one strain into the *S. mitis* species group, one strain into the *S. bovis* species group, and four into the *S. salivarius* species group, matching species identifications that were previously based on the 16S rRNA gene sequence alone ((340); Chapter 4). The genomes of two of the *S. salivarius* strains (lineage 1 and 4) were highly similar, which was expected based on their highly similar genetic typing profiles and their conserved physiological characteristics ((340); Chapter 4).

The *Streptococcus* pangenome consisted of 12,403 orthologous DNA sequences, which is double the size predicted by Lefébure and Stanhope based on 26 *Streptococcus* finished and whole genome shotgun genomes (197). However, this genome set was represented by 6 species while the current study included as many as 20 different species. The core *Streptococcus* genome was defined here as a set of 574 OG shared by all *Streptococcus* genomes, which is in line with earlier predictions (197). Analysis of the core *Streptococcus* OG revealed that the function of most genes was well defined and belonged to typically conserved cellular processes like transcription, translation and replication. Nonetheless, a significant portion of core orthologous groups were involved in metabolism, especially transport and metabolism of nucleotides and carbohydrates. Likewise, a group of 197 OG only common among the streptococci analysed here was mostly involved in metabolism and may contain streptococcal genes that contribute specifically to the lifestyle of these bacteria in the (human small) intestine.

Mining of the genomes revealed that the small-intestinal streptococci coded for two-component regulatory modules, such as those involved in natural competence. Since natural competence is a mediator for evolution and genomic plasticity (126), we focused on functions that play an important role in this system. We found gene repertoires that imply that the competence regulon in each of the streptococcal genomes analysed here is complete, which implies that these small-intestinal streptococci may become naturally transformable under specific conditions, which could contribute to the high genetic diversity of the streptococcal community in the small intestine ecosystem ((340); Chapter 4).

The small-intestinal *Streptococcus* strains had a predicted capacity to synthesize a large number of amino acids, and were expected to have capabilities to produce B-vitamins. As expected, all streptococci analysed here coded for a complete glycolytic pathway and an intact or part of the pentose phosphate pathway for energy generation, yielding lactate, acetate, formate, and possibly acetoin as fermentation products. However, the strains differed considerably in their predicted capacity to transport and metabolize specific sugars. The *Streptococcus* genomes encoded for a complement of 11 different complete PTS, which in some cases were present in multiple copies in a single strain. This indicates that some substrates may be more important for certain strains. The *S. bovis* genome encoded for 9 different PTS, which was higher compared to the number of PTS encoded by *S. mitis* (8) and the *S. salivarius* strains (4-6). In addition, this strain was also the only sequenced small-intestinal strain that appeared to encode transporters for arabinose and extracellular amylases for the degradation of starch. Nevertheless, the number of PTS was relatively low compared to a closely related strain *S. gallolyticus* UCN34, which encodes 25 PTS (280).

While the *S. bovis* strain encoded for extra- and intracellular α -amylases, the *S. mitis* and *S. salivarius* strain were found to only code for those that remain intracellular, which have been postulated to play a role in breakdown of IPS (361). However, investigations into α -amylase of *S. mutans* revealed that intracellular α -amylase was not essential for breakdown of IPS and dextrans from starch digested by exogenous

α -amylase (299). Therefore, the role of intracellular α -amylases remains to be elucidated.

The encoded carbohydrate transporters, and the reconstructions of the metabolic pathways based on genome analysis were in excellent agreement with physiological characteristics that were determined previously ((340); Chapter 4). The variation between their metabolic capacities may explain why they coexist in a harsh and fluctuating environment such as the small intestine. It also suggests that the small-intestinal genomes can accurately predict the carbohydrate utilization capacities of these bacterial strains. This may be of value in studies determining the effect of food components on the small-intestinal microbiota *in situ* with a special focus on these *Streptococcus* populations.

One prerequisite to this concept is to detect the *Streptococcus* lineages, using for example unique genes as genetic markers, in intestinal samples. To this end, PCR-based screening assays were designed for each of the small-intestinal streptococci and tested with 92 *Streptococcus* isolates. These assays correctly amplified isolates belonging to the same lineage as their target *Streptococcus* strain, based on strain-level groupings as was done with AFLP and Rep-PCR analysis ((340); Chapter 4). Although primer assays for *S. salivarius* lineage 1 and 4 isolates were developed using strain specific genes and employing strict primer design parameters to ensure primer specificity, both primer sets showed cross-reaction with isolates belonging to the non-target *S. salivarius* lineage. Determining the exact causes for this is not trivial and are likely related to the causes that potentially account for the inaccurate estimation of strain-specific genes (Table S5.6). Nonetheless, the *Streptococcus* PCR assays developed here provide a simple and rapid method for the screening of large numbers of samples from, for example, dietary intervention studies, for the unique genes of the small-intestinal *S. mitis*, *S. bovis*, and *S. salivarius* strains or lineages.

Application of the assays on 34 intestinal and 10 fecal samples collected from 19 human individuals revealed that at least two *S. salivarius* lineages were present in almost all small-intestinal samples and several fecal samples, indicating that these strains are ubiquitously present and represent an important population of, in particular, the small-intestinal microbiota. Only one ileostoma effluent sample showed no amplification within any of the assays. However, the *Streptococcus* population in this sample is most likely represented by one or more *Streptococcus* strains that do not carry the unique genes targeted by PCR-based detection assays.

In conclusion, the work presented here describes a comparative genomics study of *Streptococcus* spp. that focused on strains from the human small intestine. Comparative genomic analysis revealed that the small-intestinal strains differed in their predicted transport and metabolism of sugars, which was in agreement with physiological data. Therefore, the small-intestinal *Streptococcus* genomes are useful to construct metabolic models to predict the effect of different dietary substances on *Streptococcus* populations dynamics in the human small intestine. Furthermore, assays designed for detection of two *S. salivarius* strains were positive for most of the small-intestinal samples from different individuals, suggesting that strains,

carrying the target functional gene, represent an important population of the small-intestinal ecosystem.

Acknowledgements

This project was partially supported by the Netherlands Bioinformatics Centre (NBIC). We thank Sebastian Tims for providing fecal sample DNA, Maria Stolaki for her help with performing the PCR-based screening assays, and Roland Siezen for critical reading of the manuscript. We appreciate the help of Christopher Bauser and Julia Löcherbach of GATC-Biotech (Konstanz, Germany) for assistance in the set-up of genome sequencing.

Supplementary tables

Table S5.1. Genome statistics for small intestinal *Streptococcus*

	<i>S. mitis</i>				<i>S. bovis</i>				<i>S. salivarius</i>			
	HSISM1	ASKA000000000	HSISB1	ASKB000000000	HSISS1	ASKC000000000	HSISS2	ASKH000000000	HSISS3	ASKD000000000		
Locus tag prefix												
Accession*	118	42	41	61	170	61	151	151	151	151		
Total number of Contigs	2,219,316	1,864,835	2,074,878	2,218,190	2,103,121	2,218,190	2,034,088	2,034,088	2,034,088	2,034,088		
Contig Sum (bp)	75,679	228,131	251,655	88,297	88,297	185,834	79,375	79,375	79,375	79,375		
Max contig size (bp)	1,129	1,086	1,135	1,046	1,046	1,187	1,004	1,004	1,004	1,004		
Min contig size (bp)	18,807	44,400	50,606	12,371	12,371	36,363	13,470	13,470	13,470	13,470		
Average contig size (bp)	26,412	101,212	86,157	23,692	23,692	62,198	23,226	23,226	23,226	23,226		
contig N50	7	6	6	15	15	8	7	7	7	7		
Total number of scaffolds	2,334,368	1,906,608	2,127,962	2,360,589	2,360,589	2,399,642	2,208,371	2,208,371	2,208,371	2,208,371		
Scaffold Sum (bp)	1,227,939	1,486,018	1,616,043	702,120	702,120	1,081,630	957,480	957,480	957,480	957,480		
Max scaffold size (bp)	3,313	9,467	1,135	1,217	1,217	43,206	1,004	1,004	1,004	1,004		
Min scaffold size (bp)	333,481	317,768	354,660	157,372	157,372	299,955	315,481	315,481	315,481	315,481		
Average scaffold size (bp)	1,227,939	1,486,018	1,616,043	259,673	259,673	516,471	763,578	763,578	763,578	763,578		
Scaffold N50	41.45	37.24	40.12	39.84	39.84	39.31	40.23	40.23	40.23	40.23		
GC content (%)	2,226	2,160	1,939	2,247	2,247	2,230	2,091	2,091	2,091	2,091		
Number of predicted proteins	1828	1928	1709	1862	1862	1800	1828	1828	1828	1828		
Genes assigned to COG	1094	1117	1106	1082	1082	1110	1086	1086	1086	1086		
Genes assigned to KEGG												

*: The version described in this chapter is version XXXX01000000

Table S5.2. Characteristics of finished *Streptococcus* genomes*

Accession	Organism	Goldcard	Isolation site	Host	Comments
NC_002737	<i>Streptococcus pyogenes</i> M1 GAS	Gc00049	Patient with a wound infection	Homo sapiens	
NC_003028	<i>Streptococcus pneumoniae</i> TIGR4	Gc00058	Blood of a 30 year old male patient in Kongsvinger Norway	Homo sapiens	
NC_003098	<i>Streptococcus pneumoniae</i> R6	Gc00065	-	Homo sapiens	
NC_003485	<i>Streptococcus pyogenes</i> MGAS8232	Gc00081	throat swab from patient with acute rheumatic fever	Homo sapiens	
NC_004070	<i>Streptococcus pyogenes</i> MGAS315	Gc00094	patient with streptococcal toxic shock syndrome	Homo sapiens	
NC_004116	<i>Streptococcus agalactiae</i> 2603V/R	GC00098	clinical isolate	Homo sapiens	
NC_004350	<i>Streptococcus mutans</i> UA159	GC00109	Child with active dental caries in 1982	Homo sapiens	
NC_004368	<i>Streptococcus agalactiae</i> NEM316	Gc00100	Case of fatal septicemia	Homo sapiens	
NC_004606	<i>Streptococcus pyogenes</i> SSI-1	Gc00137	Toxic-shock patient in Japan	Homo sapiens	
NC_006086	<i>Streptococcus pyogenes</i> MGAS10394	Gc00205	pharyngeal swab from child, during a study of the epidemiology of pharyngitis in a private elementary school	Homo sapiens	
NC_006448	<i>Streptococcus thermophilus</i> LMG 18311	Gc00234	Commercial yogurt in 1974 in the United Kingdom	-	
NC_006449	<i>Streptococcus thermophilus</i> CNRZ1066	Gc00233	Isolated from yogurt in France	-	
NC_007296	<i>Streptococcus pyogenes</i> MGAS6180	Gc00284	Invasive disease in Texas in 1998	Homo sapiens	
NC_007297	<i>Streptococcus pyogenes</i> MGAS5005	Gc00285	Invasive case in Ontario	Homo sapiens	
NC_007432	<i>Streptococcus agalactiae</i> A909	Gc00302	septic human neonate	Homo sapiens	
NC_008021	<i>Streptococcus pyogenes</i> MGAS9429	Gc00379	pharyngeal swab from animal	Homo sapiens	swab from animal but host name was still homo sapiens
NC_008022	<i>Streptococcus pyogenes</i> MGAS10270	GC00378	pharyngeal swab from child	Homo sapiens	
NC_008023	<i>Streptococcus pyogenes</i> MGAS2096	Gc00377	patient with acute poststreptococcal glomerulonephritis	Homo sapiens	
NC_008024	<i>Streptococcus pyogenes</i> MGAS10750	Gc00376	human pharyngeal swab from patient with pharyngitis	Homo sapiens	
NC_008532	<i>Streptococcus thermophilus</i> LMD-9	Gc00451	-	-	
NC_008533	<i>Streptococcus pneumoniae</i> D39	Gc00437	-	Homo sapiens	
NC_009009	<i>Streptococcus sanguinis</i> SK36	Gc00509	isolated from human dental plaque	Homo sapiens	
NC_009332	<i>Streptococcus pyogenes</i> str. Manfredo	Gc00455	Patient in the 1950's in Chicago	Homo sapiens	
NC_009442	<i>Streptococcus suis</i> 05ZYH33	Gc00546	Chinese virulent strain isolated from fatal cases of STSS in 2005	Homo sapiens	
NC_009443	<i>Streptococcus suis</i> 98HAH33	Gc00547	Chinese virulent strain isolated from fatal cases of STSS in 1998	Homo sapiens	
NC_009785	<i>Streptococcus gordonii</i> str. Challis substr. CH1	Gc00643	-	Homo sapiens	
NC_010380	<i>Streptococcus pneumoniae</i> Hungary19A-6	Gc00735	Human ear, Hungary	Homo sapiens	
NC_010582	<i>Streptococcus pneumoniae</i> CGSP14	Gc00765	clinical isolate derived from a child with necrotizing pneumonia, simultaneously complicated with HUS, at Chang Gung Memorial Hospital and Children's Hospital, Taoyuan, Taiwan.	Homo sapiens	
NC_011072	<i>Streptococcus pneumoniae</i> G54	Gc00837	Genova Italy by G. Schito from a respiratory sample in 1997	Homo sapiens	

Table S5.2 (Continued). Characteristics of finished *Streptococcus* genomes*

Accession	Organism	Goldcard	Isolation site	Host	Comments
NC_011134	<i>Streptococcus equi</i> subsp. <i>zooepidemicus</i> MGCS10565	Gc00845	throat of a patient with nephritis diagnosed during an epidemic in the state of Minas Gerais, Brazil	Homo sapiens	
NC_011375	<i>Streptococcus pyogenes</i> NZ131	Gc00871	Patient with acute glomerulonephritis and was provided by Diana Martin, New Zealand Communicable Diseases Center, Porirua, New Zealand	Homo sapiens	
NC_011900	<i>Streptococcus pneumoniae</i> ATCC 700669	Gc00940	hospital, Barcelona, Spain	Homo sapiens	
NC_012004	<i>Streptococcus uberis</i> 0140J	Gc00948	clinical bovine mastitis case	Bovine	
NC_012466	<i>Streptococcus pneumoniae</i> JJA	Gc00973	-	Homo sapiens	
NC_012467	<i>Streptococcus pneumoniae</i> P1031	Gc00972	-	Homo sapiens	
NC_012468	<i>Streptococcus pneumoniae</i> 70585	Gc00969	-	Homo sapiens	
NC_012469	<i>Streptococcus pneumoniae</i> Taiwan19F-14	Gc00974	cerebrospinal fluid, Taiwan	Homo sapiens	
NC_012470	<i>Streptococcus equi</i> subsp. <i>zooepidemicus</i>				No GOLDCARD available, genbank entry: nasal swab taken from a healthy Thoroughbred racehorse
NC_012471	<i>Streptococcus equi</i> subsp. <i>equi</i> 4047	Gc00971	horse with strangles from New Forest, UK	Horse	
NC_012891	<i>Streptococcus dysgalactiae</i> subsp. <i>equisimilis</i> GGS_124	Gc01051	patients with STSS	Homo sapiens	
NC_012924	<i>Streptococcus suis</i> SC84	Gc01063	case of streptococcal toxic shock-like syndrome in Sichuan Province, China in 2005	Homo sapiens	
NC_012925	<i>Streptococcus suis</i> P1/7	Gc01061	ante-mortem blood culture from a pig dying with meningitis	Sus scrofa, Homo sapiens	
NC_012926	<i>Streptococcus suis</i> BM407	Gc01062	CSF from a human case of meningitis in Ho Chi Minh City, Vietnam in 2004	Homo sapiens	
NC_013798	<i>Streptococcus gallolyticus</i> UCN34				No GOLDCARD available
NC_013853	<i>Streptococcus mitis</i> B6	Gc01214	hospital in Bochum, Germany	Homo sapiens	
NC_013928	<i>Streptococcus mutans</i> NN2025	Gc01074	clinical serotype c strain isolated in Japan in 2002 from a patient with dental caries	Homo sapiens	
NC_014251	<i>Streptococcus pneumoniae</i> TCH8431/19A	Gc01351	respiratory tract	Homo sapiens	
NC_014494	<i>Streptococcus pneumoniae</i> AP200	Gc01425	clinical isolate from the cerebrospinal fluid of a patient with meningitis in Italy in 2003	Homo sapiens	
NC_014498	<i>Streptococcus pneumoniae</i> 670-6B	Gc01424	-	Homo sapiens	
NC_015215	<i>Streptococcus gallolyticus</i> subsp. <i>gallolyticus</i> ATCC BAA-2069	Gc01837	human blood culture; infective endocarditis	Homo sapiens	

Table S5.2 (Continued). Characteristics of finished *Streptococcus* genomes*

Accession	Organism	Goldcard	Isolation site	Host	Comments
NC_015291	<i>Streptococcus oralis</i> Uo5	Gc01712	human mouth	Homo sapiens	
NC_015433	<i>Streptococcus suis</i> ST3	Gc01735	-	Homo sapiens, Sus scrofa	
NC_015558	<i>Streptococcus parauberis</i> KCTC 11537	Gc01743	-	-	
NC_015600	<i>Streptococcus pasteurianus</i> ATCC 43144	Gc01797	human blood	Homo sapiens	
NC_015678	<i>Streptococcus parasanguinis</i> ATCC 15912	Gc01842	human throat	Homo sapiens	
NC_015760	<i>Streptococcus salivarius</i> CCHSS3	Gc01887	Human blood	Homo sapiens	
NC_015875	<i>Streptococcus pseudopneumoniae</i> IS7493	Gc01960	sputum of a patient with human immunodeficiency virus (HIV) who had documented pneumonia	Homo sapiens	
NC_016749	<i>Streptococcus macedonicus</i> ACA-DC 198	Gc02096	traditional greek kasseri cheese	-	

*: Retrieved from the Genome OnLine database (GOLD; <http://genomesonline.org>) on February 27th, 2012

Table S5.3. Primers used in this study

Primer name*	Primer sequence (5'-3')	Annotation of target gene	Expected PCR product
HSISM1_486_fwd	CGCTGCTTTAGAGGCATCAACCG	Hypothetical protein	225
HSISM1_486_rev	ACCTGGCGATCAAGCACAGAGT		
HSISB1_163_fwd	ACTTTGGTGCGTTATCCTGGTGG	Alpha-N-arabinofuranosidase (EC 3.2.1.55)	230
HSISB1_163_rev	TCAACTAAATGGCGGGCTTCGTC		
HSISS1_1164_fwd	ACTGGTTGTTCTGGCTCCTCTGG	Hypothetical protein	223
HSISS1_1164_rev	CGGTCGTACCAGATGTACCAGGC		
HSISS2_1351_fwd	GGTTGGCTTGGTTCTTTACGGGT	Hypothetical protein	230
HSISS2_1351_rev	GGCTCCAAAGCTCGAATGGTTGC		
HSISS3_521_fwd	GCTGAACCAACAAACCTCGCAGA	Hypothetical protein; possible cell wall protein, WapE	286
HSISS3_521_rev	TGGCAACCTCTTGGTCGAGTGCT		
HSISS4_733_fwd	GGCTGAACCTGATCCTCCATTTCG	Hypothetical protein	170
HSISS4_733_rev	AACGAGACGAGTCAAAGGGCTTG		

*: Target strain_locus tag_fwd/rev

Fwd: forward primer, rev: reverse primer, HSISM1 represents the small-intestinal *Streptococcus* genome for the *S. mitis* species group, HSISB1 represents the small-intestinal *Streptococcus* genome for the *S. bovis* species group, and HSISS1, HSISS2, HSISS3, and HSISS4 represents the small-intestinal *Streptococcus* genome from lineage 1, 2, 3, and 4, respectively, from the *S. salivarius* species group.

Table S5.4. Cluster specific orthologous groups

Data file with cluster-specific locus tags for *Streptococcus* clusters is available upon request.

Table S5.5. Number of shared (lower left panel) and unshared (upper right) orthologous genes between *S. salivarius* genomic lineages 1-4

	Lineage 1	Lineage 2	Lineage 3	Lineage 4
Lineage 1	1858 ^b	311	263	128
Lineage 2	1547 ^a (76)	2117 ^b	487	547
Lineage 3	1595 ^a (124)	1630 ^a (159)	2028 ^b	476
Lineage 4	1730 ^a (259)	1570 ^a (99)	1552 ^a (81)	1967 ^b

^a: Numbers in brackets indicate number of shared orthologous genes between two strains without number of shared genes among all *S. salivarius* strains (1471)

^b: Total number of orthologous groups per lineage

Table S5.6. Potential causes for inaccurate estimation of strain specific orthologous genes between *S. salivarius* lineage 1 and 4^a

Number of genes from lineage 1	Description
61	Genes did not have an ortholog, but did contain one or more conserved regions potentially encoded by the lineage 4 genome (either as part of non-orthologous genes, or in a genomic region not predicted to encode protein, including gene fragments introduced by single nucleotide insertion or deletion in lineage 4)
42	Genes were N or C-terminal fragments of frameshifts in orthologous 4 lineage genes introduced by single-nucleotide insertions or deletions in lineage 1
12	Genes were very small (<60 amino acid residues) and had not been recognized as ORFs in lineage 4
8	Genes did not have a significantly similar sequence
5	Genes were encoded in genomic regions for which only partial or no sequence data was available in lineage 4 ^b

^a: Causes that may have led to an overestimation of the strain-specific OG prediction were determined by manual analyses of strain-specific orthologous genes of the strain from *S. salivarius* lineage 1, employing OG-protein sequence detection by tblastn (7) in the genome that was predicted to lack the orthologous genes (lineage 4).

^b: Prediction based on analysis of conserved gene context of neighboring genes

Table S5.7. Number and description of two component systems predicted for small intestinal *Streptococcus* strains.

	<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>				Comment
			1	2	3	4	
Total number of two component systems	14	13	14	12	18	14	
Number of orphan two component system ^a	4	0	3	3	1	3	
CiaRH	x	x	x	x	x	x	See main text
ComDE	x	x	x	x	x	x	See main text
VraSR	x	x	x	x	x	x	See main text
CsrSR	x ^b	x	x	x	x	x	See main text
SA14-24	x	x	x	x	x	x	To the best of our knowledge, no function for this TCS has been proposed
DltR		x	x	x	x	x	involved in regulating incorporation of D-alanine in lipoteichoic acid (260).
BlpH	x						Proposed sensor histidine kinase involved in regulation of genes with potential bacteriocin-like functions (64).
FasB						x	Sensor histidine kinase involved in control of Group A streptococcal virulence factors (183).
KdpD				x			Sensor histidine kinase that promotes resistance to osmotic, oxidative, and antimicrobial stress (6).
LevQRST	x	x					Four component system that consists of the histidine kinase LevS, the response regulator LevR, and two putative extracellular sugarbinding proteins (LevQ and LevT), controlling transcriptional regulation of <i>fruA</i> (373).
NisKR	x	x		x			The chromosomes of <i>S. mitis</i> , <i>S. bovis</i> , and <i>S. salivarius</i> lineage 2 were predicted to code HK and RR controlling regulation of nisin biosynthesis by <i>L. lactis</i> (NisKR; (89, 213)). In addition, the <i>S. salivarius</i> strains also coded for TCS that was similar to SpaKR, which regulates subtilin biosynthesis by <i>Bacillus subtilis</i> (177). Both nisin and subtilin are lantibiotics and their gene clusters not only code for the nisin and subtilin precursors and proteins involved in their post-translational modification, but also contain genes involved in secretion of the modified precursors and immunity (174). However, the chromosomes of the newly sequenced <i>Streptococcus</i> strains did not appear to code for the complete nisin and subtilin biosynthetic gene clusters. These genes may have been deleted during evolution, indicating that the newly sequenced streptococci do not synthesize nisin and/or subtilin, or that the TCSs are involved in regulation of different systems.
SpaKR			x	x	x	x	
VncRS	x		x			x	Two component system playing a role in induction of multiple pathways leading to cell death in <i>S. pneumonia</i> (246). However, this finding was challenged by a study by Haas, et al. showing that this TCS was not essential for cell death, so its function remains to be further elucidated (119).
YesMN	x						To the best of our knowledge, no function for this TCS has been proposed

^a: TCS component coding for either a sensor histidine kinase or response regulator

^b: orphan response regulator

Table S5.8. Candidate bacteriocins identified by BAGEL2

Strain belonging to <i>Streptococcus</i> species group	Locus tag	Peptide size (amino acids)	Product	Class	Score*	
<i>S. mitis</i>	HSISM1_1402	111	Alkylphosphonate utilization operon proteinPhn A		1150	
	HSISM1_2084	105	hypothetical protein		1125	
	HSISM1_1626	51	hypothetical protein	IA	1100	
<i>S. bovis</i>	HSISB1_1269	101	Endonuclease III (EC 4.2.99.18)		1150	
	HSISB1_1297	116	Phenylalanyl-tRNA synthetase beta chain (EC6.1.1.20)		1125	
	HSISB1_1598	53	hypothetical protein	IID	6900	
	HSISB1_1599	50	hypothetical protein		1125	
<i>S. salivarius</i>	1	HSISS1_1504	76	hypothetical protein	IIA or IID	6500
		HSISS1_1495	76	Bacteriocin BIpU	IIA or IID	6175
	2	HSISS1_1572	66	hypothetical protein	IID	1175
		HSISS2_2019	84	pore-forming peptide, putative bacteriocin	IIA	10100
	3	HSISS2_2023	76	Bacteriocin BIpU	IIA	5100
		HSISS3_415	84	pore-forming peptide, putative bacteriocin	IIA or IID	11150
		HSISS3_402	59	hypothetical protein	IIA or IID	6175
		HSISS3_920	73	hypothetical protein	IID	6125
		HSISS3_416	79	hypothetical protein	IIA or IID	6100
		HSISS3_766	73	hypothetical protein	IID	5925
		HSISS3_769	53	hypothetical protein		1475
		HSISS3_767	75	hypothetical protein	IID	1275
		HSISS3_921	75	hypothetical protein	IID	1125
		HSISS3_410	51	hypothetical protein		1100
		HSISS3_923	53	hypothetical protein		1050
		HSISS3_45	183	hypothetical protein		1025
	4	HSISS4_689	105	Cytidine deaminase (EC 3.5.4.5)		1100
		HSISS4_1985	98	Acetyltransferase (EC 2.3.1.-)		1100
		HSISS4_1831	76	Bacteriocin BIpU	IIA	5100
		HSISS4_1840	76	hypothetical protein	IIA	5100

*: Candidates with ≥ 1800 points are highlighted in grey and are considered 'putative bacteriocins' by BAGEL2 (62) while a score below this threshold value but with a score of 1000 are considered 'interesting candidates'.

Table S5.9. Locus tags of enzymes involved in glycolysis and pentose phosphate pathway

Product	EC number	<i>S. mitis</i> (HSISM1)	<i>S. bovis</i> (HSISB1)	<i>S. salivarius</i> group			
				1 (HSISS1)	2 (HSISS2)	3 (HSISS3)	4 (HSISS4)
Glucokinase ^a	2.7.1.2	1367	1465	1788	726	2176	887
Glucose-6-phosphate isomerase ^{a,b}	5.3.1.9	281	1539	311	516	1145	187
6-phosphofructokinase 1 ^a	2.7.1.11	1443	971	1897	1468	1331	762
Fructose-bisphosphate aldolase class II ^a	4.1.2.13	392	443	140	1041	94	331
glucose-6-phosphate 1-dehydrogenase ^b	1.1.1.49	1750					
6-phosphogluconolactonase ^b	3.1.1.31	1690					
6-phosphogluconate dehydrogenase ^b	1.1.1.44	1918					
ribulose-phosphate 3-epimerase ^b	5.1.3.1	47	1612 1613	1603	2154	241	1953
Xylulose-5-phosphate phosphoketolase ^b	4.1.2.9	2010	150 151				
glyceraldehyde 3-phosphate dehydrogenase ^{a,b}	1.2.1.12	37	1622	1595	2144	292	1945
phosphoglycerate kinase ^{a,b}	2.7.2.3	35	1623 1624	1593	2143	293	1943
glyceraldehyde-3-phosphate dehydrogenase (NADP) ^{a,b}	1.2.1.9		1346	1810	696 697	2201 2202	861 862
2,3-bisphosphoglycerate-dependent phosphoglycerate mutase ^{a,b}	5.4.2.1	797	1415	1891	1477	1323	769
enolase ^{a,b}	4.2.1.11	1692	2143	1650	428	2044	1033
pyruvate kinase ^{a,b}	2.7.1.40	1444	970	1898	1467	1332	761
Phosphate acetyltransferase ^b	2.3.1.8	1190	943	1241	1771	656	1592
Acetaldehyde dehydrogenase ^b	1.2.1.10	707	260 261 262	130	1031 1032	86 87	323
Alcohol dehydrogenase ^b	1.1.1.1	2035 592 593 707	1673 260 261 262	130 756 819 820	1031 1032 1122 1123 848	1570 1621 86 87	1123 323 478

^a: Enzyme involved in glycolysis^b: Enzyme involved in pentose phosphate pathway

Table S5.10. Comparison of isolate groupings from genetic fingerprinting and results from lineage-specific PCRs

Isolate	Genus identification	Grouping according to AFLP and Rep-PCR analysis				Grouping according to strain-specific PCR assays								
		<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>				<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>				
				1	2	3	4			1	2	3	4	
2010_ileo_MS_Ia	<i>Streptococcus</i>													
2010_ileo_MS_Ib	<i>Streptococcus</i>													
2010_ileo_MS_Ic	<i>Streptococcus</i>													
2010_ileo_MS_Id	<i>Streptococcus</i>													
2010_ileo_MS_Ie	<i>Streptococcus</i>													
2010_ileo_MS>If	<i>Streptococcus</i>													
2010_ileo_MS_Ila	<i>Streptococcus</i>													
2010_ileo_MS_Ilb	<i>Streptococcus</i>													
2010_ileo_MS_Ilc	<i>Streptococcus</i>													
2010_ileo_MS_Ild	<i>Streptococcus</i>													
2010_ileo_MS_Ile	<i>Streptococcus</i>													
2010_ileo_MS_IIf	<i>Streptococcus</i>													
2010_ileo_MS_IIIa	<i>Streptococcus</i>													
2010_ileo_MS_IIIb	<i>Streptococcus</i>													
2010_ileo_MS_IIIc	<i>Streptococcus</i>													
2010_ileo_MS_IId	<i>Streptococcus</i>													
2010_ileo_MS_IIIe	<i>Streptococcus</i>													
2010_ileo_MS_IIIf	<i>Streptococcus</i>													
2010_ileo_MS_IVa	<i>Enterococcus</i>													
2010_ileo_MS_IVb	<i>Enterococcus</i>													
2010_ileo_MS_IVc	<i>Enterococcus</i>													
2010_ileo_MS_IVd	<i>Enterococcus</i>													
2010_ileo_MS_IVe	<i>Enterococcus</i>													
2010_ileo_MS_IVf	<i>Enterococcus</i>													
2010_ileo_MS_Va	<i>Streptococcus</i>													
2010_ileo_MS_Vb	<i>Enterococcus</i>													
2010_ileo_MS_Vc	<i>Streptococcus</i>													
2010_ileo_MS_Vd	<i>Streptococcus</i>													
2010_ileo_MS_Ve	<i>Streptococcus</i>													
2010_ileo_MS_Vf	<i>Streptococcus</i>													
2010_ileo_MS_VIa	<i>Enterococcus</i>													
2010_ileo_MS_VIb	<i>Streptococcus</i>													
2010_ileo_MS_VIc	<i>Streptococcus</i>													
2010_ileo_MS_VId	<i>Streptococcus</i>													
2010_ileo_MS_VIe	<i>Streptococcus</i>													
2010_ileo_MS_VIf	<i>Streptococcus</i>													
2010_ileo_MS_VIIa	<i>Streptococcus</i>													
2010_ileo_MS_VIIb	<i>Streptococcus</i>													
2010_ileo_MS_VIIc	<i>Streptococcus</i>													
2010_ileo_MS_VIId	<i>Enterococcus</i>													
2010_ileo_MS_VIIe	<i>Streptococcus</i>													
2010_ileo_MS_VIIIf	<i>Streptococcus</i>													
2010_ileo_MS_VIIIa	<i>Streptococcus</i>													
2010_ileo_MS_VIIIb	<i>Streptococcus</i>													
2010_ileo_MS_VIIIc	<i>Streptococcus</i>													
2010_ileo_MS_VIIId	<i>Streptococcus</i>													
2010_ileo_MS_VIIIe	<i>Streptococcus</i>													
2010_ileo_MS_VIIIIf	<i>Streptococcus</i>													
2010_ileo_MS_IXa	<i>Streptococcus</i>													
2010_ileo_MS_IXb	<i>Streptococcus</i>													
2010_ileo_MS_IXc	<i>Streptococcus</i>													
2010_ileo_MS_IXd	<i>Streptococcus</i>													
2010_ileo_MS_IXe	<i>Streptococcus</i>													
2010_ileo_MS_IXf	<i>Streptococcus</i>													
2010_ileo_MS_Xa	<i>Enterococcus</i>													
2010_ileo_MS_Xb	<i>Enterococcus</i>													
2010_ileo_MS_Xc	<i>Enterococcus</i>													
2010_ileo_MS_Xd	<i>Enterococcus</i>													
2010_ileo_MS_Xe	<i>Enterococcus</i>													
2010_ileo_MS_Xf	<i>Streptococcus</i>													



Table S5.10 (Continued). Comparison of isolate groupings from genetic fingerprinting and results from lineage-specific PCRs

Isolate	Genus identification	Grouping according to AFLP and Rep-PCR analysis				Grouping according to strain-specific PCR assays								
		<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>				<i>S. mitis</i>	<i>S. bovis</i>	<i>S. salivarius</i>				
				1	2	3	4			1	2	3	4	
2010_ileo_MS_XIa	<i>Streptococcus</i>													
2010_ileo_MS_XIb	<i>Streptococcus</i>													
2010_ileo_MS_XIc	<i>Streptococcus</i>													
2010_ileo_MS_XId	<i>Streptococcus</i>													
2010_ileo_MS_XIe	<i>Streptococcus</i>													
2010_ileo_MS_XIf	<i>Streptococcus</i>													
2010_ileo_MS_XIIa	<i>Enterococcus</i>													
2010_ileo_MS_XIIb	<i>Streptococcus</i>													
2010_ileo_MS_XIIc	<i>Streptococcus</i>													
2010_ileo_MS_XIId	<i>Streptococcus</i>													
2010_ileo_MS_XIIE	<i>Enterococcus</i>													
2010_ileo_MS_XIIIf	<i>Streptococcus</i>													
2010_ileo_MS_XIIIf	<i>Enterococcus</i>													
2010_ileo_MS_XIIIa	<i>Streptococcus</i>													
2010_ileo_MS_XIIIb	<i>Streptococcus</i>													
2010_ileo_MS_XIIIc	<i>Streptococcus</i>													
2010_ileo_MS_XIIId	<i>Streptococcus</i>													
2010_ileo_MS_XIIIe	<i>Streptococcus</i>													
2010_ileo_MS_XIIIf	<i>Streptococcus</i>													
2010_ileo_MS_XIVa	<i>Enterococcus</i>													
2010_ileo_MS_XIVb	<i>Streptococcus</i>													
2010_ileo_MS_XIVc	<i>Enterococcus</i>													
2010_ileo_MS_XIVd	<i>Enterococcus</i>													
2010_ileo_MS_XIVe	<i>Enterococcus</i>													
2010_ileo_MS_XIVf	<i>Streptococcus</i>													
2010_ileo_MS_XIVf	<i>Streptococcus</i>													
2010_ileo_MS_XVa	<i>Streptococcus</i>													
2010_ileo_MS_XVb	<i>Streptococcus</i>													
2010_ileo_MS_XVb	<i>Streptococcus</i>													
2010_ileo_MS_XVc	<i>Streptococcus</i>													
2010_ileo_MS_XVc	<i>Streptococcus</i>													
2010_ileo_MS_XVd	<i>Enterococcus</i>													
2010_ileo_MS_XVe	<i>Streptococcus</i>													
2010_ileo_MS_XVf	<i>Streptococcus</i>													
2010_ileo_MS_XVf	<i>Streptococcus</i>													
2010_ileo_MS_XVIa	<i>Streptococcus</i>													
2010_ileo_MS_XVIb	<i>Streptococcus</i>													
2010_ileo_MS_XVIc	<i>Streptococcus</i>													
2010_ileo_MS_XVId	<i>Streptococcus</i>													
2010_ileo_MS_XVIe	<i>Streptococcus</i>													
2010_ileo_MS_XVIf	<i>Streptococcus</i>													
2010_ileo_MS_XVIIa	<i>Streptococcus</i>													
2010_ileo_MS_XVIIa	<i>Streptococcus</i>													
2010_ileo_MS_XVIIb	<i>Streptococcus</i>													
2010_ileo_MS_XVIIc	<i>Streptococcus</i>													
2010_ileo_MS_XVIIc	<i>Streptococcus</i>													
2010_ileo_MS_XVIIId	<i>Streptococcus</i>													
2010_ileo_MS_XVIIe	<i>Streptococcus</i>													
2010_ileo_MS_XVIIe	<i>Streptococcus</i>													
2010_ileo_MS_XVIIIf	<i>Streptococcus</i>													
2010_ileo_MS_XVIIIf	<i>Streptococcus</i>													
2010_ileo_MS_XVIIIa	<i>Streptococcus</i>													
2010_ileo_MS_XVIIIa	<i>Streptococcus</i>													
2010_ileo_MS_XVIIIb	<i>Enterococcus</i>													
2010_ileo_MS_XVIIIc	<i>Enterococcus</i>													
2010_ileo_MS_XVIIIc	<i>Enterococcus</i>													
2010_ileo_MS_XVIIIId	<i>Enterococcus</i>													
2010_ileo_MS_XVIIIe	<i>Enterococcus</i>													
2010_ileo_MS_XVIIIe	<i>Enterococcus</i>													
2010_ileo_MS_XVIIIf	<i>Enterococcus</i>													
2010_ileo_MS_XIXa	<i>Streptococcus</i>													
2010_ileo_MS_XIXa	<i>Streptococcus</i>													
2010_ileo_MS_XIXb	<i>Streptococcus</i>													
2010_ileo_MS_XIXc	<i>Streptococcus</i>													
2010_ileo_MS_XIXd	<i>Streptococcus</i>													
2010_ileo_MS_XIXe	<i>Streptococcus</i>													
2010_ileo_MS_XIXf	<i>Streptococcus</i>													
2010_ileo_MS_XXa	<i>Streptococcus</i>													
2010_ileo_MS_XXa	<i>Streptococcus</i>													
2010_ileo_MS_XXb	<i>Streptococcus</i>													
2010_ileo_MS_XXb	<i>Streptococcus</i>													
2010_ileo_MS_XXc	<i>Streptococcus</i>													
2010_ileo_MS_XXc	<i>Streptococcus</i>													
2010_ileo_MS_XXd	<i>Streptococcus</i>													
2010_ileo_MS_XXd	<i>Streptococcus</i>													
2010_ileo_MS_XXe	<i>Streptococcus</i>													
2010_ileo_MS_XXe	<i>Streptococcus</i>													
2010_ileo_MS_XXf	<i>Streptococcus</i>													
2010_ileo_MS_XXf	<i>Streptococcus</i>													
Total number of isolate per lineage		3	13	1	1	8	66	3	13	57	1	8	67	

Chapter 6

Immunomodulatory properties of
Streptococcus and *Veillonella* isolates
from the human small intestine microbiota

Bartholomeus van den Bogert*, Marjolein Meijerink*, Erwin G. Zoetendal, Jerry M. Wells, and Michiel Kleerebezem

*These authors contributed equally to this work

Manuscript submitted for publication

Abstract

The human small intestine is a key site for interactions between the intestinal microbiota and the mucosal immune system. Here we investigated the immunomodulatory properties of representative species of commonly dominant small-intestinal microbial communities, including six streptococcal strains (four *Streptococcus salivarius*, one *S. bovis*, one *S. mitis*) one *Veillonella parvula* strain, one *Enterococcus gallinarum* strain, and *Lactobacillus plantarum* WCFS1 as a bench mark strain on human monocyte-derived dendritic cells. The different streptococci induced varying levels of the cytokines IL-8, TNF- α , and IL-12p70, while the *V. parvula* strain showed a strong capacity to induce IL-6. *E. gallinarum* strain was a potent inducer of cytokines and TLR2/6 signalling. As *Streptococcus* and *Veillonella* can potentially interact metabolically and frequently co-occur in ecosystems, immunomodulation by pair-wise combinations of strains were also tested for their combined immunomodulatory properties. Strain combinations induced cytokine responses in dendritic cells that differed from what might be expected on the basis of the results obtained with the individual strains. A combination of (some) streptococci with *Veillonella* appeared to negate IL-12p70 production, while augmenting IL-8, IL-6, IL-10, and TNF- α responses. This suggests that immunomodulation data obtained *in vitro* with individual strains are unlikely to adequately represent immune responses to mixtures of gut microbiota communities *in vivo*. Nevertheless, analysing the immune responses of strains representing the dominant species in the intestine may help to identify immunomodulatory mechanisms that influence immune homeostasis.

Introduction

The human intestine is home to a myriad of different microbial organisms, most of which are bacteria (341) and collectively known as microbiota. The intestinal microbiota is of particular interest because it plays an essential role in the maturation and development of the mucosal immune system in early life (170, 217) and the preferential tolerance induction to harmless antigens at mucosal sites (172, 315). The contribution of individual microbes to the mechanisms that maintain immune homeostasis are just beginning to be understood (154, 241, 341), but their importance is highlighted by the disturbances in microbiota composition associated with several intestinal-related diseases including obesity, multiple sclerosis, inflammatory bowel diseases, and type 1 diabetes (170, 196, 322, 326, 352, 359). Research on this topic has been biased towards the analysis of fecal samples that only provide information about the microbiota at the end of gastrointestinal (GI) tract (159, 194, 329), meaning that, microbial communities in the upper intestinal tract have been largely overlooked or ignored (381). This is mainly attributable to the limited accessibility of the small intestine (31, 32). Nevertheless, the Peyer's patches (PP) of the small intestine are major sites for sampling of luminal antigens, including bacteria, and the induction of adaptive immune responses. Antigen sampling by the follicle-associated epithelium (FAE) overlaying the lymphoid follicles of the PP is facilitated by the lack of mucin secreting goblet cells and the presence of specialized Microfold cells (M cells) (75, 217, 242). Bacteria sampled by M cells in the FAE are transported intact to the sub-epithelial dome of PP where dendritic cells (DCs) play a key role in bacterial handling and the induction of subsequent immune responses (see (55) for a review). Recently, PP dendritic cells were shown to sample bacteria and antigens through M cell-specific transcellular pores (199, 200). Additionally, DCs have been shown to sample luminal bacteria (and other luminal constituents) directly in the lumen by passing protrusions through the paracellular space of the epithelium without disrupting epithelial integrity (270).

While both the human small and large intestinal microbiota encompasses anaerobes belonging to the *Clostridium* clusters, the marked difference between these intestinal niches is a microbial composition predominated by facultative anaerobes, including the streptococci and *Veillonella* bacteria in the small intestine ((31, 339, 381); Chapter 3). The co-occurrence of these genera may in part depend on their potential for metabolic interaction as shown in the oral cavity (84) and previously postulated for the small intestine (381). Support for this notion comes from the high expression of genes involved primary carbohydrate transport systems by the small intestinal streptococci (381), indicating a role for the *Streptococcus* populations as primary fermentors of diet-derived simple sugars in the human small intestine. Characterization of small-intestinal bacterial streptococci revealed that the small intestine is inhabited by a variety of *Streptococcus* lineages that belong to the *S. mitis*, *S. bovis*, and *S. salivarius* species groups. These lineages displayed considerable phenotype variability in terms of carbohydrate utilization capacities ((340); Chapter 4), which was in excellent agreement with their capacities predicted

on basis of their genome sequences (Chapter 5). With the exception of streptococci, the lactic acid bacteria are generally present at low abundance in the small intestine microbiota (31) (Leimena and Van den Bogert, et al. Unpublished results; Chapter 3), but nevertheless display a substantial level of phylogenetic richness in individuals, as was the case for members of the genus *Enterococcus* ((340); Chapter 4). The enterococci are common colonizers of the GI tract, but have a less attractive reputation because of their pathogenic potential (10).

Considering the prominent role of DCs in modulation of the small-intestinal immune system the aim of the current study was to investigate the immunomodulatory properties of different small-intestinal *Enterococcus*, *Streptococcus*, and *Veillonella* isolates ((340); Chapter 4), with a special focus on the latter two genera because of their predominance in the small-intestinal ecosystem.

Materials and methods

Bacterial strains

Six *Streptococcus* strains (with known genome sequences; Chapter 5) and single *Enterococcus gallinarum* HSIEG1 and *Veillonella parvula* HSIVP1 strains (340), as well as the reference strain *Lactobacillus plantarum* WCFS1 (175) were used in the immunoassays (Table 6.1). The streptococcal strains were representative isolates of 6 distinct phylogenetic lineages, as determined by DNA fingerprinting, and for the remainder of the paper will be referred to by their species-group names: *S. mitis* (1 strain; HSISM1), *S. bovis* (1 strain; HSISB1) and *S. salivarius* (4 strains; HSISS1-4; Table 6.1; (340); Chapter 4). The streptococcal and *Enterococcus* strains were grown in *Mitis Salivarius* (MS) medium (340), while *Veillonella* was grown in medium described in the DSMZ catalogue (Medium 136) under anoxic N₂ atmosphere. *Lactobacillus plantarum* WCFS1 was grown in Mann-Rogosa Sharpe (MRS) medium (Becton Dickinson, Breda, the Netherlands). Fresh culture media did not induce any cytokine responses (data not shown). All strains were twice subcultured overnight successively, after which the streptococci and the *Enterococcus* strains had an average OD₆₀₀ of 1.3 (\pm standard deviation of 0.2), while the *V. parvula* strain and WCFS1 had OD₆₀₀ of approximately 0.5 and 2.5, respectively. The bacteria suspensions were diluted in PBS (GIBCO) to a final OD₆₀₀ of 1.

Table 6.1. Strains used in this study

Species	Strain identifier	Origin	Reference
<i>Streptococcus mitis</i>	HSISM1		
<i>Streptococcus bovis</i>	HSISB1		
<i>Streptococcus salivarius</i>	HSISS1		
<i>Streptococcus salivarius</i>	HSISS2	Ileostoma effluent	(340); Chapter 4
<i>Streptococcus salivarius</i>	HSISS3		
<i>Streptococcus salivarius</i>	HSISS4		
<i>Veillonella parvula</i>	HSIVP1		
<i>Enterococcus gallinarum</i>	HSIEG1		
<i>Lactobacillus plantarum</i>	WCFS1	Human saliva	(175)

Differentiation and maturation of dendritic cells

The study was approved by the Wageningen University Ethical Committee and was performed according to the principles of the Declaration of Helsinki. Buffy coats were obtained from the Sanquin Blood bank Nijmegen, the Netherlands. A written informed consent was obtained before sample collection. Human monocytes were isolated from blood using a combination of Ficoll density centrifugation and cell separation using CD14-specific antibody coated magnetic microbeads (Miltenyi Biotec). The purity of isolated CD14+ cell fraction was greater than 90% and cell-viability was above 95% in all experiments. To generate immature DC (iDCs), the purified CD14+ cells were cultured for 6 days in RPMI 1640 medium (Invitrogen, Breda, the Netherlands), supplemented with 100 units/ml penicillin G (Invitrogen), 100 µg/ml streptomycin (Invitrogen), 50 ng/ml IL-4 (R&D systems, Abingdon, United Kingdom) and 50 ng/ml granulocyte-macrophage colony-stimulating-factor (GM-CSF) (R&D systems). GM-CSF and IL-4 were added to differentiate the monocytes into myeloid DCs. On day 6 approximately 1×10^6 iDCs were stimulated with LPS (1 µg/ml) or the different bacteria at a cell to bacteria ratio of approximately 1:1 and 1:10 for 48 hours. As anticipated and as a consequence of the supplementation of the cell-media with antibiotics, no bacterial growth was observed during this period. Non-stimulated iDCs were used as a negative control.

Analyses of cell surface markers and measurement of cell death by flow cytometry

During the culturing period (8 days) of the CD14+ cells, cells were stained on days 3, 6 and 8 with fluorescence-conjugated monoclonal antibodies specific for CD83, CD86 or their isotype-matched controls (BD biosciences, San Diego, USA) and analysed by flow cytometry (FACSCanto II, BD, San Diego, USA) to check the maturation and activation status of the cells. CD83 and CD86 are highly expressed on DCs after stimulation with known maturation factors (e.g. LPS) compared to non-stimulated immature dendritic cells. The expression of CD83 and CD86 from different human donors can vary considerably after stimulation with different stimuli so for comparison the data was normalized to the values (100%) obtained using a standard amount of LPS added to cells from each donor.

On days 3, 6 and 8 the percentage of viable cells was measured by flow cytometry (FACSCanto II, BD, San Diego, USA). Live, apoptotic and necrotic cells were discriminated by staining with Annexin V and propidium iodide (PI) on days 3, 6 and 8 according to the manufacturer's protocol. The cells were analysed using flow cytometry (FACSCanto II, BD, San Diego, USA) and the BD FACSDiva software. Cells that are negative for both Annexin V and PI are not apoptotic or necrotic as translocation of the membrane phospholipid phosphatidylserine has not occurred and the plasma membrane is still intact. Therefore, Annexin V and PI double negative cells were considered as viable cells, whereas both single and double positive cells were regarded as non-viable (349). On days 3 to 8 the viability of the cells was between 60 and 95%. There were no significant differences in cell death between the

bacteria-stimulated cells and the non-stimulated (negative control) or LPS-stimulated (positive control) cells.

Cytokine assays

Supernatants from the DC stimulation assays were collected after stimulation for 24 hours, and analysed for the presence of cytokines (IL-1 β , IL-6, IL-8, IL-10, IL-12p70 and TNF- α) using a cytometric bead-based BD Human inflammation kit that enables multiplex measurements of soluble cytokines in the same sample (237), according to the manufacturer's protocol (BD biosciences, Breda, the Netherlands). The sensitivity-limits of detection were as follows: IL-1 β 7.2 pg/ml, IL-6 2.5 pg/ml, IL-8 3.6 pg/ml, IL-10 3.3 pg/ml, IL-12p70 1.9 pg/ml and TNF- α 0.7 pg/ml. The flow cytometry data were analysed using the BD FCAP software. Unless stated otherwise, cytokine secretion in the remainder of the paper are based on stimulation of iDCs with a DC to bacteria ratio of approximately 1:10.

TLR2/6 assay

TLR2/6 signalling capacities of the bacterial strains were determined using a reporter assay with Human Embryonic Kidney (HEK)293 (Invivogen, Toulouse, France) cells expressing human TLR2 and TLR6 heterodimers that recognize lipoteichoic acid (LTA) and lipoprotein lipid anchors of Gram-positive bacteria (4). The TLR2/6 signalling assay was performed essentially as previously described (167). Briefly, HEK293 cells were transformed with human TLR2/6 and pNIFTY, a NF- κ B luciferase reporter construct (Invivogen, Toulouse, France). HEK293 cells transformed with only the pNIFTY did not respond to Pam₂CSK (synthetic agonist of TLR2/6) demonstrating the dependency of NF- κ B activation on co-expression of hTLR2/6 receptor (167). The cells were plated at a concentration of 6×10^4 cells per well in DMEM medium (Invitrogen). Cells were then stimulated with the different bacterial strains, or Pam2CSK as a positive control or with medium alone (negative control) followed by incubation at 37°C for 6 hours under a 5% CO₂ atmosphere. Thereafter, the medium was replaced with Bright glow (Promega, Leiden, the Netherlands), and the plates shaken for 5 minutes before measuring the luminescence in a Spectramax M5 (Molecular Devices, Sunnyvale, United States). HEK293 cells not expressing TLR receptors that harbour pNIFTY were used as the negative control in the NF- κ B assays.

Statistical analysis

Mixed general linear model using restricted maximum likelihood (REML) was used to determine the statistical differences within donors between cytokine produced by DCs stimulated with the different bacterial strains. A two-sided *p*-value of 0.05 or lower was considered to be significant. The statistical analysis was performed by using SAS software (version 9.1, SAS Institute Inc., Cary, NC, USA).

Results

Small-intestinal bacteria differentially affect DC maturation and activation

S. mitis HSISM1, *S. bovis* HSISB1, four different *S. salivarius* strains (HSISS1-4), *E. gallinarum* HSIEG1, and *V. parvula* HSIVP1 strains obtained from the human small intestine were investigated for their capacity to induce maturation and activation of immature monocyte-derived DCs from donors. The DCs were stimulated for 48 hours with different strains at DC to bacteria ratios of 1 and 10. The expression of the surface marker CD83 (maturation marker) and CD86 (maturation marker and co-stimulatory molecule) were measured to determine maturation and activation status of the DCs. The mean fluorescence intensity (MFI) of dendritic cells was normalized to LPS stimulation (Figure 6.1). Stimulation of DCs with the high dose of bacterial strains (1 to 10) resulted in higher maturation and activation marker expression than with the low dose (1 to 1), except in the case of *V. parvula* stimulation. The induction of the expression of the surface markers CD86 and CD83 differed markedly among the different species used as DC stimulants, and with *S. salivarius* strain 1, 3, and 4 as well as *E. gallinarum* inducing highest expression. Moreover, the data obtained suggested that, *S. bovis* and *S. salivarius* strain 2 were the least effective at maturing DC.

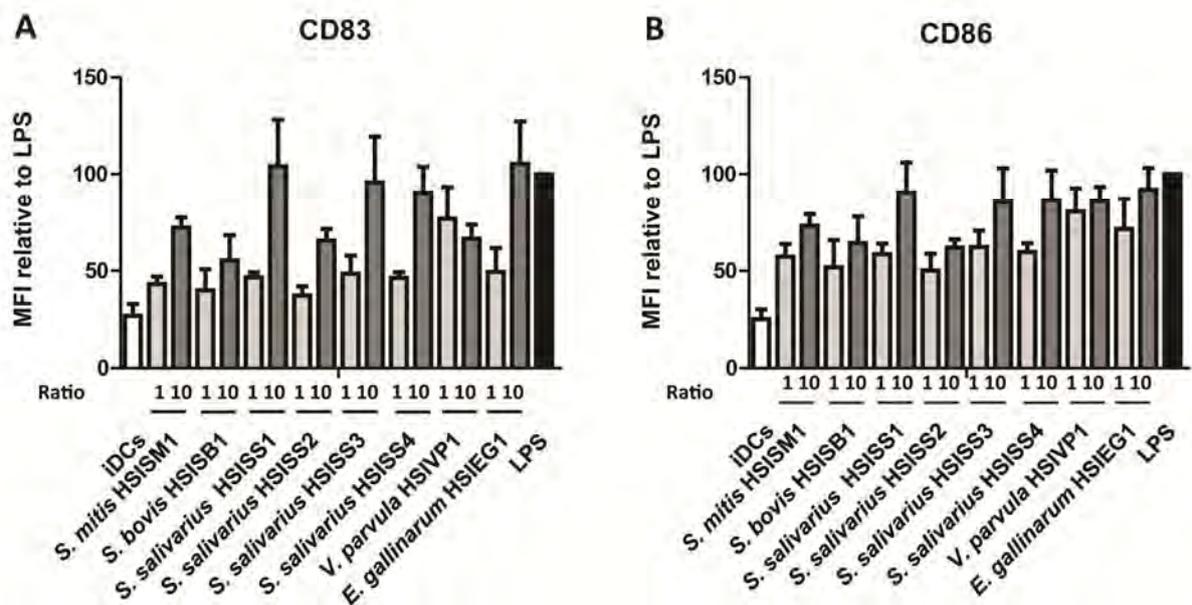


Figure 6.1. The MFI of stained cell surface markers CD83 (A) and CD86 (B) by monocyte derived dendritic cells, with immature DCs as the negative control and LPS as the positive control.

DC cytokine responses to bacterial isolates from the small intestine

The small-intestinal *Streptococcus*, *Veillonella* and *Enterococcus* strains were further investigated for their capacity to induce cytokine secretion by monocyte-derived iDCs. In addition, *L. plantarum* WCFS1 was tested and used as a benchmark strain that was analysed several times before (232, 233, 300, 336). Although cytokine responses upon stimulation with the different bacterial strains varied between the different donors, the induced immune profiles were consistent (Figure 6.2).

In agreement with what has previously been described for members of the *S. bovis* species group (27), the strain tested here induced relatively low levels of cytokines (Figure 6.2). The *V. parvula* strain elicited a moderate induction of the production of the cytokines IL-8, IL-1 β , IL-10, and TNF- α . In contrast to the *Streptococcus* strains, *V. parvula* did not stimulate IL-12p70 secretion in DCs, whereas its capacity to induce IL-6 was substantially higher.

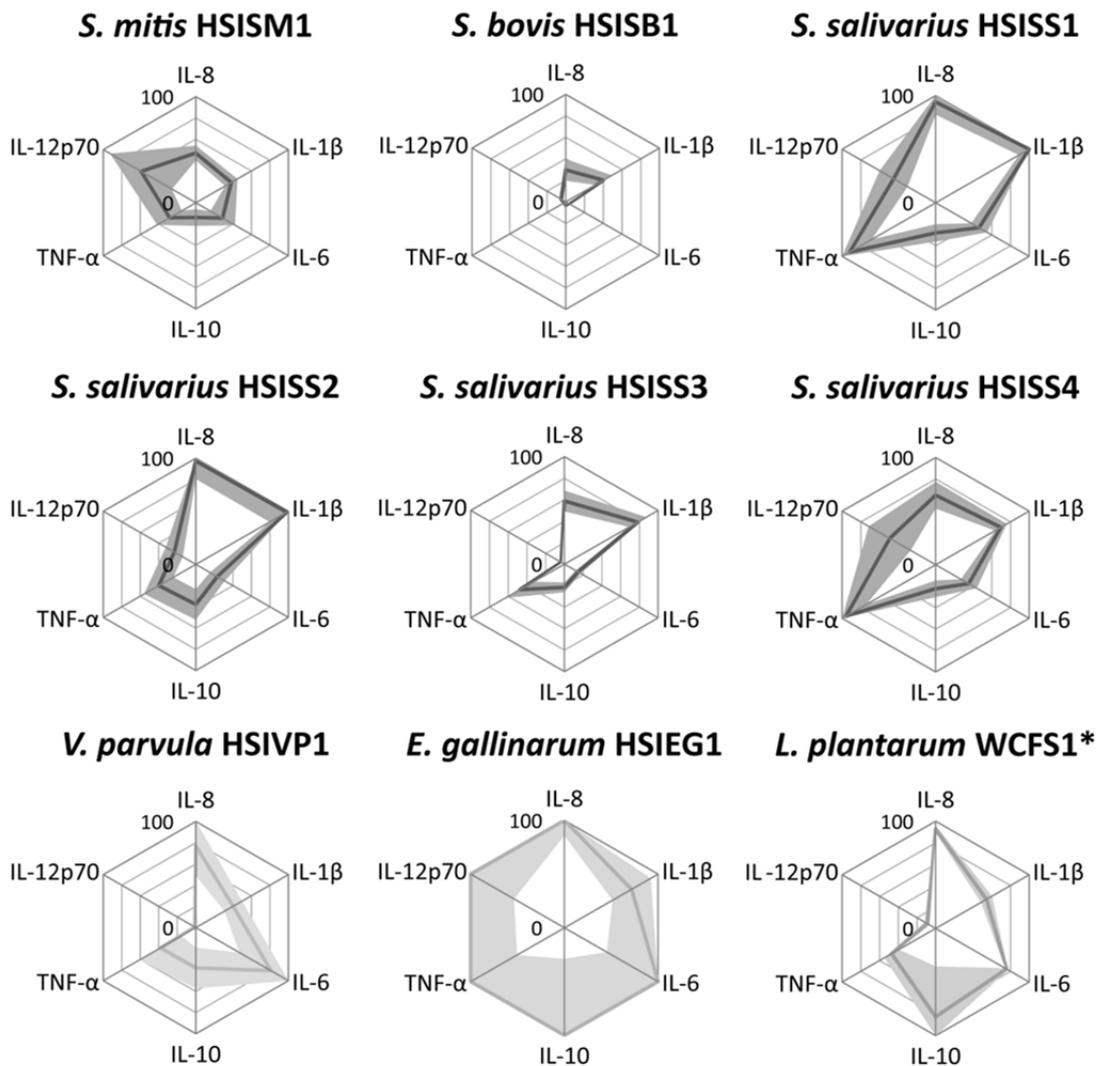


Figure 6.2. Cytokine secretion by dendritic cells derived from 5 human donors after stimulation with bacterial strains. Cytokine levels are expressed as relative values of the highest inducing strain (100% cytokine levels (pg/ml): IL-8: 17598 ; IL-1 β : 41; IL-6: 4775; IL-10: 206; TNF- α : 5151; IL-12p70: 2397; Table S1). Lines represents the average secreted cytokine amounts and faded colours represent the interval between the upper and the lower SEM. SEM values higher than 100% are not visualized.

*: Cytokine responses determined using DCs derived from 2 different human donors (Figure 5, Table S2)

Although, the cytokine response between donors for the *E. gallinarum* strain varied, the induced cytokine amounts by *E. gallinarum* strain were the highest among the tested strains (Figure 6.2), indicating that the immune system response is more pronounced if triggered with this strain compared to the streptococci. Furthermore,

the IL-10 and TNF- α levels induced by *L. plantarum* WCFS1 were comparable to previous study (232). Noteworthy, *L. plantarum* WCFS1 induced considerably higher amounts of IL-8, IL-6, and IL-10, higher than the streptococci, albeit that this was based on DCs derived from 2 donors (Table S6.2).

The *Streptococcus* strains induced relatively consistent levels of IL-1 β (16-41 pg/ml), IL-6 (161-2221 pg/ml), and IL-10 (7-78 pg/ml), while showing substantial differences in their ability to induce the production of the chemokine IL-8 (5231-17147 pg/ml) and the pro-inflammatory cytokines TNF- α (86-4933 pg/ml) and IL-12p70 (81-1416 pg/ml) (Figure 6.2; Table S6.1). This illustrates that the strains tested here elicited distinct cytokine profiles, which is in agreement with earlier observations that revealed distinct DC responses to closely related species and strains (233, 335). Notably, the immune response profiles elicited by the *S. salivarius* strains 1 and 4, were similar, which corroborates earlier observations on the genetic and physiological similarity of these strains ((340); Chapter 4 and 5).

***S. bovis* is not immunosuppressive**

As the *S. bovis* strain elicited a low immune response compared to the other strains tested (Figure 6.2), we hypothesized that this strain might possess an immunomodulatory component, which suppresses cytokine secretion. Therefore, we co-stimulated DCs with LPS (10 ng/ml) and *S. bovis* or *S. salivarius* strain 4. The *S. bovis* strain did not reduce the cytokine levels induced by LPS stimulation. In fact, the amount of IL-6 produced by DCs stimulated with *S. bovis* and LPS together, was greater than the sum of the individual IL-6 responses, indicating a synergistic effect of these stimuli on secretion of this cytokine by DCs. A synergistic effect on IL-6 production was also observed when DCs were co-stimulated with *S. salivarius* and LPS (Figure 6.3). Furthermore, co-stimulation of DCs with LPS in combination with spent culture supernatant from either of the two bacterial strains also elevated cytokine production as compared to LPS alone (Figure 6.3), with the exception of TNF- α which was produced in lower amounts compared to the LPS alone. Although, spent bacterial culture supernatant elevated cytokine secretion in combination with LPS it was less than that measured when bacteria were combined with LPS as a stimulus. Together these results indicate that *S. bovis* is not actively immunosuppressing dendritic cell cytokine production and can in fact enhance cytokine production in response to LPS.

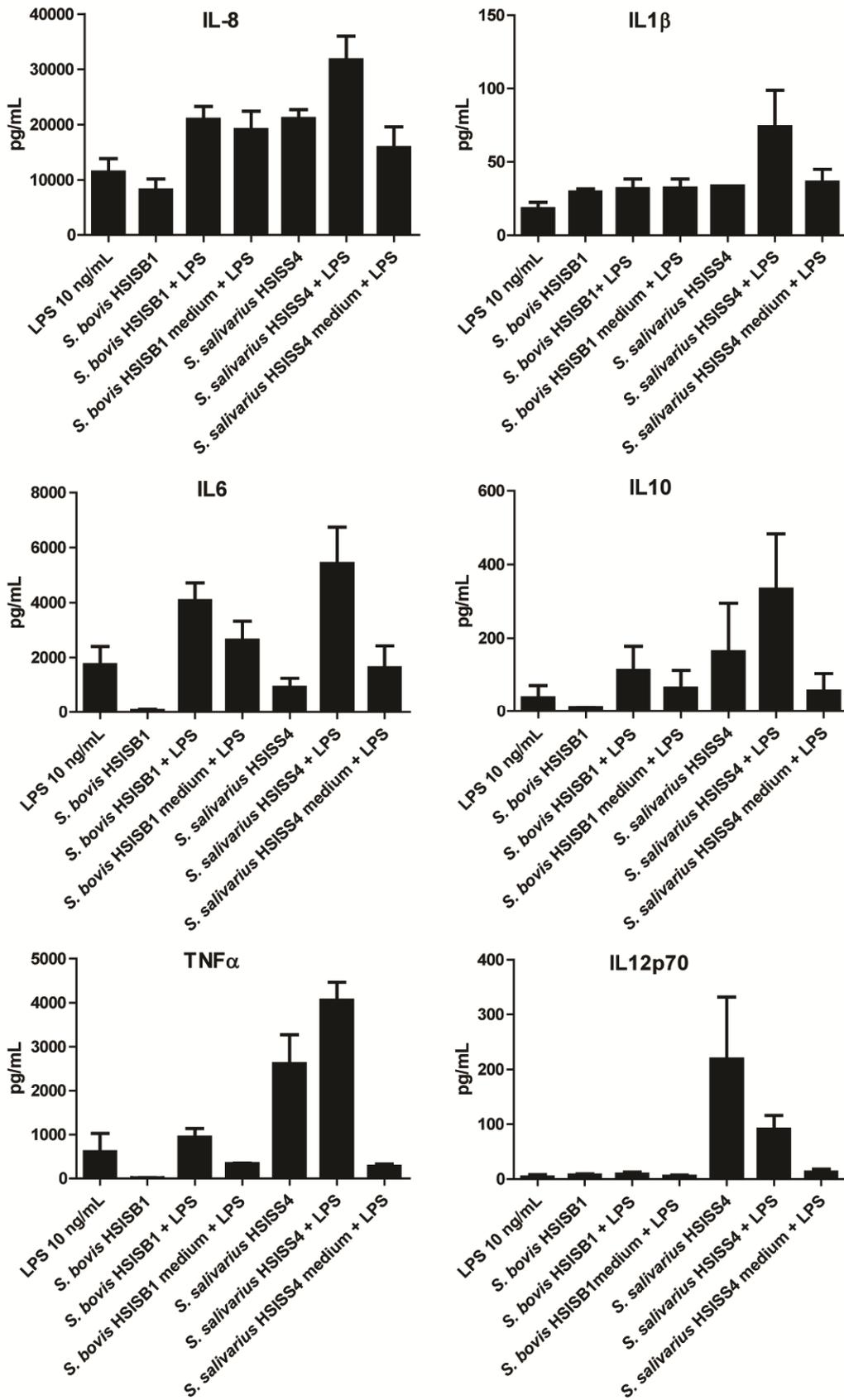


Figure 6.3. Cytokine secretion by dendritic cells derived from 2 human donors after mono-stimulation with tested strains, disrupted strains, medium strains with or without co-stimulation with LPS.

Involvement of TLR2 and TLR6 in innate immune signalling by small-intestinal *Streptococcus* and *Veillonella* strains

TLR2/6-mediated activation of NF- κ B is potentially one of the major pathways for DC activation via LTA or lipoproteins in the cell envelope of bacteria. Therefore, we tested the TLR2/6 signalling capacities of the *Streptococcus* and *Veillonella* strains in a reporter assay using HEK293 cells expressing human TLR2 and TLR6 heterodimers that recognize lipoteichoic acid (LTA) and lipoprotein lipid anchors of Gram-positive bacteria (4). The results demonstrated that most strains are capable of triggering NF- κ B activation via TLR2/6 dependent signalling. However, the *S. bovis* strain did not activate significant TLR2/6 signalling in this reporter assay (Figure 6.4), which is analogous to its failure to induce high levels of cytokine production in DCs (see above). Similarly, the strong DC-response elicited by *E. gallinarum* stimulation was reflected in its TLR2/6 signalling capacity, where this strain classified as one of the strongest TLR2/6 stimulators among the strains tested. These results suggest that there is at least a certain degree of congruency between the capacity of individual strains to elicit TLR2/6 signalling in HEK293 NF- κ B reporter cells, and their capacity to stimulate high levels of cytokine production in iDCs. However, *S. salivarius* strain 2 hardly induced responses in DCs but strongly activated TLR2/6 signalling, illustrating that the congruency between the two cell-assays is not universally supported and may vary, depending on the stimulus tested.

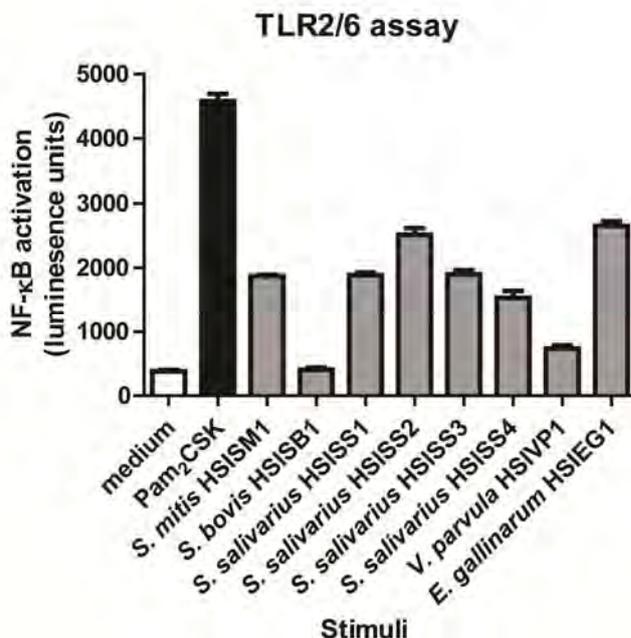


Figure 6.4. hTLR2/6 assay. HEK293 cells were incubated with the small-intestinal strains at a cell to bacteria ratio of 1:10, Pam₂CSK as a positive control and medium as a negative control. This figure is representative out of two hTLR2/6 assays.

Co-stimulation of dendritic cells with streptococci and *Veillonella*

Based on the frequent co-occurrence of the *Streptococcus* and *Veillonella* spp. in various habitats associated with the human body (84, 381), we considered it relevant to evaluate the cytokine responses that were elicited in iDCs that were co-stimulated with one of the small intestinal *Streptococcus* strains in combination with the *V. parvula* strain.

The amounts of cytokines IL-8, IL-6, IL-10, TNF- α , and IL-12p70 produced in DCs stimulated with a combination of the two species was consistently different from the levels anticipated on basis of mono-stimulation with the individual strains (Figure 6.5). However, this was not the case for IL-1 β that was secreted in low amounts by mono- and co-stimulated DCs, in all stimulation assays (Table S6.2). As an example of the specific co-stimulatory effects, the production of cytokine IL-12p70 by DCs could be elicited by stimulation with several of the streptococci, but this stimulation effect was suppressed by co-stimulation with *V. parvula*. This suppression may be influenced by the amounts of IL-10 induced in the co-stimulation assays and autocrine signalling via the IL-10R (56). This observation was especially clear in the comparison of the stronger IL-12p70 inducing streptococcal mono-stimulations (*S. mitis* and the strains representing *S. salivarius* lineage 1 and 4) with the corresponding co-stimulations that included *V. parvula* (Figure 6.5). This result implies that *V. parvula* can suppress the pro-inflammatory stimulation elicited by the streptococci. This proposed suppression of pro-inflammatory responses by *V. parvula* co-stimulation, appeared specific for the *Streptococcus* co-stimulations in this study, as the strong IL-12p70 induction elicited by mono-stimulation with *E. gallinarum* appeared unaffected by co-stimulation with *V. parvula* (Figure 6.5; Table S6.2). Furthermore, the amounts of TNF- α secreted upon co-stimulation of iDCs with *V. parvula* and the streptococci (except for *S. salivarius* strain 1 and 4) were considerably increased in comparison to those observed with mono-stimulations with the streptococci or *V. parvula* alone. Finally, while mono-stimulation with the *Streptococcus* strains induced variable amounts of IL-8 and generally low amounts of IL-6 production in DCs, and *V. parvula* mono-stimulation only induced these cytokines only in moderate amounts (Figure 6.5), co-stimulation of DCs commonly led to higher amounts of secreted IL-8, IL-10, and IL-6 (Figure 6.5; Table S6.2). The latter observation was especially obvious for co-stimulation with *V. parvula* and the *S. bovis* or *S. salivarius* strains 1, 3, and 4, which by themselves induced among the smallest amounts of IL-8 and IL-6 of all tested bacteria (Table S6.1 and S6.2), but in combination with *V. parvula* induced high amounts of these cytokines in iDCs (Figure 6.5; Table S6.2). Interestingly, the postulated synergy between *V. parvula* and the streptococci with respect to stimulation of production of IL-8, IL-6, IL-10, and TNF- α might again be relatively specific for these combinations of bacteria, as co-stimulation of iDCs with *V. parvula* and *L. plantarum* WCFS1 suppressed production of these cytokines, leading to the lowest IL-8, IL-6, IL-10, and TNF- α amounts observed in these co-stimulation analyses (Figure 6.5; Table S6.2). These observations suggest that immune cell stimulation with combinations of some streptococci and *V. parvula* leads to responses that are specific for the combined bacterial stimuli, including both (autocrine signalling-mediated) immune-suppressive and -synergistic effects that could not be predicted from respective mono-stimulations with either of the bacteria. Extrapolation of these *in vitro* immune (co-)stimulation profiles to the *in vivo* situation that encompasses the exposure of the immune system to bacterial communities rather than single strains is far from trivial.

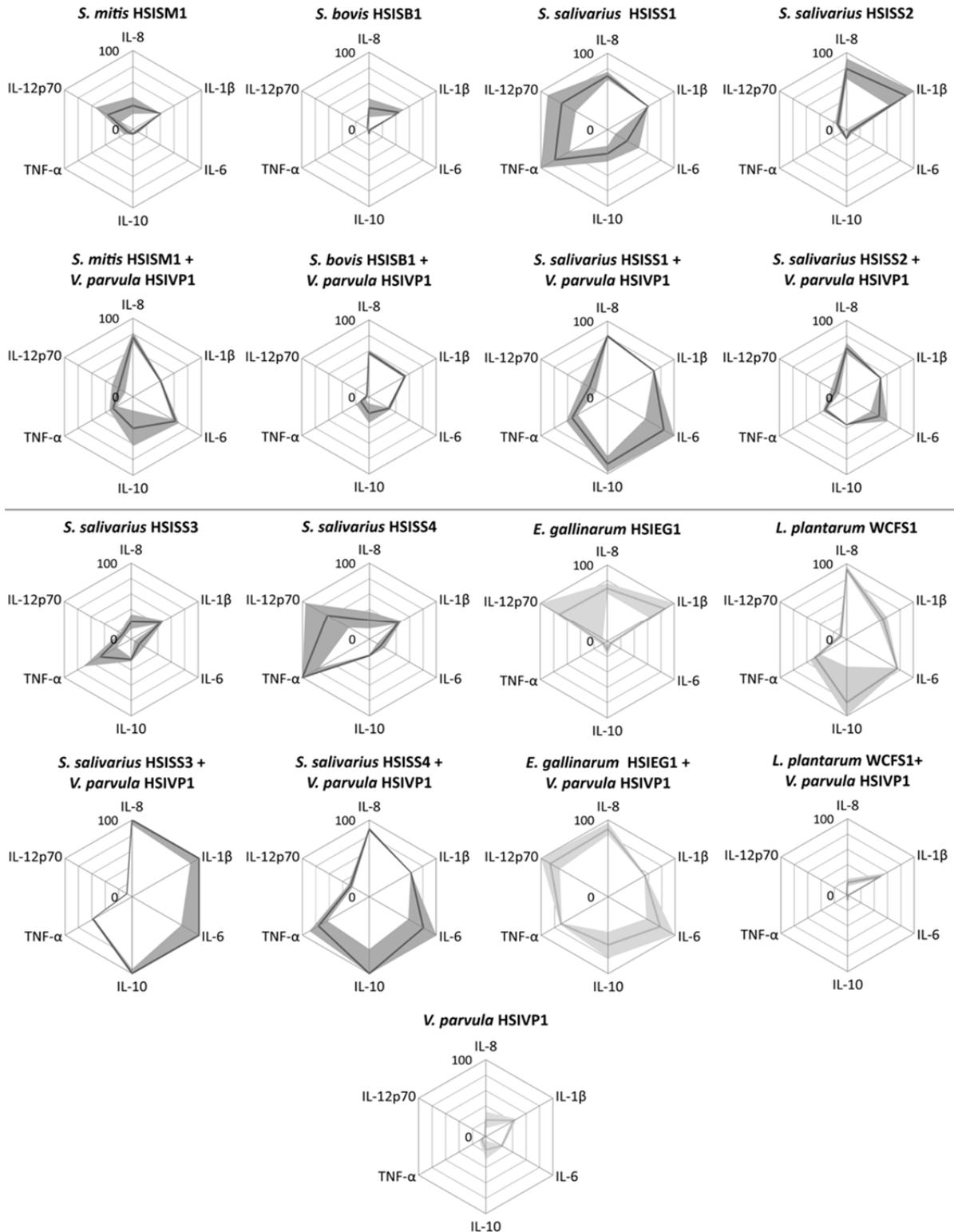


Figure 6.5. Cytokine secretion by dendritic cells derived from 2 human donors after mono-stimulation with tested strains and co-stimulation with *V. parvula*. Cytokine levels are expressed as relative values of the highest inducing strain (100% cytokine levels (pg/ml): IL-8: 30054 ; IL-1 β : 55; IL-6: 5451; IL-10: 155; TNF- α : 2613; IL-12p70: 300; Table S2). Lines represents the average secreted cytokine amounts and faded colours represent the interval between the upper and the lower SEM. SEM values higher than 100% are not visualized.

*: Cytokine responses determined using DCs derived from 2 different human donors

Nonetheless, certain trends could be seen in our *in vitro* results of streptococcal and *V. parvula* co-stimulation, suggesting at least a partial consistency in the co-stimulatory capacities of two species. This notion is further illustrated by the high similarity of the immune profiles elicited by co-stimulation with *V. parvula* and *S. salivarius* strain 1 or 4 (Figure 6.5), which is in good agreement with the close relatedness of these streptococcal strains ((340); Chapter 4 and 5; see also above). This observation indicates that immune-cell co-stimulations with different bacterial strains are robust and reproducible.

Discussion

Individual GI commensals (e.g. *Faecalibacterium prauznitsii* (301) and *Bacteroides fragilis* (312, 327)) affect the host immune system in specific ways (see (153) for a recent review). Given that the human small intestine is an important region to study host-microbe interactions, we evaluated the immunomodulatory properties of *Streptococcus*, *Veillonella*, and *Enterococcus* strains isolated from the small intestine. The strains used (especially valid for the streptococci (340); Chapter 4) can be regarded as representatives of distinct phylogenetic lineages that were identified among a large panel of isolates obtained from the human small intestine ecosystem. The *Streptococcus* strains tested here, have previously been subjected to in depth analysis, including physiological studies focussing on their carbohydrate utilizing capacities ((340); Chapter 4), but also the determination of their complete genome sequences (Chapter 5). The current study revealed that these *Streptococcus* strains differ significantly in their ability to elicit cytokine production responses in iDCs as well as their capacity to activate NF- κ B responses via TLR2/6. These findings corroborate previous reports that conclude that significantly different immunomodulatory properties can be observed in the comparison of closely related species (335) and strains (233). However, stimulation of iDCs with *S. salivarius* strain 1 and 4 induced similar amounts of different cytokines, which is in agreement with their highly conserved genetic content and physiological characteristics ((340); Chapter 4 and 5)). Among the strains tested, *S. salivarius* strain 2 was the least effective in activating and maturing responses in iDCs, but at the same time was identified as one of the strongest inducers of TLR2/6 signalling, which is likely due to the difference in phagocytosis capacity between dendritic cells and HEK293 cells or that bacterial components are shielding certain MAMPs. The cytokine responses of the small intestinal streptococci were quite similar to other *Streptococcus* strains, including pathogenic *S. suis* strains although these elicited higher IL-12 (up to 6948 pg/ml) in DCs (231). However, the small intestinal *Streptococcus* strains tested here are not known to be virulent, although remnants of the virulence genes were found in their genomes (Chapter 5). Similarly, remnants of virulence related genes were also encountered in the genomes of strains of the yoghurt-associated species *S. thermophilus* (28), suggesting that benign streptococci may share functions with related pathogens.

Compared to the streptococci, the small intestinal *E. gallinarum* strain appeared to be consistently more potent in inducing cytokine production in iDC and was one of the strongest inducers of TLR2/6 signalling, which is in agreement with earlier studies that report on the highly immune-stimulating capacities of enterococci (20, 287).

In contrast to the other streptococci tested in this study, DCs were relatively unresponsive to the *S. bovis* strain, which also induced negligible TLR2/6-mediated signalling. Interestingly, amounts of cytokines produced by DCs co-stimulated with *S. bovis* and LPS were higher compared to stimulation with LPS alone, indicating a synergistic immunostimulatory effect. The low immune response to *S. bovis* may therefore be due to the modification of conserved MAMPs reducing their capacity to signal through TLRs and NLRs or shielding effects (e.g. due to capsule polysaccharide). Close relatives of the *S. bovis* strain (e.g. *S. gallolyticus* subsp *gallolyticus* UCN34 (280)) have a less attractive reputation and are known to evade the host immune system and have been associated with GI tract malignancies (27). Notably, genome mining of the *S. bovis* strain (Van den Bogert, et al. Unpublished results; Chapter 5) revealed gene repertoires similar to the capsular operon encoded by *S. gallolyticus* subsp *gallolyticus* UCN34 (280) (data not shown), which was postulated to shield the bacterial cell from the host immune system (27). Further comparative analyses could elucidate the genetic relatedness (e.g. coding capacities for virulence factors) between the *S. bovis* strain tested here and potentially pathogenic close relatives.

As *Streptococcus* and *Veillonella* spp. have been found to co-occur in various microbial ecosystems associated with humans and are proposed to have metabolic interactions (84, 381), the small-intestinal isolates from both genera were tested in co-stimulation experiments. This study revealed that combinations of streptococcal and *Veillonella* strains elicited an immune response profile that was distinct from the profile expected on basis of the corresponding mono-stimulations. The IL-12p70 stimulatory effect of the *Streptococcus* strains appeared to be negated by the *Veillonella* strain, while (some) streptococci when combined with *Veillonella* substantially augmented IL-8, IL-6, IL-10, and TNF- α responses. Determining the exact mechanism underlying these co-stimulation effects is not trivial. Nonetheless, these observations may bear relevance for mucosal areas that encompass dense populations of immature DC that are stimulated by simultaneous stimulation by multiple bacteria derived from the intestine lumen (e.g., within Peyer's Patches' follicle areas), and where cross-talk between DCs may be an important determinant of the overall cytokine concentrations in the vicinity of these DC populations. Alternatively, if DCs (e.g., within the Peyer's Patches) take up multiple bacteria or multiple fragments of bacteria via M cells then the co-stimulation results may imply that this would result in different cytokine responses as compared to those predicted on basis of single strain immune profiles measured in *in vitro* models. Our current knowledge and understanding of these interactions within the microbiota community as well as their interaction with the host (immune) system is still in its infancy, underpinning the need for further mechanistic studies in this area. Immunomodulation analyses with a variety of well characterized bacterial isolates

from the microbiota, would be a good starting point to identify potential immunomodulatory effects (including immunosuppression) for members of the microbiota. Deciphering of the underlying molecular mechanisms and identification of the bacterial effector molecules is a necessary subsequent step to unravel the molecular basis for individual bacteria-immune interactions. Insights in these individual molecular mechanisms of interaction for various bacterial species and strains could accelerate the deciphering of the complex and multifactorial interplay between the microbiota and the host immune system *in vivo*. In addition, high resolution *in vivo* measurements of the molecular responses to specific microbes can complement mechanistic *in vitro* studies by providing the necessary *in vivo* support for the molecular mechanisms unravelled with the help of *in vitro* systems. Mono-association (or simplified community colonization) studies in gnotobiotic animal models could provide an attractive reductionist model to extrapolate *in vitro* findings to an *in vivo* situation (101, 137, 286, 303, 306, 369). Subsequent mono-association studies with derivatives of the same bacterial species or strains that lack one or more of their (immune) effector molecules could enable the *in vivo* establishment of the molecular interaction mechanisms proposed on basis of *in vitro* observations. As an example, approaches like this have elucidated how *B. fragilis* and its zwitterionic polysaccharide PSA are able to shape the host immune system (see (312, 327) for recent reviews). These reductionist *in vivo* and *in vitro* models offer a unique set-up to take the essential initial steps towards understanding the complexity of the interplay between the microbiota and the host in the intestine and its possible consequences for the overall physiology of the host organism, including its immune system status. Alternatively, the molecular responses elicited in the human intestine mucosa by specific bacteria can in some cases directly be determined *in vivo*, which is exemplified by the in depth analysis of transcriptional responses in the duodenal mucosa of healthy human volunteers upon the consumption of dietary lactobacilli (335, 336). Such measurements may serve to guide *in vitro* studies that aim to decipher the underlying molecular mechanisms. The latter approach has the considerable advantage that the starting point for the *in vitro* work is based on relevant *in vivo* observations in humans, and may therefore suffer less from the potentially poor extrapolation of molecular responses from animal models to humans. To this end, it would be of great interest to determine the transcriptional responses elicited in the human small intestine mucosa upon their exposure to the endogenous small intestinal streptococci, *Veillonella* and other bacterial groups residing in the lumen of this part of the human gut (31, 381). Such *in vivo* datasets could be employed to evaluate their possible alignment with the *in vitro* immunomodulatory observations described here.

Acknowledgements

This project was partially supported by the Netherlands Bioinformatics Centre (NBIC). We thank Eddy Smid for directing the TIFN project.

Supplementary information

Supplementary tables

Table S6.1. Average and SEM cytokine response values from monocyte derived iDCs* stimulated with bacterial strains.

Bacterial strain	IL-8	IL-1 β	IL-6	IL-10	TNF- α	IL-12p70
<i>S. mitis</i> HSISM1	8247 \pm 1229	16 \pm 3	1381 \pm 634	29 \pm 16	1457 \pm 691	1416 \pm 799
<i>S. bovis</i> HSISB1	5231 \pm 1783	17 \pm 4	161 \pm 54	7 \pm 2	86 \pm 30	115 \pm 53
HSISS1	16555 \pm 2115	41 \pm 5	2221 \pm 403	58 \pm 17	4732 \pm 1220	1076 \pm 358
<i>S. salivarius</i> HSISS2	17147 \pm 3018	41 \pm 8	1104 \pm 380	78 \pm 30	2058 \pm 811	556 \pm 219
HSISS3	10337 \pm 1718	32 \pm 4	696 \pm 132	45 \pm 10	2452 \pm 783	81 \pm 23
HSISS4	11406 \pm 2202	29 \pm 3	1690 \pm 400	46 \pm 16	4933 \pm 1134	1176 \pm 539
<i>V. parvula</i> HSIVP1	13547 \pm 4944	17 \pm 4	3869 \pm 1616	78 \pm 40	1969 \pm 1166	23 \pm 14
<i>E. gallinarum</i> HSIEG1	17598 \pm 2512	30 \pm 8	4775 \pm 2619	206 \pm 147	5151 \pm 2550	2397 \pm 1109
<i>L. plantarum</i> WCFS1**	27782 \pm 1065	30 \pm 3	4132 \pm 85	128 \pm 73	1229 \pm 202	31 \pm 9

*: iDCs were obtained from five different healthy human donors

** : Cytokine responses determined using DCs derived from 2 different human donors

Table S6.2. Average and SEM cytokine response values from monocyte derived iDCs* stimulated with bacterial strains with and without *V. parvula* co-stimulation.

Bacterial strain	IL-8	IL-1 β	IL-6	IL-10	TNF- α	IL-12p70	
<i>S. mitis</i> HSISM1	9076 \pm 3231	22 \pm 0	301 \pm 149	9 \pm 2	197 \pm 138	127 \pm 64	
<i>S. bovis</i> HSISB1	8377 \pm 3648	24 \pm 4	66 \pm 38	9 \pm 3	16 \pm 3	8 \pm 2	
Single stimulation <i>S. salivarius</i>	HSISS1	21125 \pm 1561	33 \pm 0	1597 \pm 983	49 \pm 16	2063 \pm 576	239 \pm 74
	HSISS2	23782 \pm 3977	49 \pm 14	280 \pm 200	18 \pm 1	151 \pm 74	46 \pm 12
	HSISS3	6933 \pm 2972	24 \pm 3	666 \pm 414	41 \pm 4	1208 \pm 644	44 \pm 14
	HSISS4	7571 \pm 3399	24 \pm 2	931 \pm 322	33 \pm 2	2613 \pm 653	219 \pm 113
<i>V. parvula</i> HSIVP1	6497 \pm 3187	23 \pm 2	1301 \pm 97	26 \pm 19	158 \pm 91	4 \pm 4	
<i>E. gallinarum</i> HSIEG1	20916 \pm 2055	47 \pm 13	279 \pm 119	18 \pm 10	146 \pm 77	248 \pm 224	
<i>L. plantarum</i> WCFS1	27782 \pm 1065	30 \pm 3	4132 \pm 85	128 \pm 73	1229 \pm 202	31 \pm 9	
<i>S. mitis</i> HSISM1+							
<i>V. parvula</i> HSIVP1	22783 \pm 1981	22 \pm 1	3391 \pm 260	62 \pm 35	760 \pm 139	73 \pm 34	
<i>S. bovis</i> HSISB1+							
Co-stimulation <i>V. parvula</i> HSIVP1	HSISS1	17253 \pm 933	29 \pm 1	1621 \pm 121	33 \pm 19	315 \pm 183	10 \pm 4
	HSISS2	24158 \pm 606	38 \pm 0	4604 \pm 1537	135 \pm 16	1391 \pm 229	94 \pm 36
	<i>S. salivarius</i> + HSISS2	19222 \pm 2100	28 \pm 0	2637 \pm 692	54 \pm 0	838 \pm 87	58 \pm 22
	<i>V. parvula</i> HSIVP1 HSISS3	30054 \pm 1937	55 \pm 7	5451 \pm 1446	154 \pm 9	1516 \pm 29	26 \pm 1
HSISS4	26390 \pm 670	34 \pm 0	4404 \pm 1241	155 \pm 50	1982 \pm 347	98 \pm 11	
<i>E. gallinarum</i> HSIEG1+							
<i>V. parvula</i> HSIVP1	26498 \pm 2349	30 \pm 1	4228 \pm 1129	97 \pm 26	1834 \pm 35	300 \pm 49	
<i>L. plantarum</i> WCFS1+							
<i>V. parvula</i> HSIVP1	4954 \pm 1417	27 \pm 4	35 \pm 24	9 \pm 6	5 \pm 5	4 \pm 4	

*: iDCs were obtained from two different healthy human donors



Chapter 7

General discussion

General discussion

The different anatomical regions of the human gastrointestinal (GI) tract harbor distinct bacterial communities that vary in density and diversity (203, 341). The fecal microbiome composition, reflecting the distal end of the colon, has been associated with host health and the dynamics of the microbial system residing in this readily accessible region of the GI tract has been well documented (66, 203, 382). However, similar information is largely missing for the small intestinal microbiome, while its influence on host physiology is highly plausible (58). Therefore, the work in this thesis aimed at providing a better understanding of the composition and dynamics of the human small intestinal microbiota and to provide insight in the metabolic potential as well as immunomodulatory properties of some of its typical commensal inhabitants.

Evaluation of ileostomy subjects as small intestinal model system

The small intestine, located between the stomach and the colon, is poorly accessible, and sample collection from this part of the human intestine for characterization of its microbiota is challenging (32). Ileostomy subjects, employed in the work described in this thesis, have their terminal ileum connected to an abdominal stoma as a result of colectomy. This stoma allows non-invasive and repetitive sampling of the contents of the ileum (31, 32). Moreover, the microbiota present in ileostoma samples is likely to correspond better to the small intestine samples from an individual with a normal GI tract, as compared to samples collected during (emergency) surgery (3) or samples collected from sudden death victims at autopsy (128).

Ileostoma effluent has been used since the 1960s in studies focusing on the small intestinal microbiota and revealed a high number of lactic acid bacteria (e.g. streptococci) relative to fecal samples based on estimations from classical plating (94, 111). Later studies, employing culture-independent technologies, confirmed the abundance of lactic acid bacteria in ileostoma effluent (31, 125). The ileostoma does seem to be potentially vulnerable to oxygen penetration, leading to disruption of the normal anaerobic environment (125). However, the results described in **chapter 3** and earlier findings from our laboratory (31, 381) showed that the small intestine of ileostomy subjects supports substantial abundances and activity of strict anaerobes (e.g. *Clostridium*, *Coprococcus*, *Veillonella*, *Turicibacter*, and members of the Peptostreptococcaceae). Nonetheless, in-depth characterization of small intestinal samples from individuals with a normal GI tract is essential to define a “healthy” human small intestinal microbiota. Such samples were successfully obtained using an intraluminal naso-ileal catheter of which the tip is placed nasogastrically and moves to the small intestine by peristaltic movement. In this way, luminal fluid could be obtained by aspiration following flushing of the small intestine with physiological salt solution through a port of the catheter. However, it is important to note that consecutive flushing of the small intestinal lumen drastically influenced the outcome of microbial profiling (Figure 7.1, Van den Bogert, et al. Unpublished results), suggesting that frequency and volume of flushing should be minimized to obtain the

information about the microbial composition that resides in the ‘normal’ small intestine. Notably, subjects are overnight fasted before the positioning of the catheter in the small intestine, which may take as much as 6 hours. This procedure requires gastroenterologist supervision, and its position needs to be verified by short-interval fluoroscopic control. As a consequence, the catheter sampling methodology is quite invasive and is not suitable for the collection of multiple samples from the same individual over time, which is required to address the effect of host diet on microbiota diversity and dynamics. Nevertheless, the samples obtained by this method enabled the comparative analysis of microbiota composition, revealing that jejunal and proximal-ileum samples resembled the microbiota in ileostoma effluent, while samples obtained from the terminal ileum were more similar to those obtained from the colon or fecal origin. This analysis indicated that the ileostoma effluent appears to be a suitable reflection of bacterial communities that can be encountered in the proximal regions of the small intestine (341, 381) (see also figure 1.3, **Chapter 1**).

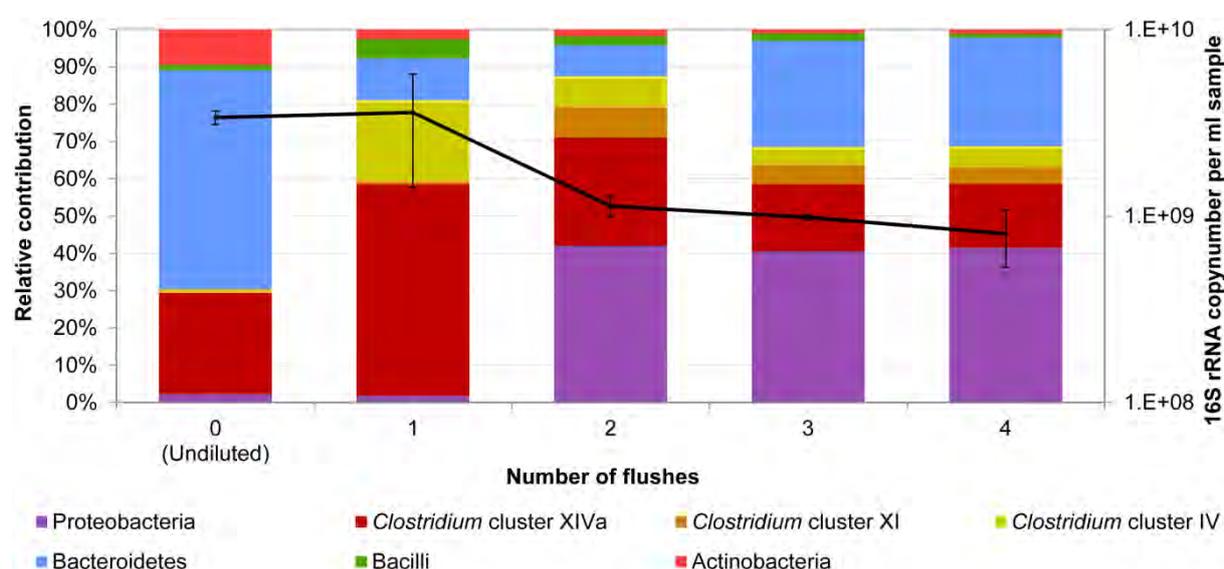


Figure 7.1. Relative contribution of detected phylogenetic groups and 16S rRNA copy numbers in ileal fluid samples collected after consecutive flushing of the small intestine with 10 ml physiological salt solution through a port of an intraluminal naso-ileal catheter. Profiles were generated by phylogenetic microarray analysis using the Human Intestinal Tract Chip (HITChip) (266). In the legend, phylogenetic groups that contribute at least 1% to one of the profiles are indicated. The 16S rRNA copy numbers were quantified by total qPCR as describe in **chapter 2** (Van den Bogert, et al. Unpublished results).

A recent innovation in the area of micro-electronics, the IntelliCap (<http://www.research.philips.com/initiatives/intellicap/index.html>), enables delivery of drugs or food components to specific sites in the GI tract as well as has sensors to measure pH and temperature. This technology is currently being adapted for *in vivo* sampling and has great potential to collect (small) intestinal samples in a non-invasive manner.

In conclusion, the catheter sampling methodology is instrumental for the determination and definition of the “healthy” small intestinal microbiota. The ileostomy

model, on the other hand, provides a suitable *in vivo* model system that offers non-invasive and repetitive sampling of luminal effluent, enabling the analysis of microbial community dynamics of the small intestinal microbiota over time, or as a result of dietary interventions.

Advanced typing of intestinal microbiota

Technological as well as computational advances during the past decades provided a suite of methodologies that enabled culture-independent and comprehensive analysis of the microbial ecology of the GI tract (see (219) and (358) for recent reviews). **Chapter 2** describes the comparison of two popular technologies, barcoded 454-pyrosequencing and phylogenetic microarray analysis using the HITChip (266). Both technologies generated consistent microbial composition profiles for samples obtained of terminal ileum and fecal origin, reiterating that both technologies facilitate robust microbial profiling and essentially generate equivalent biological conclusions in terms of microbial composition. However, analogous to other phylogenetic microarrays (49, 219, 339), the HITChip analyses offer a broader dynamic range of detection relative to the commonly applied depth of analysis in 454-pyrosequencing-based microbial profiling. Therefore, HITChip analysis appears to remain the preferred methodology for profiling of the microbiota in more distal regions of the human GI tract. In contrast, the comparison of microbiota composition profiles generated by HITChip analysis and pyrosequencing using ileostoma effluent samples revealed a relatively poor correlation between the two methods. Since the small intestinal microbiota is less well characterized (see above), this discrepancy is most likely attributable to the fact that microarrays are constrained to the detection of phylogenetic groups that were included in the database used for microarray-probe design. By contrast, 454-pyrosequencing enables *de novo* community profiling, and was therefore used for subsequent profiling efforts of the small intestinal communities in this thesis (**Chapter 3**). To enhance the depth of analysis of sequence-based *de novo* microbiota profiling, the Ion Torrent and Illumina sequencing platforms could be considered. Though the former is in its infancy and produces shorter read-lengths relative to 454-pyrosequencing (~200 bp), it is becoming an attractive technology for 16S rRNA based ecological surveys due to the substantially lower cost (108, 160, 185, 362). The Illumina sequencing platform produces a number of sequence reads that is several orders of magnitude higher as compared to 454-pyrosequencing and thus offers a greater depth of analysis. However, the current relatively short sequence read-length generated by the Illumina platform provides only limited reliability in subsequent taxonomic classifications and diversity assessments, which is further compromised by the increasing sequence-error rates for reads longer than 60 bp (39, 50). The continuous improvement of sequence read-length quality for this sequencing platform may change this situation in the upcoming years.

Community profiling approaches, including those mentioned above, that are PCR-based strongly depend on the phylogenetic coverage of oligonucleotide primers used for amplification (16, 98). This notion is supported by the assessments of the coverage published for 'universal' 16S rRNA primers (16, 127, 138, 341) as well as

by our findings in **chapter 2**, which demonstrated that a widely used primer (27F-Nondeg) does not adequately amplify the 16S rRNA gene of bifidobacterial origin. Nevertheless, our study demonstrated that the choice of primer did not strongly affect the overall outcome of microbial profiling and did not consistently skew richness and diversity estimates (**Chapter 2**), although it should be noted that the four primer pairs used in our survey targeted the same region of 16S rRNA genes. This notion has been extended by several comparative studies that indicated that the region of the 16S rRNA gene is a critical determinant for the outcome of microbial composition profiling as well as estimates of ecological metrics (37, 88, 186). At present, a consensus with respect to the most suitable 16S rRNA gene region for microbial ecosystem fingerprinting in environmental samples is still lacking (50, 120). As a consequence, direct comparison of the results obtained in different studies are hampered by the different regions of the 16S rRNA gene they targeted. This issue may be overcome by further advances in sequence technologies that provide longer sequence-read lengths, such as the XL+ (up to 1,000 bp and beyond) and XLR70 (up to 600 bp) chemistry kits for the GS FLX+ platform (Roche; www.454.com), although these kits have not yet been tested for 16S rRNA based community profiling. The Pacific Biosystems (PacBio) sequencing platform offers similar sequence-read lengths, but the high error-rate of this technology hampers appropriate data interpretation (108).

With the application of pyrosequencing in molecular ecology came the need for visualization and interpretation software suites to handle the vast datasets that are obtained. The Quantitative Insights Into Microbial Ecology (QIIME; see **chapter 3** and **4**) pipeline (38) and Mother (290) are two popular analysis pipelines that integrate preexisting software tools (e.g. the Ribosomal Database Project (RDP) classifier (355)) into a single workflow that enables assessment and visualization of the microbial composition and diversity analysis within and between samples using numerous ecological metrics, including the phylogeny-based distance metric, UniFrac (210, 211). Microbiologists are facing an important challenge to keep these software suites up-to-date to meet the advances in sequencing technologies and complexity of ecological studies.

The human small intestinal microbiota

The diversity and dynamics of the small intestinal microbiota were investigated through 16S rRNA (gene) based community profiling using ileostoma effluent from several subjects (**Chapter 3** and **4**). In line with earlier findings from our laboratory (31, 339, 381), the subjects harbored a personal ileostoma effluent microbiota that displayed pronounced compositional fluctuations during the course of several days and even between samples obtained within a day (**Chapter 3**). Fluctuations were observed at the genus and species level, but also expand to the genetic lineage level as was demonstrated for members of the *Streptococcus* genus (**Chapter 4**). These observations contrast the more stable microbiota composition encountered in fecal samples (229, 266, 376), which is commonly dominated by members of the genera *Bacteroides*, *Prevotella*, and the *Clostridium* clusters IV and XIVa (11, 31, 266, 381).

There appears to be an ileostoma core microbiota consisting of few bacterial genera, including the *Streptococcus* and *Veillonella* (**Chapter 3**), of which the former can display a substantial level of phylogenetic richness in individuals (**Chapter 4**). These genera were also commonly detected in the active fraction of ileostoma community, indicating that they are typical active inhabitants of the human small intestine (**Chapter 3**). Next to the mentioned core microbiota members, most ileostoma samples (but not all) contained representatives of the genera *Actinomyces*, *Clostridium*, *Escherichia*, *Haemophilus*, and *Turicibacter*, which could actually constitute a sizeable fraction of the overall population in specific individuals, as was the case for members of the genus *Escherichia* (**Chapter 3**). The concept of the existence of a phylogenetic core in the small intestinal microbiota should be explored using a larger group of (ileostomy) subjects, but is unlikely to be sustained when individuals with different health status, geographic origins, and/or of different age groups are included (370) (see **chapter 1** and (212) for a recent review). It is more likely that an individual small intestine microbiota core is existent, which is the fraction of the microbiota that is consistently present within a specific niche associated with a particular healthy individual, and which has previously been shown to exist for fecal microbial communities (267). Such an individual core is supported by the findings described in **chapter 4** and **5**, showing that subjects harbour the same streptococcal lineages for at least a year.

The microbiota in the human small intestine benefits from a high nutrient availability derived from the host's diet, which can sustain a microbiota of which the greater part actively participates in this microbial ecosystem, and thus may affect the host's mucosal functions. This notion is supported by the strong congruence between the phylogenetic profiles of the total and active fractions within the ileostoma microbiota as was assessed by pyrosequencing of 16S rDNA and rRNA, respectively (**Chapter 3**). This appears to be a characteristic that is specific for the small intestinal microbiota, since substantial compositional differences were reported between the active and total communities (257) as well as the viable (active) and injured or dead (non-active) communities (18) within the fecal microbiota. Further investigation of the activity of the small intestinal microbiota by metatranscriptome analysis of ileostoma effluent from different subjects revealed host-specific activity patterns. However, there was a strong conservation in expressed functions between subjects, supporting the notion of a functional core (188, 329). A considerable fraction of the shared expressed genes belonged to typically conserved functions associated to translation, ribosomal structure, and biogenesis as well as functions related to transport and metabolism of 'simple' carbohydrate substrates. The latter functions were predominantly assigned to the streptococcal members of the microbiota in samples where this genus was highly predominant, but in samples where the streptococci were less abundant was shown to be taken over by other members of the microbial community (e.g. members of the *Escherichia* genus; Leimena and Van den Bogert, et al. Unpublished results). The latter supports the importance of the capacity to effectively ferment diet-derived simple carbohydrate substrates for the small intestinal microbes, which is not necessarily restricted to, but commonly predominated by a

single phylogenetic group like the members of the common *Streptococcus* genus (**Chapter 3**; Leimena and Ramiro-Garcia, et al. Submitted).

Refined analysis of the representatives of the *Streptococcus* genus revealed that the small-intestinal streptococci predominantly belong to the *S. mitis*, *S. bovis*, and *S. salivarius* species groups and that they exhibited a substantial divergence in their carbohydrate import and utilization capacities (**Chapter 4** and **5**). Thereby, the functional characteristics of these streptococci reflect their adaptation to the small-intestinal habitat, in which a variety of diet-derived carbohydrates can be encountered. *Veillonella* do not have the capacity to ferment these primary carbohydrates, but seem to be important members of the small intestine as they are renowned utilizers of lactic acid as a carbon and energy source, and thus seem to depend on the lactate production by the carbohydrate fermenting streptococci (**Chapter 3** and **5**). This consistent metabolic contribution by *Veillonella* also requires less flexibility of this population, which may explain why their populations display a reduced level of diversity or richness (**Chapter 4**).

The findings presented in **Chapter 3**, **4**, and **5** demonstrate that comprehensive microbiota analysis necessitates ecosystems biology approaches (264) that combine data from “(meta-) omics” approaches to provide further insights into the function and biodiversity of the intestinal microbiota and interactions between its community members. The wealth of knowledge this provides can be utilized to predict the effect of environmental changes, such as a change of diet, on the intestinal microbiota composition, and function, as well as their metabolic output as a community, which may impact on the host mucosal functions. As the human small intestine represents the first region where food components and the intestinal microbiota meet, this region of the intestinal tract appears to be an important region to study these host-microbe interactions (see below).

Phylogenetic relatedness between small intestinal and oral cavity derived *Streptococcus* and *Veillonella* populations

The abundant streptococci and *Veillonella* in the small intestine and other sections of the upper GI tract (1, 9, 23, 256) could originate from their populations in the oral cavity (353). To support this suggested origin, a total of 120 bacterial isolates were obtained from saliva, teeth, and tongue samples from the same ileostomy subject employed in **chapter 4**. The oral bacterial isolates were subsequently assigned to the genera *Streptococcus* [93], *Veillonella* [26], and *Lactobacillus* [1] (Table 7.1), which were each represented by a single phylotype. Genetic fingerprinting (**Chapter 4**) differentiated the oral streptococci into 5 lineages, 2 of which were common in all of the obtained oral samples and the remaining lineages were obtained from either saliva (1 lineage) or tooth surfaces (2 lineages). However, none of the saliva, teeth, and tongue samples contained all of the *Streptococcus* lineages. This supports the notion that anatomic sites and their associated micro-environments support a distinct microbial community (57, 147), and site specific sampling is required to adequately describe the entire oral microbiota and its localized distribution.

Table 7.1. Number of *Streptococcus*, *Veillonella* and *Lactobacillus* isolates per phylotype, lineage, and sample origin based on identification employing 16S rRNA gene classification, AFLP analysis and (GTG)5-PCR fingerprinting^a

16S rRNA gene classification (genus)	Phylotype	Lineage	Species group identification	Total isolates ^b	Total isolates per sample origin							
					ileostoma effluent t = 0 ^c	ileostoma effluent t = 1 ^c	Oral cavity t = 1 ^d	Saliva	Teeth	Tongue		
<i>Streptococcus</i>	1 ^e	1	<i>S. mitis</i>	3	3							
	2 ^e	2	<i>S. bovis</i>	17	13	4						
	3 ^e	3			64		64					
		4		<i>S. salivarius</i>	1 ^h	1						
		5			1			1			1	
		6			4			4				
		7			24	1		23			23	
		8			44	8		36	12		4	20
		9			95	66		29	14		3	12
<i>Veillonella</i>	4 ^f	N.D. ^g	<i>V. parvula</i>	63	21	16	26	9	6	11		
<i>Lactobacillus</i>	8 ^f	N.D. ^g	<i>L. fermentum</i>	5		4	1			1		

^a: A graphic representation of the data included in this table is added as figure 7.2

^b: All isolates were obtained from the same ileostomist at 2 time points (t = 0 and, one year later, at t = 1) (**Chapter 4**; Table 1; subject 1)

^c: Data based on **chapter 4**

^d: The near full-length 16S rRNA gene sequences were deposited in the Genbank database and are available under accession numbers JQ680146 to JQ680198

^e: Isolates obtained from *Mitis Salivarius* agar

^f: Isolates obtained from *Veillonella* Selective Agar

^g: N.D. Not determined because AFLP analysis and/or Rep-PCR genomic fingerprinting did not reveal discriminative lineages

^h: AFLP analysis identified a single *Streptococcus* isolate as a separate genomic lineage, while Rep-PCR genome fingerprinting grouped this isolate to *Streptococcus* lineage 9 (**chapter 4**).

Although the number of samples that was obtained for this study was limited, these findings are in line with results published by Zaura, et al. (372), where it was shown that the microbial composition of three individuals differed between samples obtained from various sites in the oral cavity. Similarly, studies by Aas, et al. (1) and Diaz, et al. (71) demonstrated that some bacterial species in the oral cavity are common to most intra-oral sites, while others are site specific. Remarkably, the oral bacterial isolates cluster together with *salivarius* group streptococci, *Veillonella parvula*, and *L. fermentum* that were isolated from ileostoma effluent (**Chapter 4**). Among the 6 genetically distinct lineages of the *Streptococcus salivarius* phylotype that were identified, one lineage was only obtained from ileostoma effluent while another two were only isolated from the oral cavity. Notably, the remaining three lineages appeared to be present in both ecosystems (Table 7.1; Figure 7.2 and 7.3). This suggests that the microbiota of the small intestine and the oral cavity both contain distinct as well as shared microbial lineages when they are classified at the sub-phylotype level.

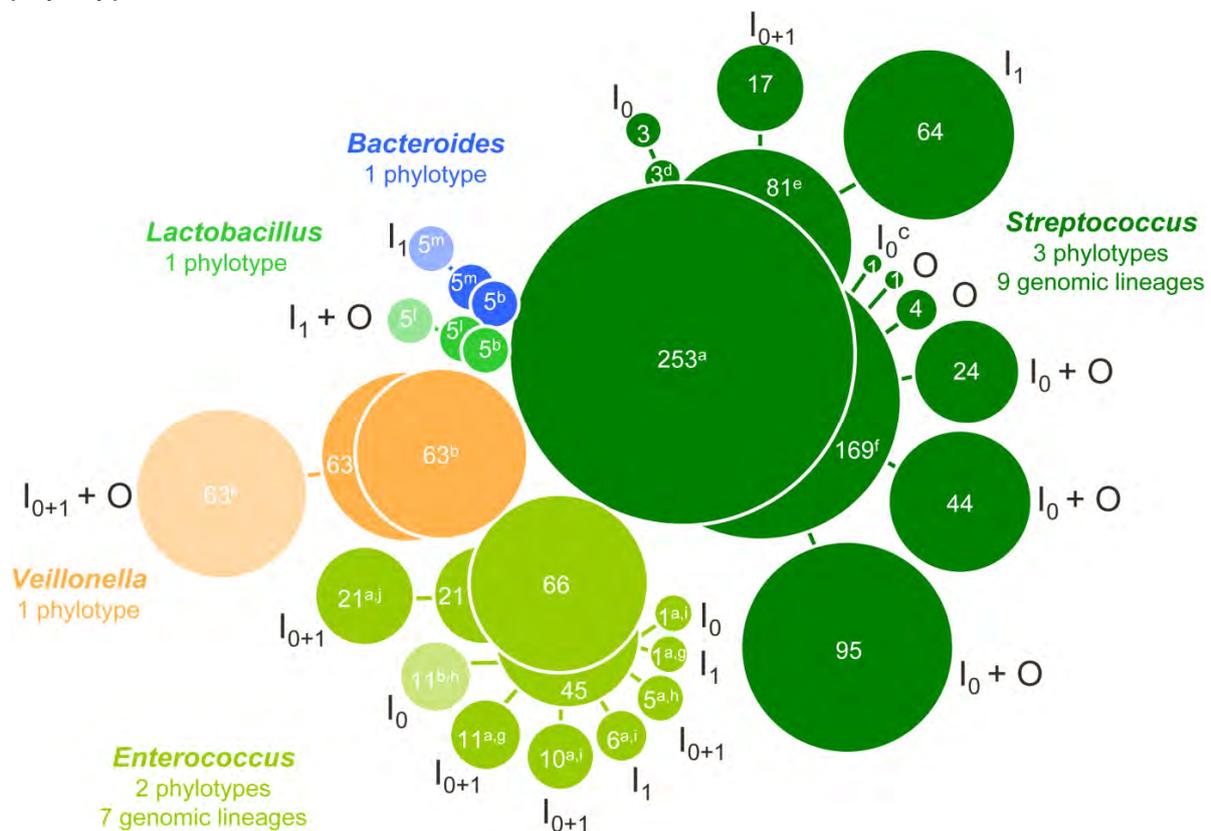


Figure 7.2. Groupings of *Streptococcus*, *Enterococcus*, *Veillonella*, *Lactobacillus*, and *Bacteroides* spp. (inner circles) isolates obtained from ileostoma effluent and the oral cavity into phylotypes (middle circles) and genomic lineages (outer circles). Numbers in faded colored groupings represent bacterial isolates for which AFLP analysis and/or Rep-PCR genomic fingerprinting did not reveal discriminative lineages. I₀/I₁: isolates obtained from ileostoma effluent at time point 0 and/or 1. O: isolates obtained from oral cavity. a: isolates obtained from *Mitis Salivarius* agar (**Chapter 4**); b: isolates obtained from *Veillonella* selective agar ((274, 275); Chapter 4); c: AFLP analysis identified a single *Streptococcus* isolate as a separate genomic lineage, while Rep-PCR genome fingerprinting grouped this isolate to *Streptococcus* lineage 9 (Table 7.1; Figure 7.3; **Chapter 4**); d: *S. mitis* group; e: *S. bovis* group; f: *S. salivarius* group; g: *E. faecium* group; h: *E. gallinarum* group; i: *E. avium* group; j: *E. faecalis*; k: *V. parvula*; l: *L. fermentum*; m: *B. fragilis*.

Although speculative, this ‘genomic-lineage’ level comparison between the oral and small intestinal microbiota indicates that bacteria do not necessarily persist exclusively in a specific body habitat, but can also pass from one body site to another within their host. This observation may suggest that the *Streptococcus* and *Veillonella* populations are not necessarily autochthonous members of the small intestine microbiota (353), but the observation that they are not only abundantly present but also highly active (**Chapter 3 and 4**), implies that these *Streptococcus* populations do play a prominent role in the primary carbohydrate metabolism occurring in the small intestinal ecosystem.

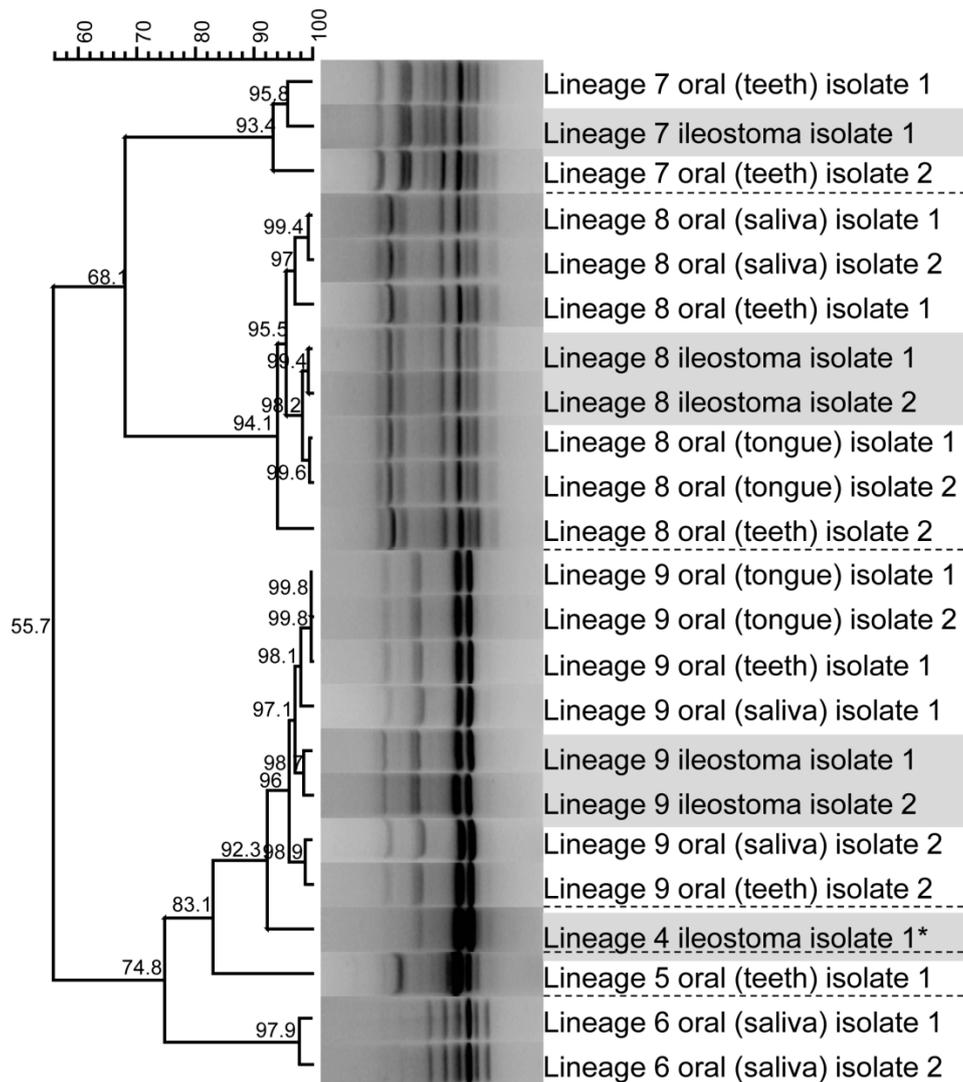


Figure 7.3. UPGMA clustering of Rep-PCR profiles with Pearson correlation coefficients from randomly selected *salivarius* group streptococci lineages cultured from ileostoma effluent and the oral cavity. Similarities among profiles were calculated with an optimization of 0.69%. Based on analysis of replicates with the Rep-PCR protocol, 84% similarity was used as a threshold for the discrimination between individual genetic lineages (data not shown). Numbering of lineages is based on table 1. *: AFLP analysis identified a single *Streptococcus* isolate as a separate genomic lineage, while Rep-PCR genome fingerprinting grouped this isolate to *Streptococcus* lineage 9 (Table 7.1; Figure 7.2; **Chapter 4**).

Screening for novel small intestinal bacteria

In agreement with the earlier findings by Booijink, et al. (Booijink 2010), the discordant profiles from pyrosequencing and HITChip analysis for ileostoma effluent (see above and **Chapter 2**), indicated that the small intestinal ecosystem harbors a range of bacteria with 16S rRNA gene sequences that have not been detected previously in the human gut. This notion was further explored employing classical cloning and sequencing of random 16S rRNA genes from ileostoma effluent samples (**Chapter 2**). Though fewer than expected (31), sequences from novel intestinal phylotypes were obtained that mainly belonged to *Veillonella* spp.. The lower number of novel intestinal phylotypes may be explained by the fact that this approach can be biased by DNA extraction procedure, coverage of primers used in amplification of the 16S rRNA gene, differential PCR amplification, and/or cloning efficiencies. This may especially be true when one aims to amplify and clone 16S rRNA genes of lowly abundant microbial communities (151, 378). The cloned 16S rRNA gene sequences of the novel intestinal *Veillonella* phylotype showed $\leq 97.2\%$ identity to those from type strains represented in the SILVA reference database release 111 (261). Although not the primary objective of the study, the selective cultivation conditions employed in **chapter 4** failed to obtain isolates of this novel intestinal *Veillonella* phylotype from ileostoma effluent (Van den Bogert, et al. Unpublished results). One explanation for this apparent discrepancy could be that the cloned 16S rRNA gene sequences were derived from dead cells that originate from the upper GI tract (353). However, isolates from the novel intestinal *Veillonella* phylotype were also not obtained from oral samples (see above) that were obtained from the same ileostomy subject (Van den Bogert, et al. Unpublished results). An alternative explanation may be that it is well known that oxygen sensitivity and culture conditions vary between anaerobic bacterial species (295). The novel *Veillonella* phylotype may have been selectively eliminated from cultivation due to different nutrient requirements or by detrimental oxygen exposure in the ileostoma bag and/or during transport to the laboratory. One alternative to obtain novel intestinal isolates, which circumvents molecular approaches, was recently introduced by Lagier, et al. (190) and termed microbial culturomics. Analogous to other “omics”-approaches (see **Chapter 1** and (341, 382) for reviews), culturomics gains insights into the microbial composition by high-throughput comprehensive culture-dependent methods (114). In a proof-of-concept focusing on 3 fecal samples, culturomics by means of 212 different culture conditions enabled cultivation of an astonishing 32,500 bacterial colonies that could be differentiated into 340 species. These included 174 species that, to date, have not been associated with the human GI tract, and encompassing 31 new species (190). This shows that this approach complemented by innovative cultivation technologies (**Chapter 1**) have great potential to isolate rare and so far uncultured bacteria. However, it should be noted that this is time-consuming, laborious, and costly and at the same time may not provide a reliable representation of the overall microbial composition in the original samples. The latter is related to cultivation biases associated with the use of selective media that favor growth of specific microbial groups (**Chapter 4**; (76, 239)).

Exploring the uncultured fraction of microbial communities is by no means a trivial endeavor, but the new concept of culturomics and other cultivation technologies (**Chapter 1**) combined with molecular approaches, offers avenues to study the physiological characteristics and functional properties of rare and so far uncultured species, and may shed light on the (host-) microbial relationships that shape the (small intestinal) ecosystem.

Metabolic relationship between small intestinal community members

In agreement with metagenome analyses of small intestinal microbial communities (381), small intestinal *Streptococcus* isolates displayed a substantial variation in their encoded carbohydrate transport systems and metabolic capacities dedicated to the degradation of simple carbohydrates (**Chapter 4 and 5**). Furthermore, genes related to lactate fermentation (e.g. lactate-import permease) encoded by *Veillonella* were commonly expressed in ileostoma effluent (**Chapter 3**; Leimena, et al. Unpublished results), supporting the previously postulated notion (381) that small intestinal *Veillonella* may utilize the fermentation end-products of streptococci as energy source. The draft genome sequence was determined for the small-intestinal *V. parvula* isolate we were able to culture (**Chapter 4**) to investigate its metabolic capacity with a special focus on lactate utilization.

Classification of the genes encoded by the *V. parvula* genome to Cluster of Orthologous groups (COG; (321)), revealed that metabolic functions accounted for 28.3% of the *V. parvula* genes. Bearing in mind that most *Veillonella* species cannot ferment carbohydrates (65) it is no surprise that the genome of the *V. parvula* strain encoded less genes assigned to carbohydrate transport and metabolism (2.4%) compared to the small intestinal streptococci (3.5-6.9%; Van den Bogert, et al. Unpublished results). Notably, relative to the small intestinal *Streptococcus* genomes, a large fraction of the *V. parvula* genome was assigned to functions in energy production and conversion. The latter COG category encompasses most of the genes necessary for conversion of lactate to propanoate (Figure 7.4). This pathway also includes the characteristic methylmalonyl-CoA decarboxylase that is composed of five different polypeptides α , δ , ϵ , γ , β (132, 141). The α and β subunits function as a carboxyltransferase and decarboxylase, respectively (133). Furthermore, the β subunit is tightly bound to the membrane and presumably involved in Na^+ translocation (131, 141). The γ subunit is a biotin containing protein. The δ subunit is necessary for assembly and functioning of the methylmalonyl-CoA decarboxylase complex (140). The ϵ subunit has a high sequence homology to the C-terminal end of the δ subunit and is likely the result of a gene duplication event, but other than stabilizing the enzyme complex has no known catalytic function (140, 141). The methylmalonyl-CoA decarboxylase complex generates a transmembrane electrochemical (Na^+) gradient (131). *Propionigenium modestum* utilizes a similarly established sodium gradient for ATP synthesis employing a Na^+ translocating ATP synthase. However, *Veillonella* spp. are not known to have such a system and possibly use the sodium gradient for lactate import analogous to citrate transport by *Klebsiella pneumoniae* (33, 72, 131).

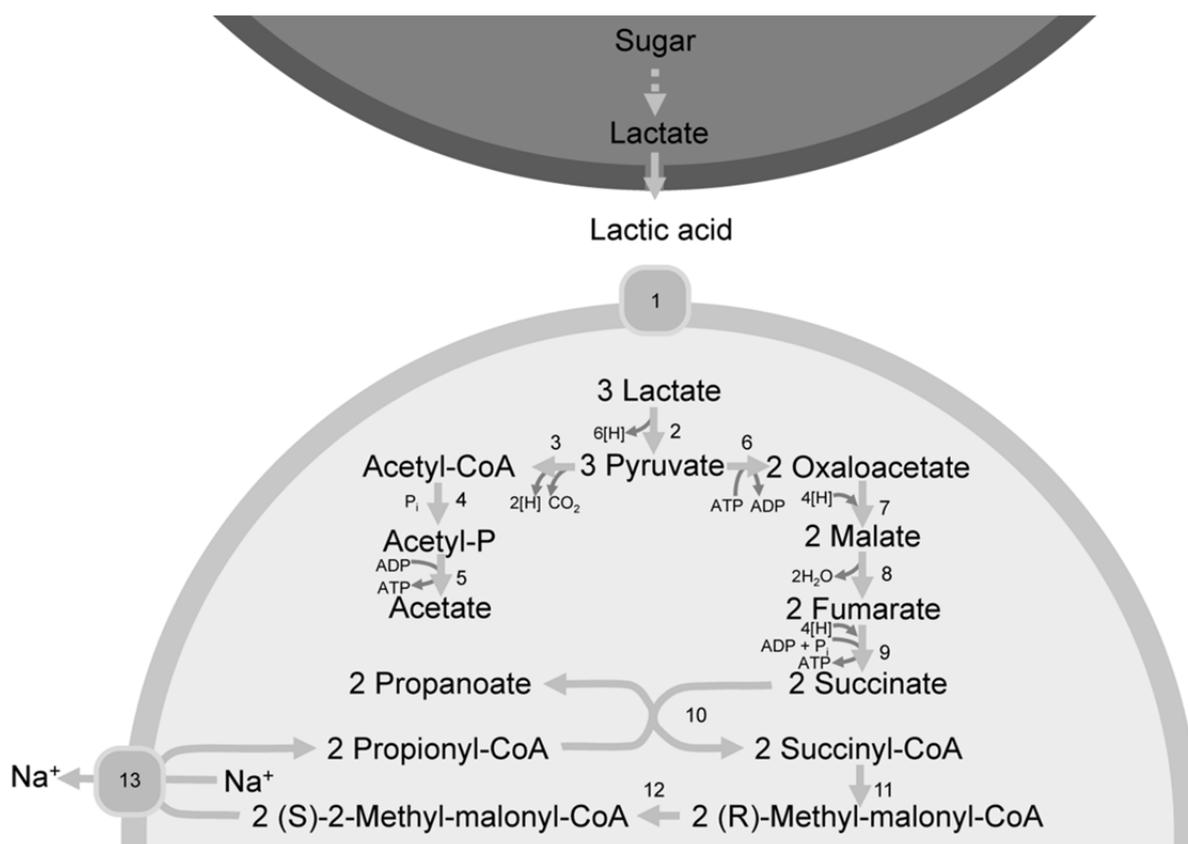


Figure 7.4. Metabolic pathway for propanoate fermentation by *V. parvula* HSIVP1

Enzymes: 1, Lactate permease; 2, Lactate dehydrogenase (HSIVP1_1868); 3, Formate lyase (HSIVP1_1027); 4, Phosphotransacetylase (HSIVP1_1771); 5, Acetate kinase (HSIVP1_664); 6, Pyruvate carboxylase (HSIVP1_150); 7, Malate dehydrogenase (HSIVP1_1554); 8, Fumarase (HSIVP1_1491); 9, Succinate dehydrogenase (HSIVP1_1758/1759); 10, Succinate-CoA transferase (HSIVP1_684); 11, Methylmalonyl-CoA mutase (HSIVP1_1810/1811/682/683); 12, Methylmalonyl-CoA racemase (HSIVP1_680); 13, Methylmalonyl-CoA decarboxylase (HSIVP1_675/676/677/678/679).

The metabolic relationship (food-chain) between small intestinal streptococci and *Veillonella* may be part of a more elaborate relationship between these bacteria. *In vitro* investigations, using members of both genera (*S. gordonii* and *V. atypica*) found in the oral cavity in dental plaque biofilms, revealed that *V. atypica* produces a diffusible signal that, through a pathway involving the transcriptional regulator *ccpA*, enhances expression of the *S. gordonii amyB* gene. This gene encodes an α -amylase that is postulated to play a role in the breakdown of intracellular glycogen (**Chapter 5**), releasing sugar for fermentation and, as a result, increases lactate production for *Veillonella*. The precise signaling mechanism involved in this cross-species microbial communication is not completely understood, but the sugar maltose, which is possibly incorporated in the lipopolysaccharide present in the *V. atypica* outer membrane, may play an important role since it has been shown to induce *amyB* transcription (84, 158). It is of obvious interest to investigate whether this signaling interaction also occurs between the small intestinal *Streptococcus* and *Veillonella* communities. Preliminary analysis of the metatranscriptome data did not

reveal a specific association between the relative abundance of the *Veillonella* genus and the level of expression measured for the amylase gene. The quantitative resolution of the metatranscriptome datasets may not be sufficient to eventually detect such a correlation, while also a substantially larger sample-set may be needed to conclude on such a correlation with statistical significance.

Metabolic potential of small intestinal enterococci

Besides the *Streptococcus* and *Veillonella* isolates recovered from ileostoma effluent, a total of 66 *Enterococcus* isolates were also obtained (**Chapter 4**). Most *Enterococcus* isolates were obtained from *Mitis Salivarius* (MS-agar) plates (**Chapter 4**) and encompassed 7 distinct genetic lineages. An additional 11 *Enterococcus* isolates were cultivated from *Veillonella* selective agar (VSA) and belonged to a distinct group of *gallinarum* species enterococci (**Chapter 4**). Interestingly, evaluation of the carbohydrate fermentation capabilities of one randomly picked isolate from this *E. gallinarum* group (*Enterococcus* sp. HSIEG1; Table 7.2), showed that this isolate was able to ferment a wide variety of carbohydrates, including ribose, xylose, and tagatose, which were not fermented by any of the small intestinal *Streptococcus* isolates (**Chapter 4**). The broad fermentation capacity of the *E. gallinarum* strain was also reflected by the genetic repertoire represented in its genome sequence (Table 7.3), which encoded almost 2-fold more genes (9.8%) that were annotated to function in carbohydrate transport and metabolism relative to the genomes of the small-intestinal *Streptococcus* genomes (3.5-6.9%; Van den Bogert, et al. Unpublished results).

The *E. gallinarum* genome encoded single copies of the generic cytoplasmic factors enzyme I (EI), and phosphor-carrier protein (HPr; see also **chapter 5**) that are involved in phospho-transfer to over 30 PTS transporter functions with predicted specificities that include glucose/maltose, mannose, fructose, galactose, mannitol, lactose, sucrose, and β -glucosides (Van den Bogert, et al. Unpublished results). Moreover, the *E. gallinarum* genome encoded several ABC sugar transporters, including those involved in maltose/maltodextrin transport. Next to these transport associated functions, the genome also encodes the necessary enzymes pathways to metabolize most of these sugars, which is in good agreement with the growth capacities determined for this strain (Table 7.2), showing that the genome predictions are accurately matching with the observed physiological characteristics. The *Enterococcus gallinarum* strain encoded the required enzymes for glycolytic conversion energy generation, and also encoded a complete and intact pentose phosphate pathway. Pyruvate dissipation pathways predicted for *E. gallinarum* include the capacity to produce L-lactate (using lactate dehydrogenase [EC 1.1.1.27]) or several other fermentation metabolites, like formate, acetate, ethanol, acetoin, and 2,3-butanediol.

Table 7.2. API strip scores for isolate from *E. gallinarum* HSIEG1 recovered from ileostoma effluent using *Veillonella* selective agar*

Test	Reaction
CONTROL	0
Glycerol	1
Erythritol	2
D-Arabinose	3
L-Arabinose**	4
D-Ribose**	5
D-Xylose**	6
L-Xylose	7
D-Adonitol	8
Methyl-βD-Xylopyranoside	9
D-Galactose**	10
D-Glucose**	11
D-Fructose**	12
D-Mannose**	13
L-Sorbose	14
L-Rhamnose	15
Dulcitol	16
Inositol	17
D-Mannitol	18
D-Sorbitol	19
Methyl-αD-Mannopyranoside	20
Methyl-αD-glucopyranoside	21
N-AcetylGlucosamine	22
Amygdalin	23
Arbutin	24
Esculin ferric citrate	25
Salicin	26
D-Cellobiose**	27
D-Maltose**	28
D-Lactose (bovine origin)	29
D-Melibiose**	30
D-Saccharose (sucrose) **	31
D-Trehalose**	32
Inulin	33
D-MeLeZitose	34
D-Raffinose	35
AmiDon (starch)	36
Glycogen	37
Xylitol	38
Gentiobiose	39
D-Turanose	40
D-Lyxose	41
D-Tagatose**	42
D-Fucose	43
L-Fucose	44
D-Arabitol	45
L-arabitol	46
Potassium GlucoNaTe	47
Potassium 2-KetoGluconate	48
Potassium 5-KetoGluconate	49

Black: positive reaction; Grey w: weak reaction
d: variable reaction between duplicate tests of the same isolate; White: negative reaction

*: See **chapter 4** for API 50 CH test methodology

** : The genome of *E. gallinarum* HSIEG1 coded for necessary enzymes to metabolize this sugar

Despite the large metabolic capacity of the *Enterococcus* strain described here, small intestinal enterococci were generally present at low abundance in the small intestine microbiota (31) with relative abundances ranging from <0.1% to 1.5% (Leimena and Van den Bogert, et al. Unpublished results; **Chapter 3**). Possibly, the generally low abundance of enterococci in the oral cavity ((372); see above), esophagus, and stomach microbiota (23, 256), limits their influx in the proximal small intestine whereby they may be outcompeted by the streptococci that are continuously entering the small intestine ecosystem from more proximal regions of the digestive tract (353). Alternatively the common low abundance of enterococci in the small intestine may be due to incompatibility of these microbes with the physicochemical conditions of this habitat, which may relate to interactions between small-intestinal community members, efficiency in uptake and conversion of nutrient components, and bile-sensitivity or susceptibility to the pancreatic enzymes that are secreted in this habitat. This latter explanation may readily be tested by comparative analyses of the compatibility of the small intestinal *Streptococcus* and *Enterococcus* isolates (**Chapter 4**) with *in vitro* assays that aim to mimic specific small intestinal conditions (337), or in more advanced small intestine model systems like the TNO intestinal model (TIM) (222) or Simulator of Human Intestinal Microbial Ecosystems (SHIME) (338).

Table 7.3. Genome statistics for *V. parvula* HSIVP1 and *Enterococcus* sp. HSIEG1 isolates

	<i>V. parvula</i> HSIVP1*	<i>Enterococcus</i> sp. HSIEG1*
DDBJ/EMBL/GenBank Accession	ASKE00000000	ASKG01000000
Total number of Contigs	27	157
Contig Sum (bp)	2,144,824	3,257,043
Max contig size	759,850	143,405
Min contig size	1,153	1,017
Average contig size	79,437	20,745
contig N50	171,200	37,770
Total number of scaffolds	3	6
Scaffold Sum (bp)	2,177,813	3,447,543
Max scaffold size	1,454,078	1,848,264
Min scaffold size	31,789	41,260
Average scaffold size	725,937	574,590
Scaffold N50	1,454,078	1,848,264
GC content (%)	38.51	40.45
Number of predicted proteins	2,014	3,901
Genes assigned to COG	1,633	3,087
Genes assigned to KEGG	1,087	1,566

*: The version described in this chapter is version XXXX01000000

Influence of small intestinal bacteria on the human host

Previous studies have shown that individual GI commensals (e.g. *Faecalibacterium prauznitsii* (301) and *Bacteroides fragilis* (312, 327)) affect the host immune system in specific ways and can thereby play an important role in homeostasis (**Chapter 1**; see (153) for a recent review). To obtain an impression of the capacity of the small-intestinal streptococci and *Veillonella* to affect the host immune system, their immunomodulatory properties were evaluated using *in vitro* models that employ human blood-derived immature dendritic cells (**Chapter 6**). These assays revealed that the different streptococci induced distinct cytokine profiles with variable induction capacities of the pro-inflammatory cytokines IL-8, TNF- α , and IL-12p70, which may be related to their varying capacity to activate NF- κ B responses via TLR2/6 (**Chapter 6**). Notably, the *V. parvula* isolate was also tested in these assays and was shown to induce only low amounts of IL-12p70 in iDCs, but strongly induced the production of IL-6. Considering the co-occurrence (**Chapter 3** and **4**) and the proposed metabolic relationship of *Streptococcus* and *Veillonella* (see above), we hypothesized that the host immune system may react differently to a combination of strains from both genera. Indeed, stimulation of iDCs with pairs of *Streptococcus* and *Veillonella* isolates appeared to negate IL-12p70 production by dendritic cells, while augmented IL-8, IL-6, IL-10, and TNF- α responses. These observations support the idea that the composition of the intestinal microbiota may modulate an individual's immune status and homeostasis, which may affect their responsiveness during infection, or other conditions that involve immunological responses like cancer, (food-)allergies, or autoimmunity.

Future perspectives

The work in this thesis deepens our understanding of the human small intestinal microbiota in terms of composition, dynamics, and activity as well as provides an initial analysis of the physiological, genetic, and immunomodulatory characteristics of streptococci and *Veillonella* as typical members of the small intestinal ecosystem. These studies are valuable additions to the accumulated data that generally focuses on the homeostatic interaction between the gut microbiota and the host in the terminal part of the GI tract.

As we speculated that diet composition is a major driver of the low degree of temporal stability of the small intestine microbiota, one of the main questions to be addressed is if dietary components can be used to effectively modulate the microbiota and thereby the host immune status (**Chapter 3**). *In vitro* model systems are instrumental in the first steps unraveling the underlying mechanisms (see (193) for a review), but they do not accurately mimic the *in vivo* situation that preserves the contributions of other microbial members in the consortium, cross-talk between immune cells (DCs) within densely populated immune follicles like the Peyer's Patches (55, 271), and simultaneous exposure of immune cells to MAMPs derived from multiple rather than single microbial cells or species (168). Therefore, *in vivo* models are crucial to decipher the mechanisms underlying the immunomodulatory responses elicited by individual microbes, as well as by combinations of them, in order to understand the molecular communication between the small intestine microbiota and the mucosal immune system. Such studies are seldom performed with human subjects (335, 336) due to the experimental and ethical limitations. These issues may be overcome by using gnotobiotic animals that provide an attractive model to extrapolate *in vitro* findings to an *in vivo* situation (101, 137, 286, 303, 306, 369). Conventionalization studies by El Aidy, et al. (85, 86) in germfree mouse models demonstrated that the jejunal mucosa quickly and strongly responds to the colonizing microbiota with a repression of processes involved in lipid metabolism and gluconeogenesis and induction of anabolic metabolism, including lipogenesis, nucleotide genesis, amino acid synthesis and glycolysis. Interestingly, transcriptional signature genes of the mucosal response to the colonizing microbiota included genes of which human orthologs play a role in metabolic disorders (e.g. type 2 diabetes and insuline resistance), suggesting that the proximal small-intestinal host-microbe interactions are critical modulators of our overall well-being. The findings reported by El Aidy, et al. may relate to the improvement in insulin sensitivity of patients with metabolic syndrome after infusion with the intestinal microbiota from lean individuals (352) and explain the mechanism underlying the observation that diabetic obese patients show a dramatic improvement in glycemia after implantation of a duodenal-jejunal bypass liner (EndoBarrier) that hinders nutrient jejunal nutrient absorption and exposure of the local microbiota and its metabolites (63).

The genomic and physiological characteristics of the small-intestinal isolates (**Chapter 4 and 5**) can be of significant value to predict the effect of different dietary substances on *Streptococcus* population dynamics in the human small intestine. Only a minority of the small-intestinal streptococcal isolates had the capacity to utilize

arabinose, which is richly present in for example oatmeal. Therefore, we hypothesize that consumption of oatmeal causes a distinct shift in composition and activity of the (streptococcal) populations in the small intestine compared to subjects consuming foods low in arabinose like yoghurt. To address this, ileostomy subjects are an adequate *in vivo* model system for a dietary intervention study, because of the possibilities for repetitive and non-invasive sampling of luminal effluent that can be executed by the ileostoma volunteers themselves. The intervention can be focused on breakfast in which study participants are asked to consume a yoghurt breakfast for one week, followed by an oatmeal breakfast for one week. After these weeks the collected effluent samples are analyzed to assess microbial diversity and population dynamics (by 16S rRNA gene sequence based community profiling), while activity patterns of the microbiota can be determined by metatranscriptome analysis of mRNA enriched samples from the effluent. The study-design would benefit from a large group of (at least 10) subjects that also includes individuals with an intact intestinal tract, to further establish a consensus on the 'normal' composition and activity of the microbiota in the human small intestine. Catheter and capsule (e.g. IntelliCap; see above) sampling methodologies could be instrumental in such endeavour, albeit that they come with the challenge to collect and avoid deterioration of a sufficient volume of luminal content that accurately represents the small-intestinal composition and transcriptome. Improvements in sample preparation protocols and sequencing technologies potentially could eventually overcome these issues of technical limitations and may enable sequencing of minimal amounts of nucleic acids (171). Alongside these advancements, sequencing technologies will generate more sequences with longer read-lengths at lower costs, allowing the reconstruction of microbial ecosystems at greater depth of analysis than previously possible (288).

To further our knowledge of the interplay between diet, microbiota and the mucosal tissues in the small intestine (85), it would be particularly interesting to determine the change in transcriptional responses in the small intestine mucosa during the dietary intervention. To this end, mucosal expression profiles can be generated by whole-genome microarray based analysis of biopsies taken by gastrointestinal endoscopy similar to the approach used by van Baarlen, et al. (335, 336) to study *in vivo* mucosal responses in the duodenum of healthy human volunteers upon the consumption of dietary lactobacilli. Interestingly, these studies revealed that the bacteria induced mucosal responses that resembled that induced by pharmaceutical compounds used to, for example, regulate immune responses, treat high blood pressure, or stimulate tissue vascularity (335). DNA extracted from the biopsies can be used to analyse the diversity of the mucosa-associated microbiota that can significantly differ from that of the intestinal lumen (80, 201, 383).

Furthermore, food components and the intestinal microbiota influence intestinal morphology in terms of crypt depth, villus height, and vascularization (see (302) for a recent review), as well as tight junction structures that influence intestinal permeability (167) (see (332) for a review). The latter is postulated to play a role in development of intestinal disorders (e.g. inflammatory bowel diseases; (51)).

Histological and histochemical analyses of the collected biopsies will further our understanding of the morphological and molecular changes of the small intestinal mucosa as a result of exposure to dynamics in microbiota and food composition.

The next challenge will be to identify mechanistic connections between multi-disciplinary data sets from human intervention studies and to draw conclusions about the homeostatic relationship between diet, microbiota, and immune responses. The field of bioinformatics will be especially important to provide algorithms that reduce computational complexity for efficient data mining. Ultimately the research in this area may open avenues towards rational design of food compositions that aim to modulate the small intestine microbiota composition and activity and thereby aiming to regulate mucosal (and systemic) immunity.



Appendices

References	186
List of publications	204
Co-author affiliations	205
Samenvatting	206
Acknowledgements – Dankwoord	209
About the author	211
Overview of completed training activities	212

References

1. **Aas, J. A., B. J. Paster, L. N. Stokes, I. Olsen, and F. E. Dewhirst.** 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**:5721-5732.
2. **Abubucker, S., N. Segata, J. Goll, A. M. Schubert, J. Izard, B. L. Cantarel, B. Rodriguez-Mueller, J. Zucker, M. Thiagarajan, B. Henrissat, O. White, S. T. Kelley, B. Methe, P. D. Schloss, D. Gevers, et al.** 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8**:e1002358.
3. **Ahmed, S., G. T. Macfarlane, A. Fite, A. J. McBain, P. Gilbert, and S. Macfarlane.** 2007. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* **73**:7435-7442.
4. **Akira, S., K. Takeda, and T. Kaisho.** 2001. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol* **2**:675-680.
5. **Albesharat, R., M. A. Ehrmann, M. Korakli, S. Yazaji, and R. F. Vogel.** 2011. Phenotypic and genotypic analyses of lactic acid bacteria in local fermented food, breast milk and faeces of mothers and their babies. *Syst Appl Microbiol* **34**:148-155.
6. **Alegado, R. A., C. Y. Chin, D. M. Monack, and M. W. Tan.** 2011. The two-component sensor kinase KdpD is required for *Salmonella typhimurium* colonization of *Caenorhabditis elegans* and survival in macrophages. *Cell Microbiol* **13**:1618-1637.
7. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-410.
8. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological reviews* **59**:143-169.
9. **Andersson, A. F., M. Lindberg, H. Jakobsson, F. Backhed, P. Nyren, and L. Engstrand.** 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS one* **3**:e2836.
10. **Arias, C. A., and B. E. Murray.** 2012. The rise of the *Enterococcus*: beyond vancomycin resistance. *Nat Rev Microbiol* **10**:266-278.
11. **Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J. M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, et al.** 2011. Enterotypes of the human gut microbiome. *Nature* **473**:174-180.
12. **Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman.** 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**:7724-7736.
13. **Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman.** 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72**:5734-5741.
14. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, et al.** 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**:75.
15. **Backhed, F., R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon.** 2005. Host-bacterial mutualism in the human intestine. *Science* **307**:1915-1920.
16. **Baker, G. C., J. J. Smith, and D. A. Cowan.** 2003. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**:541-555.
17. **Baldrian, P., M. Kolarik, M. Stursova, J. Kopecky, V. Valaskova, T. Vetrovsky, L. Zifcakova, J. Snajdr, J. Ridl, C. Vlcek, and J. Voriskova.** 2011. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition.
18. **Ben-Amor, K.** 2004. Microbial Eco-physiology of the human intestinal tract: A flow cytometric approach. Wageningen University, Wageningen.
19. **Ben-Amor, K., H. Heilig, H. Smidt, E. E. Vaughan, T. Abee, and W. M. de Vos.** 2005. Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. *Appl Environ Microbiol* **71**:4679-4689.
20. **Benyacoub, J., P. F. Perez, F. Rochat, K. Y. Saudan, G. Reuteler, N. Antille, M. Humen, G. L. De Antoni, C. Cavadini, S. Blum, and E. J. Schiffrin.** 2005. *Enterococcus faecium* SF68 enhances the immune response to *Giardia intestinalis* in mice. *J Nutr* **135**:1171-1176.
21. **Berridge, B. R., J. D. Fuller, J. de Azavedo, D. E. Low, H. Bercovier, and P. F. Frelief.** 1998. Development of specific nested oligonucleotide PCR primers for the *Streptococcus iniae* 16S-23S ribosomal DNA intergenic spacer. *J Clin Microbiol* **36**:2778-2781.

22. **Bevins, C. L., and N. H. Salzman.** 2011. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol* **9**:356-368.
23. **Bik, E. M., P. B. Eckburg, S. R. Gill, K. E. Nelson, E. A. Purdom, F. Francois, G. Perez-Perez, M. J. Blaser, and D. A. Relman.** 2006. Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci U S A* **103**:732-737.
24. **Binnewies, T. T., Y. Motro, P. F. Hallin, O. Lund, D. Dunn, T. La, D. J. Hampson, M. Bellgard, T. M. Wassenaar, and D. W. Ussery.** 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics* **6**:165-185.
25. **Boekhorst, J., R. J. Siezen, M. C. Zwahlen, D. Vilanova, R. D. Pridmore, A. Mercenier, M. Kleerebezem, W. M. de Vos, H. Brussow, and F. Desiere.** 2004. The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content. *Microbiology* **150**:3601-3611.
26. **Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano.** 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**:578-579.
27. **Boleij, A., C. M. Muytjens, S. I. Bukhari, N. Cayet, P. Glaser, P. W. Hermans, D. W. Swinkels, A. Bolhuis, and H. Tjalsma.** 2011. Novel clues on the specific association of *Streptococcus gallolyticus* subsp *gallolyticus* with colorectal cancer. *J Infect Dis* **203**:1101-1109.
28. **Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, et al.** 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* **22**:1554-1558.
29. **Booijink, C. C.** 2009. Analysis of Diversity and Function of the Human Small Intestinal Microbiota. Wageningen University, Wageningen, The Netherlands.
30. **Booijink, C. C., J. Boekhorst, E. G. Zoetendal, H. Smidt, M. Kleerebezem, and W. M. de Vos.** 2010. Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. *Appl Environ Microbiol* **76**:5533-5540.
31. **Booijink, C. C., S. El-Aidy, M. Rajilić-Stojanović, H. G. Heilig, F. J. Troost, H. Smidt, M. Kleerebezem, W. M. De Vos, and E. G. Zoetendal.** 2010. High temporal and inter-individual variation detected in the human ileal microbiota. *Environ Microbiol* **12**:3213-3227.
32. **Booijink, C. C., E. G. Zoetendal, M. Kleerebezem, and W. M. de Vos.** 2007. Microbial communities in the human small intestine: coupling diversity to metagenomics. *Future Microbiol* **2**:285-295.
33. **Bott, M., M. Meyer, and P. Dimroth.** 1995. Regulation of anaerobic citrate metabolism in *Klebsiella pneumoniae*. *Mol. Microbiol.* **18**:533-546.
34. **Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer.** 2003. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* **185**:6220-6223.
35. **Brown, C. T., A. G. Davis-Richardson, A. Giongo, K. A. Gano, D. B. Crabb, N. Mukherjee, G. Casella, J. C. Drew, J. Ilonen, M. Knip, H. Hyoty, R. Veijola, T. Simell, O. Simell, J. Neu, et al.** 2011. Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PloS one* **6**:e25792.
36. **Busuioac, M., K. Mackiewicz, B. A. Buttaro, and P. J. Piggot.** 2009. Role of intracellular polysaccharide in persistence of *Streptococcus mutans*. *J Bacteriol* **191**:7315-7322.
37. **Cai, L., L. Ye, A. H. Tong, S. Lok, and T. Zhang.** 2013. Biased Diversity Metrics Revealed by Bacterial 16S Pyrotags Derived from Different Primer Sets. *PloS one* **8**:e53649.
38. **Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, et al.** 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**:335-336.
39. **Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight.** 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* **108 Suppl 1**:4516-4522.
40. **Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight.** 2010. Microbes and Health Sackler Colloquium: Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*.
41. **Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* **21**:3422-3423.

42. **Case, R. J., Y. Boucher, I. Dahllof, C. Holmstrom, W. F. Doolittle, and S. Kjelleberg.** 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* **73**:278-288.
43. **Chavez de Paz, L., G. Svensater, G. Dahlen, and G. Bergenholtz.** 2005. Streptococci from root canals in teeth with apical periodontitis receiving endodontic treatment. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **100**:232-241.
44. **Chen, K., and L. Pachter.** 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput Biol* **1**:106-112.
45. **Cheng, Q., D. Staflieni, S. S. Purushothaman, and P. Cleary.** 2002. The group B streptococcal C5a peptidase is both a specific protease and an invasin. *Infect Immun* **70**:2408-2413.
46. **Cheon, K., S. A. Moser, J. Whiddon, R. C. Osgood, S. Momeni, J. D. Ruby, G. R. Cutter, D. B. Allison, and N. K. Childers.** 2011. Genetic diversity of plaque *mutans* streptococci with rep-PCR. *J Dent Res* **90**:331-335.
47. **Chitsaz, H., J. L. Yee-Greenbaum, G. Tesler, M. J. Lombardo, C. L. Dupont, J. H. Badger, M. Novotny, D. B. Rusch, L. J. Fraser, N. A. Gormley, O. Schulz-Trieglaff, G. P. Smith, D. J. Evers, P. A. Pevzner, and R. S. Lasken.** 2011. Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotechnol* **29**:915-921.
48. **Claesson, M. J., S. Cusack, O. O'Sullivan, R. Greene-Diniz, H. de Weerd, E. Flannery, J. R. Marchesi, D. Falush, T. Dinan, G. Fitzgerald, C. Stanton, D. van Sinderen, M. O'Connor, N. Harnedy, K. O'Connor, et al.** 2011. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A* **108 Suppl 1**:4586-4591.
49. **Claesson, M. J., O. O'Sullivan, Q. Wang, J. Nikkila, J. R. Marchesi, H. Smidt, W. M. de Vos, R. P. Ross, and P. W. O'Toole.** 2009. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PloS one* **4**:e6669.
50. **Claesson, M. J., Q. Wang, O. O'Sullivan, R. Greene-Diniz, J. R. Cole, R. P. Ross, and P. W. O'Toole.** 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**:e200.
51. **Clayburgh, D. R., L. Shen, and J. R. Turner.** 2004. A porous defense: the leaky epithelial barrier in intestinal disease. *Laboratory investigation; a journal of technical methods and pathology* **84**:282-291.
52. **Cloonan, N., and S. M. Grimmond.** 2008. Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol* **9**:234.
53. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**:D141-145.
54. **Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, and J. A. Farrow.** 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* **44**:812-826.
55. **Coombes, J. L., and F. Powrie.** 2008. Dendritic cells in intestinal immune regulation. *Nat Rev Immunol* **8**:435-446.
56. **Corinti, S., C. Albanesi, A. la Sala, S. Pastore, and G. Girolomoni.** 2001. Regulatory activity of autocrine IL-10 on dendritic cell functions. *J Immunol* **166**:4312-4318.
57. **Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight.** 2009. Bacterial community variation in human body habitats across space and time. *Science* **326**:1694-1697.
58. **Cotter, P. D.** 2011. Small intestine and microbiota. *Curr Opin Gastroenterol* **27**:99-105.
59. **Cotter, P. D., C. Hill, and R. P. Ross.** 2005. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol* **3**:777-788.
60. **Daims, H., A. Bruhl, R. Amann, K. H. Schleifer, and M. Wagner.** 1999. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**:434-444.
61. **Darby, A. C., and N. Hall.** 2008. Fast forward genetics. *Nat Biotechnol* **26**:1248-1249.
62. **de Jong, A., A. J. van Heel, J. Kok, and O. P. Kuipers.** 2010. BAGEL2: mining for bacteriocins in genomic data. *Nucleic Acids Res* **38**:W647-651.

63. **de Moura, E. G., B. C. Martins, G. S. Lopes, I. R. Orso, S. L. de Oliveira, M. P. Galvao Neto, M. A. Santo, P. Sakai, A. C. Ramos, A. B. Garrido Junior, M. C. Mancini, A. Halpern, and I. Ceconello.** 2012. Metabolic improvements in obese type 2 diabetes subjects implanted for 1 year with an endoscopically deployed duodenal-jejunal bypass liner. *Diabetes Technol Ther* **14**:183-189.
64. **de Saizieu, A., C. Gardes, N. Flint, C. Wagner, M. Kamber, T. J. Mitchell, W. Keck, K. E. Amrein, and R. Lange.** 2000. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J Bacteriol* **182**:4696-4703.
65. **De Vos, P., G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. Schleifer, and W. B. Whitman.** 2009. *Bergey's Manual of Systematic Bacteriology*, 2nd ed, vol. 3: The Firmicutes. Springer, New York.
66. **de Vos, W. M., and E. A. de Vos.** 2012. Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr Rev* **70 Suppl 1**:S45-56.
67. **Derrien, M., E. E. Vaughan, C. M. Plugge, and W. M. de Vos.** 2004. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **54**:1469-1476.
68. **DeSantis, T. Z., E. L. Brodie, J. P. Moberg, I. X. Zubietta, Y. M. Piceno, and G. L. Andersen.** 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* **53**:371-383.
69. **Dethlefsen, L., S. Huse, M. L. Sogin, and D. A. Relman.** 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**:e280.
70. **Detmers, F. J., E. R. Kunji, F. C. Lanfermeijer, B. Poolman, and W. N. Konings.** 1998. Kinetics and specificity of peptide uptake by the oligopeptide transport system of *Lactococcus lactis*. *Biochemistry* **37**:16671-16679.
71. **Diaz, P. I., A. K. Dupuy, L. Abusleme, B. Reese, C. Obergfell, L. Choquette, A. Dongari-Bagtzoglou, D. E. Peterson, E. Terzi, and L. D. Strausbaugh.** 2012. Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Molecular oral microbiology* **27**:182-201.
72. **Dimroth, P., and B. Schink.** 1998. Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria. *Arch. Microbiol.* **170**:69-77.
73. **Do, T., K. A. Jolley, M. C. Maiden, S. C. Gilbert, D. Clark, W. G. Wade, and D. Beighton.** 2009. Population structure of *Streptococcus oralis*. *Microbiology* **155**:2593-2602.
74. **Dridi, B., M. Henry, A. El Khechine, D. Raoult, and M. Drancourt.** 2009. High prevalence of *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* detected in the human gut using an improved DNA detection protocol. *PLoS one* **4**:e7063.
75. **Duerkop, B. A., S. Vaishnava, and L. V. Hooper.** 2009. Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* **31**:368-376.
76. **Dunbar, J., S. White, and L. Forney.** 1997. Genetic Diversity through the Looking Glass: Effect of Enrichment Bias. *Appl Environ Microbiol* **63**:1326-1331.
77. **Duncan, S. H., G. Holtrop, G. E. Lobley, A. G. Calder, C. S. Stewart, and H. J. Flint.** 2004. Contribution of acetate to butyrate formation by human faecal bacteria. *Br J Nutr* **91**:915-923.
78. **Duncan, S. H., P. Louis, and H. J. Flint.** 2004. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* **70**:5810-5817.
79. **Dunne, E. M., E. K. Ong, R. J. Moser, P. M. Siba, S. Phuanukoonnon, A. R. Greenhill, R. M. Robins-Browne, E. K. Mulholland, and C. Satzke.** 2011. Multilocus sequence typing of *Streptococcus pneumoniae* by use of mass spectrometry. *J Clin Microbiol* **49**:3756-3760.
80. **Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman.** 2005. Diversity of the human intestinal microbial flora. *Science* **308**:1635-1638.
81. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792-1797.
82. **Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight.** 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**:2194-2200.
83. **Egert, M., A. A. de Graaf, A. Maathuis, P. de Waard, C. M. Plugge, H. Smidt, N. E. Deutz, C. Dijkema, W. M. de Vos, and K. Venema.** 2007. Identification of glucose-fermenting bacteria present in an *in vitro* model of the human intestine by RNA-stable isotope probing. *FEMS Microbiol Ecol* **60**:126-135.

84. **Egland, P. G., R. J. Palmer, Jr., and P. E. Kolenbrander.** 2004. Interspecies communication in *Streptococcus gordonii*-*Veillonella atypica* biofilms: signaling in flow conditions requires juxtaposition. *Proc Natl Acad Sci U S A* **101**:16917-16922.
85. **El Aidy, S., C. A. Merrifield, M. Derrien, P. van Baarlen, G. Hooiveld, F. Levenez, J. Dore, J. Dekker, E. Holmes, S. P. Claus, D. J. Reijngoud, and M. Kleerebezem.** 2012. The gut microbiota elicits a profound metabolic reorientation in the mouse jejunal mucosa during conventionalisation. *Gut*, In press.
86. **El Aidy, S., P. van Baarlen, M. Derrien, D. J. Lindenbergh-Kortleve, G. Hooiveld, F. Levenez, J. Dore, J. Dekker, J. N. Samsom, E. E. Nieuwenhuis, and M. Kleerebezem.** 2012. Temporal and spatial interplay of microbiota and intestinal mucosa drive establishment of immune homeostasis in conventionalized mice. *Mucosal Immunol* **5**:567-579.
87. **Emrich, S. J., W. B. Barbazuk, L. Li, and P. S. Schnable.** 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res* **17**:69-73.
88. **Engelbrekton, A., V. Kunin, K. C. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, and P. Hugenholtz.** 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *Isme J* **4**:642-647.
89. **Engelke, G., Z. Gutowski-Eckel, P. Kiesau, K. Siegers, M. Hammelmann, and K. D. Entian.** 1994. Regulation of nisin biosynthesis and immunity in *Lactococcus lactis* 6F3. *Appl Environ Microbiol* **60**:814-825.
90. **Enright, M. C., B. G. Spratt, A. Kalia, J. H. Cross, and D. E. Bessen.** 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect Immun* **69**:2416-2427.
91. **Ferraris, L., J. Aires, A. J. Waligora-Dupriet, and M. J. Butel.** 2010. New selective medium for selection of bifidobacteria from human feces. *Anaerobe* **16**:469-471.
92. **Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, G. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najjar, et al.** 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98**:4658-4663.
93. **Fey, A., S. Eichler, S. Flavier, R. Christen, M. G. Hofle, and C. A. Guzman.** 2004. Establishment of a real-time PCR-based approach for accurate quantification of bacterial RNA targets in water, using *Salmonella* as a model organism. *Appl Environ Microbiol* **70**:3618-3623.
94. **Finegold, S. M., V. L. Sutter, J. D. Boyle, and K. Shimada.** 1970. The normal flora of ileostomy and transverse colostomy effluents. *J Infect Dis* **122**:376-381.
95. **Flamm, H.** 2011. [The Austrian pediatrician Theodor Escherich as bacteriologist and social hygienist: The 100th anniversary of his death on February 15th, 1911]. *Wien Klin Wochenschr* **123**:157-171.
96. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, and et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496-512.
97. **Fontaine, L., C. Boutry, M. H. de Frahan, B. Delplace, C. Fremaux, P. Horvath, P. Boyaval, and P. Hols.** 2010. A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Bacteriol* **192**:1444-1454.
98. **Frank, J. A., C. I. Reich, S. Sharma, J. S. Weisbaum, B. A. Wilson, and G. J. Olsen.** 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* **74**:2461-2470.
99. **Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. Delong.** 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**:3805-3810.
100. **Furet, J. P., P. Quenee, and P. Tailliez.** 2004. Molecular quantification of lactic acid bacteria in fermented milk products using real-time quantitative PCR. *Int J Food Microbiol* **97**:197-207.
101. **Gaboriau-Routhiau, V., S. Rakotobe, E. Lecuyer, I. Mulder, A. Lan, C. Bridonneau, V. Rochet, A. Pisi, M. De Paepe, G. Brandi, G. Eberl, J. Snel, D. Kelly, and N. Cerf-Bensussan.** 2009. The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**:677-689.
102. **Garrity, G. M., J. A. Bell, and T. G. Lilburn.** 2004. Taxonomic outline of the prokaryotes, p. 401, *Bergey's manual of systematic bacteriology*, 2nd ed. Release 5.0, Springer-Verlag, New York, NY.

103. **Giammarinaro, P., M. Sicard, and A. M. Gasc.** 1999. Genetic and physiological studies of the CiaH-CiaR two-component signal-transducing system involved in cefotaxime resistance and competence of *Streptococcus pneumoniae*. *Microbiology* **145** (Pt 8):1859-1869.
104. **Giannoukos, G., D. M. Ciulla, K. Huang, B. J. Haas, J. Izard, J. Z. Levin, J. Livny, A. M. Earl, D. Gevers, D. V. Ward, C. Nusbaum, B. W. Birren, and A. Gnrke.** 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* **13**:r23.
105. **Gilbert, J. A., D. Field, Y. Huang, R. Edwards, W. Li, P. Gilna, and I. Joint.** 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS one* **3**:e3042.
106. **Gilbert, J. A., and M. Hughes.** 2011. Gene expression profiling: metatranscriptomics. *Methods Mol Biol* **733**:195-205.
107. **Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, et al.** 2001. Comparative genomics of *Listeria* species. *Science* **294**:849-852.
108. **Glenn, T. C.** 2011. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* **11**:759-769.
109. **Gonzalez, J. M., M. C. Portillo, P. Belda-Ferre, and A. Mira.** 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PLoS one* **7**:e29973.
110. **Goodman, A. L., G. Kallstrom, J. J. Faith, A. Reyes, A. Moore, G. Dantas, and J. I. Gordon.** 2011. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A* **108**:6252-6257.
111. **Gorbach, S. L., L. Nahas, L. Weinstein, R. Levitan, and J. F. Patterson.** 1967. Studies of intestinal microflora. IV. The microflora of ileostomy effluent: a unique microbial ecology. *Gastroenterology* **53**:874-880.
112. **Gosalbes, M. J., A. Durban, M. Pignatelli, J. J. Abellan, N. Jimenez-Hernandez, A. E. Perez-Cobas, A. Latorre, and A. Moya.** 2011. Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS one* **6**:e17447.
113. **Gourse, R. L., T. Gaal, M. S. Bartlett, J. A. Appleman, and W. Ross.** 1996. rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. *Annu Rev Microbiol* **50**:645-677.
114. **Greub, G.** 2012. Culturomics: a new approach to study the human microbiome. *Clin Microbiol Infect* **18**:1157-1159.
115. **Guarner, F.** 2006. Enteric flora in health and disease. *Digestion* **73 Suppl 1**:5-12.
116. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**:696-704.
117. **Guss, A. M., G. Roeselers, I. L. Newton, C. R. Young, V. Klepac-Ceraj, S. Lory, and C. M. Cavanaugh.** 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *Isme J* **5**:20-29.
118. **Haas, B. J., D. Gevers, A. M. Earl, M. Feldgarden, D. V. Ward, G. Giannoukos, D. Ciulla, D. Tabbaa, S. K. Highlander, E. Sodergren, B. Methe, T. Z. DeSantis, C. Human Microbiome, J. F. Petrosino, R. Knight, et al.** 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**:494-504.
119. **Haas, W., J. Sublett, D. Kaushal, and E. I. Tuomanen.** 2004. Revising the role of the pneumococcal vex-vncRS locus in vancomycin tolerance. *J. Bacteriol.* **186**:8463-8471.
120. **Hamady, M., and R. Knight.** 2009. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res* **19**:1141-1152.
121. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**:235-237.
122. **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**:669-685.
123. **Handelsman, J., M. Liles, D. Mann, C. Riesenfeld, R. M. Goodman, and D. Brendan Wren and Nick.** 2002. Cloning the metagenome: Culture-independent access to the diversity and functions of the uncultivated microbial world, p. 241-255, *Methods in Microbiology*, vol. 33. Academic Press.
124. **Harmsen, H. J., G. C. Raangs, T. He, J. E. Degener, and G. W. Welling.** 2002. Extensive set of 16S rRNA-based probes for detection of bacteria in human feces. *Appl Environ Microbiol* **68**:2982-2990.

125. **Hartman, A. L., D. M. Lough, D. K. Barupal, O. Fiehn, T. Fishbein, M. Zasloff, and J. A. Eisen.** 2009. Human gut microbiome adopts an alternative state following small bowel transplantation. *Proc Natl Acad Sci U S A* **106**:17187-17192.
126. **Havarstein, L. S.** 2010. Increasing competence in the genus *Streptococcus*. *Mol Microbiol* **78**:541-544.
127. **Hayashi, H., M. Sakamoto, and Y. Benno.** 2004. Evaluation of three different forward primers by terminal restriction fragment length polymorphism analysis for determination of fecal *bifidobacterium* spp. in healthy subjects. *Microbiol Immunol* **48**:1-6.
128. **Hayashi, H., R. Takahashi, T. Nishi, M. Sakamoto, and Y. Benno.** 2005. Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol* **54**:1093-1101.
129. **He, Z., Y. Deng, and J. Zhou.** 2012. Development of functional gene microarrays for microbial community analysis. *Curr Opin Biotechnol* **23**:49-55.
130. **Henne, A., R. A. Schmitz, M. Bomeke, G. Gottschalk, and R. Daniel.** 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66**:3113-3116.
131. **Hilpert, W., and P. Dimroth.** 1991. On the mechanism of sodium ion translocation by methylmalonyl-CoA decarboxylase from *Veillonella alcalescens*. *Eur J Biochem* **195**:79-86.
132. **Hilpert, W., and P. Dimroth.** 1983. Purification and characterization of a new sodium-transport decarboxylase. Methylmalonyl-CoA decarboxylase from *Veillonella alcalescens*. *Eur J Biochem* **132**:579-587.
133. **Hoffmann, A., W. Hilpert, and P. Dimroth.** 1989. The carboxyltransferase activity of the sodium-ion-translocating methylmalonyl-CoA decarboxylase of *Veillonella alcalescens*. *Eur J Biochem* **179**:645-650.
134. **Homan, W. L., D. Tribe, S. Poznanski, M. Li, G. Hogg, E. Spalburg, J. D. Van Embden, and R. J. Willems.** 2002. Multilocus sequence typing scheme for *Enterococcus faecium*. *J Clin Microbiol* **40**:1963-1971.
135. **Hooper, L. V.** 2009. Do symbiotic bacteria subvert host immunity? *Nat Rev Microbiol* **7**:367-374.
136. **Hooper, L. V., T. Midtvedt, and J. I. Gordon.** 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr* **22**:283-307.
137. **Hooper, L. V., M. H. Wong, A. Thelin, L. Hansson, P. G. Falk, and J. I. Gordon.** 2001. Molecular analysis of commensal host-microbial relationships in the intestine. *Science* **291**:881-884.
138. **Horz, H. P., M. E. Vianna, B. P. Gomes, and G. Conrads.** 2005. Evaluation of universal probes and primer sets for assessing total bacterial load in clinical samples: general implications and practical use in endodontic antimicrobial therapy. *J Clin Microbiol* **43**:5332-5337.
139. **Huber, T., G. Faulkner, and P. Hugenholtz.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**:2317-2319.
140. **Huder, J. B., and P. Dimroth.** 1995. Expression of the sodium ion pump methylmalonyl-coenzyme A-decarboxylase from *Veillonella parvula* and of mutated enzyme specimens in *Escherichia coli*. *J. Bacteriol.* **177**:3623-3630.
141. **Huder, J. B., and P. Dimroth.** 1993. Sequence of the sodium ion pump methylmalonyl-CoA decarboxylase from *Veillonella parvula*. *J Biol Chem* **268**:24564-24571.
142. **Huebner, A., S. Sharma, M. Srisa-Art, F. Hollfelder, J. B. Edel, and A. J. Demello.** 2008. Microdroplets: a sea of applications? *Lab Chip* **8**:1244-1254.
143. **Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67**:4399-4406.
144. **Huijsdens, X. W., R. K. Linskens, M. Mak, S. G. Meuwissen, C. M. Vandenbroucke-Grauls, and P. H. Savelkoul.** 2002. Quantification of bacteria adherent to gastrointestinal mucosa by real-time PCR. *J Clin Microbiol* **40**:4423-4427.
145. **Human Microbiome Jumpstart Reference Strains, C., K. E. Nelson, G. M. Weinstock, S. K. Highlander, K. C. Worley, H. H. Creasy, J. R. Wortman, D. B. Rusch, M. Mitreva, E. Sodergren, A. T. Chinwalla, M. Feldgarden, D. Gevers, B. J. Haas, R. Madupu, et al.** 2010. A catalog of reference genomes from the human microbiome. *Science* **328**:994-999.
146. **Human Microbiome Project, C.** 2012. A framework for human microbiome research. *Nature* **486**:215-221.
147. **Human Microbiome Project, C.** 2012. Structure, function and diversity of the healthy human microbiome. *Nature* **486**:207-214.

148. **Hungate, R. E.** 1963. A roll tube method for cultivation of strict anaerobes, vol. 3B. Academic press, London.
149. **Huse, S. M., J. A. Huber, H. G. Morrison, M. L. Sogin, and D. M. Welch.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**:R143.
150. **Hutkins, R. W., and C. Ponne.** 1991. Lactose Uptake Driven by Galactose Efflux in *Streptococcus thermophilus*: Evidence for a Galactose-Lactose Antiporter. *Appl Environ Microbiol* **57**:941-944.
151. **Huys, G., T. Vanhoutte, and P. Vandamme.** 2008. Application of sequence-dependent electrophoresis fingerprinting in exploring biodiversity and population dynamics of human intestinal microbiota: what can be revealed? *Interdiscip Perspect Infect Dis* **2008**:597603.
152. **Ingham, C. J., A. Sprenkels, J. Bomer, D. Molenaar, A. van den Berg, J. E. van Hylckama Vlieg, and W. M. de Vos.** 2007. The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proc Natl Acad Sci U S A* **104**:18217-18222.
153. **Ivanov, II, and K. Honda.** 2012. Intestinal commensal microbes as immune modulators. *Cell host & microbe* **12**:496-508.
154. **Ivanov, II, and D. R. Littman.** 2010. Segmented filamentous bacteria take the stage. *Mucosal Immunol* **3**:209-212.
155. **Jacobs, D. M., E. Gaudier, J. van Duynhoven, and E. E. Vaughan.** 2009. Non-digestible food ingredients, colonic microbiota and the impact on gut health and immunity: a role for metabolomics. *Curr Drug Metab* **10**:41-54.
156. **Jalanka-Tuovinen, J., A. Salonen, J. Nikkila, O. Immonen, R. Kekkonen, L. Lahti, A. Palva, and W. M. de Vos.** 2011. Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PloS one* **6**:e23035.
157. **Jarvie, T., and T. Harkins.** 2008. Transcriptome sequencing with the Genome Sequencer FLX system. *Nat Methods* **5**:vi-vii
158. **Johnson, B. P., B. J. Jensen, E. M. Ransom, K. A. Heinemann, K. M. Vannatta, K. A. Eglund, and P. G. Eglund.** 2009. Interspecies signaling between *Veillonella atypica* and *Streptococcus gordonii* requires the transcription factor CcpA. *J Bacteriol* **191**:5563-5565.
159. **Jumpertz, R., D. S. Le, P. J. Turnbaugh, C. Trinidad, C. Bogardus, J. I. Gordon, and J. Krakoff.** 2011. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am J Clin Nutr* **94**:58-65.
160. **Jünemann, S., K. Prior, R. Szczepanowski, I. Harks, B. Ehmke, A. Goesmann, J. Stoye, and D. Harmsen.** 2012. Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PloS one* **7**:e41606.
161. **Kaddurah-Daouk, R., B. S. Kristal, and R. M. Weinshilboum.** 2008. Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol* **48**:653-683.
162. **Kadioglu, A., J. N. Weiser, J. C. Paton, and P. W. Andrew.** 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* **6**:288-301.
163. **Kandler, O.** 1983. Carbohydrate metabolism in lactic acid bacteria. *Antonie Van Leeuwenhoek* **49**:209-224.
164. **Kanehisa, M.** 2002. The KEGG database. *Novartis Foundation symposium* **247**:91-101; discussion 101-103, 119-128, 244-152.
165. **Kanehisa, M., and S. Goto.** 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**:27-30.
166. **Kant, R., J. Blom, A. Palva, R. J. Siezen, and W. M. de Vos.** 2011. Comparative genomics of *Lactobacillus*. *Microb Biotechnol* **4**:323-332.
167. **Karczewski, J., F. J. Troost, I. Konings, J. Dekker, M. Kleerebezem, R. J. Brummer, and J. M. Wells.** 2010. Regulation of human epithelial tight junction proteins by *Lactobacillus plantarum* *in vivo* and protective effects on the epithelial barrier. *Am J Physiol Gastrointest Liver Physiol* **298**:G851-859.
168. **Kawai, T., and S. Akira.** 2011. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34**:637-650.
169. **Keijser, B. J., E. Zaura, S. M. Huse, J. M. van der Vossen, F. H. Schuren, R. C. Montijn, J. M. ten Cate, and W. Crielaard.** 2008. Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* **87**:1016-1020.
170. **Kelly, D., M. I. Delday, and I. Mulder.** 2012. Microbes and microbial effector molecules in treatment of inflammatory disorders. *Immunol Rev* **245**:27-44.
171. **Kircher, M., and J. Kelso.** 2010. High-throughput DNA sequencing--concepts and limitations. *Bioessays* **32**:524-536.

172. **Kiyono, H., J. R. McGhee, M. J. Wannemuehler, and S. M. Michalek.** 1982. Lack of oral tolerance in C3H/HeJ mice. *J Exp Med* **155**:605-610.
173. **Klaassens, E. S., W. M. de Vos, and E. E. Vaughan.** 2007. Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl Environ Microbiol* **73**:1388-1392.
174. **Kleerebezem, M.** 2004. Quorum sensing control of lantibiotic production; nisin and subtilin autoregulate their own biosynthesis. *Peptides* **25**:1405-1414.
175. **Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, et al.** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* **100**:1990-1995.
176. **Kleerebezem, M., and E. E. Vaughan.** 2009. Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity. *Annu Rev Microbiol* **63**:269-290.
177. **Klein, C., C. Kaletta, and K. D. Entian.** 1993. Biosynthesis of the lantibiotic subtilin is regulated by a histidine kinase/response regulator system. *Appl Environ Microbiol* **59**:296-303.
178. **Kolmeder, C. A., M. de Been, J. Nikkila, I. Ritamo, J. Matto, L. Valmu, J. Salojarvi, A. Palva, A. Salonen, and W. M. de Vos.** 2012. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PloS one* **7**:e29913.
179. **Koonin, E. V., and Y. I. Wolf.** 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**:6688-6719.
180. **Kopylova, E., L. Noe, and H. Touzet.** 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**:3211-3217.
181. **Kramer, J. G., and F. L. Singleton.** 1992. Variations in rRNA content of marine *Vibrio* spp. during starvation-survival and recovery. *Appl Environ Microbiol* **58**:201-207.
182. **Kreft, J., and J. A. Vazquez-Boland.** 2001. Regulation of virulence genes in *Listeria*. *Int J Med Microbiol* **291**:145-157.
183. **Kreikemeyer, B., M. D. Boyle, B. A. Buttaro, M. Heinemann, and A. Podbielski.** 2001. Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. *Mol Microbiol* **39**:392-406.
184. **Kubista, M., J. M. Andrade, M. Bengtsson, A. Forootan, J. Jonak, K. Lind, R. Sindelka, R. Sjoback, B. Sjogreen, L. Strombom, A. Stahlberg, and N. Zoric.** 2006. The real-time polymerase chain reaction. *Mol Aspects Med* **27**:95-125.
185. **Kuczynski, J., C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers, and R. Knight.** 2012. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* **13**:47-58.
186. **Kumar, P. S., M. R. Brooker, S. E. Dowd, and T. Camerlengo.** 2011. Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PloS one* **6**:e20956.
187. **Kunji, E. R., G. Fang, C. M. Jeronimus-Stratingh, A. P. Bruins, B. Poolman, and W. N. Konings.** 1998. Reconstruction of the proteolytic pathway for use of beta-casein by *Lactococcus lactis*. *Mol Microbiol* **27**:1107-1118.
188. **Kurokawa, K., T. Itoh, T. Kuwahara, K. Oshima, H. Toh, A. Toyoda, H. Takami, H. Morita, V. K. Sharma, T. P. Srivastava, T. D. Taylor, H. Noguchi, H. Mori, Y. Ogura, D. S. Ehrlich, et al.** 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**:169-181.
189. **Kutahya, O. E., M. J. Starrenburg, J. L. Rademaker, C. H. Klaassen, J. E. van Hylckama Vlieg, E. J. Smid, and M. Kleerebezem.** 2011. High-resolution amplified fragment length polymorphism typing of *Lactococcus lactis* strains enables identification of genetic markers for subspecies-related phenotypes. *Appl. Environ. Microbiol.* **77**:5192-5198.
190. **Lagier, J. C., F. Armougom, M. Million, P. Hugon, I. Pagnier, C. Robert, F. Bittar, G. Fournous, G. Gimenez, M. Maraninchi, J. F. Trape, E. V. Koonin, B. La Scola, and D. Raoult.** 2012. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* **18**:1185-1193.
191. **Lamy, M. C., M. Zouine, J. Fert, M. Vergassola, E. Couve, E. Pellegrini, P. Glaser, F. Kunst, T. Msadek, P. Trieu-Cuot, and C. Poyart.** 2004. CovS/CovR of group B *streptococcus*: a two-component global regulatory system involved in virulence. *Mol Microbiol* **54**:1250-1268.
192. **Lane, D. J.** 1991. 16S/23S rRNA sequencing, p. 115-175. *In* M. G. E. Stackebrandt (ed.), *Nucleic acid techniques in bacterial systematics*. Academic Press, Chichester, England.

193. **Langerholc, T., P. A. Maragkoudakis, J. Wollgast, L. Gradisnik, and A. Cencic.** 2011. Novel and established intestinal cell line models – An indispensable tool in food science and nutrition. *Trends Food Sci. Technol.* **22**, Supplement 1:S11-S20.
194. **Larsen, N., F. K. Vogensen, F. W. van den Berg, D. S. Nielsen, A. S. Andreasen, B. K. Pedersen, W. A. Al-Soud, S. J. Sorensen, L. H. Hansen, and M. Jakobsen.** 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS one* **5**:e9085.
195. **Lawson, R. D., and W. J. Coyle.** 2010. The noncolonic microbiome: does it really matter? *Curr Gastroenterol Rep* **12**:259-262.
196. **Lee, Y. K., J. S. Menezes, Y. Umesaki, and S. K. Mazmanian.** 2011. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A* **108** Suppl 1:4615-4622.
197. **Lefebvre, T., and M. J. Stanhope.** 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol* **8**:R71.
198. **Leimena, M. M., M. Wels, R. S. Bongers, E. J. Smid, E. G. Zoetendal, and M. Kleerebezem.** 2012. Comparative analysis of *Lactobacillus plantarum* WCFS1 transcriptomes by using DNA microarray and next-generation sequencing technologies. *Appl Environ Microbiol* **78**:4141-4148.
199. **Lelouard, H., M. Fallet, B. de Bovis, S. Meresse, and J. P. Gorvel.** 2012. Peyer's patch dendritic cells sample antigens by extending dendrites through M cell-specific transcellular pores. *Gastroenterology* **142**:592-601 e593.
200. **Lelouard, H., S. Henri, B. De Bovis, B. Mugnier, A. Chollat-Namy, B. Malissen, S. Meresse, and J. P. Gorvel.** 2010. Pathogenic bacteria and dead cells are internalized by a unique subset of Peyer's patch dendritic cells that express lysozyme. *Gastroenterology* **138**:173-184 e171-173.
201. **Lepage, P., P. Seksik, M. Sutren, M. F. de la Cochetiere, R. Jian, P. Marteau, and J. Dore.** 2005. Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm Bowel Dis* **11**:473-480.
202. **Lepš, J., and P. Šmilauer.** 2003. *Multivariate Analysis of Ecological Data using CANOCO.* Cambridge University Press.
203. **Leser, T. D., and L. Molbak.** 2009. Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host. *Environ Microbiol* **11**:2194-2206.
204. **Lessard, C., A. Cochu, J. D. Lemay, D. Roy, K. Vaillancourt, M. Frenette, S. Moineau, and C. Vadeboncoeur.** 2003. Phosphorylation of *Streptococcus salivarius* lactose permease (LacS) by HPr(His ~ P) and HPr(Ser-P)(His ~ P) and effects on growth. *J Bacteriol* **185**:6764-6772.
205. **Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon.** 2008. Evolution of mammals and their gut microbes. *Science* **320**:1647-1651.
206. **Ley, R. E., C. A. Lozupone, M. Hamady, R. Knight, and J. I. Gordon.** 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* **6**:776-788.
207. **Li, M., B. Wang, M. Zhang, M. Rantalainen, S. Wang, H. Zhou, Y. Zhang, J. Shen, X. Pang, M. Zhang, H. Wei, Y. Chen, H. Lu, J. Zuo, M. Su, et al.** 2008. Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci U S A* **105**:2117-2122.
208. **Liu, Y. Q., and Y. J. Pan.** 2009. Application of FT-ICR MS for the Study of Protein Complexes. *Appl. Spectrosc. Rev.* **44**:231-244.
209. **Louis, P., and H. J. Flint.** 2009. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett* **294**:1-8.
210. **Lozupone, C., and R. Knight.** 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**:8228-8235.
211. **Lozupone, C. A., M. Hamady, S. T. Kelley, and R. Knight.** 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73**:1576-1585.
212. **Lozupone, C. A., J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight.** 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**:220-230.
213. **Lubelski, J., R. Rink, R. Khusainov, G. N. Moll, and O. P. Kuipers.** 2008. Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cell Mol Life Sci* **65**:455-476.
214. **Luca-Harari, B., J. Darenberg, S. Neal, T. Siljander, L. Strakova, A. Tanna, R. Creti, K. Ekelund, M. Koliou, P. T. Tassios, M. van der Linden, M. Straut, J. Vuopio-Varkila, A.**

- Bouvet, A. Efstratiou, et al. 2009. Clinical and microbiological characteristics of severe *Streptococcus pyogenes* disease in Europe. *J Clin Microbiol* **47**:1155-1165.
215. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res* **32**:1363-1371.
216. Lukjancenko, O., D. W. Ussery, and T. M. Wassenaar. 2012. Comparative genomics of *Bifidobacterium*, *Lactobacillus* and related probiotic genera. *Microb Ecol* **63**:651-673.
217. Macpherson, A. J., and N. L. Harris. 2004. Interactions between commensal intestinal bacteria and the immune system. *Nat Rev Immunol* **4**:478-485.
218. Mane, S. P., C. Evans, K. L. Cooper, O. R. Crasta, O. Folkerts, S. K. Hutchison, T. T. Harkins, D. Thierry-Mieg, J. Thierry-Mieg, and R. V. Jensen. 2009. Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing. *BMC Genomics* **10**:264.
219. Marchesi, J. R. 2011. Human distal gut microbiome. *Environ Microbiol* **13**:3088-3102.
220. Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-380.
221. Maron, P. A., L. Ranjard, C. Mougel, and P. Lemanceau. 2007. Metaproteomics: a new approach for studying functional microbial ecology. *Microb Ecol* **53**:486-493.
222. Marteau, P., M. Minekus, R. Havenaar, and J. H. Huis in't Veld. 1997. Survival of lactic acid bacteria in a dynamic model of the stomach and small intestine: validation and the effects of bile. *J. Dairy Sci.* **80**:1031-1037.
223. Martin, F. P., N. Sprenger, I. K. Yap, Y. Wang, R. Bibiloni, F. Rochat, S. Rezzi, C. Cherbut, S. Kochhar, J. C. Lindon, E. Holmes, and J. K. Nicholson. 2009. Panorganismal gut microbiome-host metabolic crosstalk. *J Proteome Res* **8**:2090-2105.
224. Mascher, T., D. Zahner, M. Merai, N. Balmelle, A. B. de Saizieu, and R. Hakenbeck. 2003. The *Streptococcus pneumoniae* cia regulon: CiaR target sites and transcription profile analysis. *J Bacteriol* **185**:60-70.
225. Mashburn-Warren, L., D. A. Morrison, and M. J. Federle. 2010. A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol Microbiol* **78**:589-606.
226. Maslowski, K. M., and C. R. Mackay. 2011. Diet, gut microbiota and immune responses. *Nat Immunol* **12**:5-9.
227. Matsheka, M. I., A. J. Lastovica, H. Zappe, and B. G. Elisha. 2006. The use of (GTG)₅ oligonucleotide as an RAPD primer to type *Campylobacter concisus*. *Lett. Appl. Microbiol.* **42**:600-605.
228. Matsuki, T., K. Watanabe, J. Fujimoto, Y. Miyamoto, T. Takada, K. Matsumoto, H. Oyaizu, and R. Tanaka. 2002. Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Appl Environ Microbiol* **68**:5445-5451.
229. Matsuki, T., K. Watanabe, J. Fujimoto, T. Takada, and R. Tanaka. 2004. Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces. *Appl Environ Microbiol* **70**:7220-7228.
230. McCartney, A. L., W. Wenzhi, and G. W. Tannock. 1996. Molecular analysis of the composition of the bifidobacterial and *Lactobacillus* microflora of humans. *Appl Environ Microbiol* **62**:4608-4613.
231. Meijerink, M., M. L. Ferrando, G. Lammers, N. Taverne, H. E. Smith, and J. M. Wells. 2012. Immunomodulatory effects of *Streptococcus suis* capsule type on human dendritic cell responses, phagocytosis and intracellular survival. *PloS one* **7**:e35849.
232. Meijerink, M., S. van Hemert, N. Taverne, M. Wels, P. de Vos, P. A. Bron, H. F. Savelkoul, J. van Bilsen, M. Kleerebezem, and J. M. Wells. 2010. Identification of genetic loci in *Lactobacillus plantarum* that modulate the immune response of dendritic cells using comparative genome hybridization. *PloS one* **5**:e10632.
233. Meijerink, M., J. M. Wells, N. Taverne, M. L. de Zeeuw Brouwer, B. Hilhorst, K. Venema, and J. van Bilsen. 2012. Immunomodulatory effects of potential probiotics in a mouse peanut sensitization model. *FEMS Immunol Med Microbiol* **65**:488-496.
234. Metzker, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* **11**:31-46.

235. **Monteils, V., L. Cauquil, S. Combes, J. J. Godon, and T. Gidenne.** 2008. Potential core species and satellite species in the bacterial community within the rabbit caecum. *FEMS Microbiol Ecol* **66**:620-629.
236. **Morales, S. E., and W. E. Holben.** 2011. Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? *FEMS Microbiol. Ecol.* **75**:2-16.
237. **Morgan, E., R. Varro, H. Sepulveda, J. A. Ember, J. Apgar, J. Wilson, L. Lowe, R. Chen, L. Shivraj, A. Agadir, R. Campos, D. Ernst, and A. Gaur.** 2004. Cytometric bead array: a multiplexed assay platform with applications in various areas of biology. *Clin Immunol* **110**:252-266.
238. **Morozova, O., M. Hirst, and M. A. Marra.** 2009. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* **10**:135-151.
239. **Muniesa, M., A. R. Blanch, F. Lucena, and J. Jofre.** 2005. Bacteriophages may bias outcome of bacterial enrichment cultures. *Appl Environ Microbiol* **71**:4269-4275.
240. **Murphy, N. R., and R. J. Hellwig.** 1996. Improved nucleic acid organic extraction through use of a unique gel barrier material. *BioTechniques* **21**:934-936, 938-939.
241. **Nagano, Y., K. Itoh, and K. Honda.** 2012. The induction of Treg cells by gut-indigenous *Clostridium*. *Curr Opin Immunol* **24**:392-397.
242. **Neish, A. S.** 2009. Microbes in gastrointestinal health and disease. *Gastroenterology* **136**:65-80.
243. **Ng, S. K., and I. R. Hamilton.** 1971. Lactate metabolism by *Veillonella parvula*. *J Bacteriol* **105**:999-1005.
244. **Nicholson, J. K., and J. C. Lindon.** 2008. Systems biology: Metabonomics. *Nature* **455**:1054-1056.
245. **Niu, B., L. Fu, S. Sun, and W. Li.** 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11**:187.
246. **Novak, R., E. Charpentier, J. S. Braun, and E. Tuomanen.** 2000. Signal transduction by a death signal peptide: uncovering the mechanism of bacterial killing by penicillin. *Mol. Cell* **5**:49-57.
247. **O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**:4007-4021.
248. **Oresic, M.** 2009. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutr Metab Cardiovasc Dis* **19**:816-824.
249. **Ozsolak, F., A. R. Platt, D. R. Jones, J. G. Reifengerger, L. E. Sass, P. McInerney, J. F. Thompson, J. Bowers, M. Jarosz, and P. M. Milos.** 2009. Direct RNA sequencing. *Nature* **461**:814-818.
250. **Page, R. D.** 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**:357-358.
251. **Paliy, O., H. Kenche, F. Abernathy, and S. Michail.** 2009. High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol* **75**:3572-3579.
252. **Palmer, C., E. M. Bik, M. B. Eisen, P. B. Eckburg, T. R. Sana, P. K. Wolber, D. A. Relman, and P. O. Brown.** 2006. Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res* **34**:e5.
253. **Park, J., A. Kerner, M. A. Burns, and X. N. Lin.** 2011. Microdroplet-enabled highly parallel co-cultivation of microbial communities. *PloS one* **6**:e17019.
254. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, et al.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**:848-852.
255. **Passalacqua, K. D., A. Varadarajan, B. D. Ondov, D. T. Okou, M. E. Zwick, and N. H. Bergman.** 2009. Structure and complexity of a bacterial transcriptome. *J Bacteriol* **191**:3203-3211.
256. **Pei, Z., E. J. Bini, L. Yang, M. Zhou, F. Francois, and M. J. Blaser.** 2004. Bacterial biota in the human distal esophagus. *Proc Natl Acad Sci U S A* **101**:4250-4255.
257. **Peris-Bondia, F., A. Latorre, A. Artacho, A. Moya, and G. D'Auria.** 2011. The active human gut microbiota differs from the total microbiota. *PloS one* **6**:e22448.
258. **Peterson, J., S. Garges, M. Giovanni, P. McInnes, L. Wang, J. A. Schloss, V. Bonazzi, J. E. McEwen, K. A. Wetterstrand, C. Deal, C. C. Baker, V. Di Francesco, T. K. Howcroft, R. W. Karp, R. D. Lunsford, et al.** 2009. The NIH Human Microbiome Project. *Genome Res* **19**:2317-2323.

259. **Picariello, G., P. Ferranti, G. Mamone, I. Klouckova, Y. Mechref, M. V. Novotny, and F. Addeo.** 2012. Gel-free shotgun proteomic analysis of human milk. *J Chromatogr A* **1227**:219-233.
260. **Poyart, C., M. C. Lamy, C. Boumaila, F. Fiedler, and P. Trieu-Cuot.** 2001. Regulation of D-alanyl-lipoteichoic acid biosynthesis in *Streptococcus agalactiae* involves a novel two-component regulatory system. *J Bacteriol* **183**:6324-6334.
261. **Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glockner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**:7188-7196.
262. **Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, et al.** 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**:59-65.
263. **Quach, D., N. M. van Sorge, S. A. Kristian, J. D. Bryan, D. W. Shelver, and K. S. Doran.** 2009. The CiaR response regulator in group B *Streptococcus* promotes intracellular survival and resistance to innate immune defenses. *J Bacteriol* **191**:2023-2032.
264. **Raes, J., and P. Bork.** 2008. Molecular eco-systems biology: towards an understanding of community function. *Nat Rev Microbiol* **6**:693-699.
265. **Rajilić-Stojanović, M.** 2007. Diversity of Human Gastrointestinal Microbiota - Novel Perspectives from High Throughput Analyses. Wageningen University, Wageningen, The Netherlands.
266. **Rajilić-Stojanović, M., H. G. Heilig, D. Molenaar, K. Kajander, A. Surakka, H. Smidt, and W. M. de Vos.** 2009. Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* **11**:1736-1751.
267. **Rajilić-Stojanović, M., H. G. Heilig, S. Tims, E. G. Zoetendal, and W. M. de Vos.** 2012. Long-term monitoring of the human intestinal microbiota composition. *Environ Microbiol*, In press.
268. **Rajilić-Stojanović, M., H. Smidt, and W. M. de Vos.** 2007. Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol* **9**:2125-2136.
269. **Reeder, J., and R. Knight.** 2009. The 'rare biosphere': a reality check. *Nat Methods* **6**:636-637.
270. **Rescigno, M., M. Urbano, B. Valzasina, M. Francolini, G. Rotta, R. Bonasio, F. Granucci, J. P. Kraehenbuhl, and P. Ricciardi-Castagnoli.** 2001. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat Immunol* **2**:361-367.
271. **Rimoldi, M., M. Chieppa, V. Salucci, F. Avogadri, A. Sonzogni, G. M. Sampietro, A. Nespoli, G. Viale, P. Allavena, and M. Rescigno.** 2005. Intestinal immune homeostasis is regulated by the crosstalk between epithelial cells and dendritic cells. *Nat Immunol* **6**:507-514.
272. **Rinttila, T., A. Kassinen, E. Malinen, L. Krogius, and A. Palva.** 2004. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* **97**:1166-1177.
273. **Rogers, A. H.** 1973. The vitamin requirements of some oral streptococci. *Arch Oral Biol* **18**:227-232.
274. **Rogosa, M.** 1956. A selective medium for the isolation and enumeration of the *Veillonella* from the oral cavity. *J Bacteriol* **72**:533-536.
275. **Rogosa, M., R. J. Fitzgerald, M. E. Mackintosh, and A. J. Beaman.** 1958. Improved medium for selective isolation of *Veillonella*. *J Bacteriol* **76**:455-456.
276. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, et al.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**:2541-2547.
277. **Rooijers, K., C. Kolmeder, C. Juste, J. Dore, M. de Been, S. Boeren, P. Galan, C. Beauvallet, W. M. de Vos, and P. J. Schaap.** 2011. An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics* **12**:6.
278. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**:365-386.
279. **Rudney, J. D., Y. Pan, and R. Chen.** 2003. Streptococcal diversity in oral biofilms with respect to salivary function. *Arch Oral Biol* **48**:475-493.
280. **Rusniok, C., E. Couve, V. Da Cunha, R. El Gana, N. Zidane, C. Bouchier, C. Poyart, R. Leclercq, P. Trieu-Cuot, and P. Glaser.** 2010. Genome sequence of *Streptococcus gallolyticus*: insights into its adaptation to the bovine rumen and its ability to cause endocarditis. *J Bacteriol* **192**:2266-2276.

281. **Russell, R. R., J. Aduse-Opoku, I. C. Sutcliffe, L. Tao, and J. J. Ferretti.** 1992. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. *J Biol Chem* **267**:4631-4637.
282. **Said, H. M., A. Ortiz, M. P. Moyer, and N. Yanagawa.** 2000. Riboflavin uptake by human-derived colonic epithelial NCM460 cells. *Am J Physiol Cell Physiol* **278**:C270-276.
283. **Salonen, A., J. Nikkila, J. Jalanka-Tuovinen, O. Immonen, M. Rajilić-Stojanović, R. A. Kekkonen, A. Palva, and W. M. de Vos.** 2010. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods* **81**:127-134.
284. **Salonen, A., J. Salojarvi, L. Lahti, and W. M. de Vos.** 2012. The adult intestinal core microbiota is determined by analysis depth and health status. *Clin Microbiol Infect* **18 Suppl 4**:16-20.
285. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: A laboratory manual*, Second Edition ed, vol. 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
286. **Samuel, B. S., E. E. Hansen, J. K. Manchester, P. M. Coutinho, B. Henrissat, R. Fulton, P. Latreille, K. Kim, R. K. Wilson, and J. I. Gordon.** 2007. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci U S A* **104**:10643-10648.
287. **Sava, I. G., E. Heikens, and J. Huebner.** 2010. Pathogenesis and immunity in enterococcal infections. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and In* **16**:533-540.
288. **Sboner, A., X. J. Mu, D. Greenbaum, R. K. Auerbach, and M. B. Gerstein.** 2011. The real cost of sequencing: higher than you think! *Genome Biol* **12**:125.
289. **Scanlan, P. D., and J. R. Marchesi.** 2008. Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. *Isme J* **2**:1183-1193.
290. **Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**:7537-7541.
291. **Schmeisser, C., H. Steele, and W. R. Streit.** 2007. Metagenomics, biotechnology with non-culturable microbes. *Appl. Microbiol. Biotechnol.* **75**:955-962.
292. **Schmidt, D., M. D. Wilson, C. Spyrou, G. D. Brown, J. Hadfield, and D. T. Odom.** 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**:240-248.
293. **Schwartz, A., G. Le Blay, and M. Blaut.** 2000. Quantification of different *Eubacterium* spp. in human fecal samples with species-specific 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* **66**:375-382.
294. **Sherstnev, A., C. Duc, C. Cole, V. Zacharaki, C. Hornyik, F. Oszolak, P. M. Milos, G. J. Barton, and G. G. Simpson.** 2012. Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. *Nat Struct Mol Biol* **19**:845-852.
295. **Shimamura, S., F. Abe, N. Ishibashi, H. Miyakawa, T. Yaeshima, T. Araya, and M. Tomita.** 1992. Relationship between oxygen sensitivity and oxygen metabolism of *Bifidobacterium* species. *J Dairy Sci* **75**:3296-3306.
296. **Shokralla, S., J. L. Spall, J. F. Gibson, and M. Hajibabaei.** 2012. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* **21**:1794-1805.
297. **Siezen, R. J., J. R. Bayjanov, G. E. Felis, M. R. van der Sijde, M. Starrenburg, D. Molenaar, M. Wels, S. A. van Hijum, and J. E. van Hylckama Vlieg.** 2011. Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. *Microb Biotechnol* **4**:383-402.
298. **Siggins, A., E. Gunnigle, and F. Abram.** 2012. Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol. Ecol.* **80**:265-280.
299. **Simpson, C. L., and R. R. Russell.** 1998. Intracellular alpha-amylase of *Streptococcus mutans*. *J Bacteriol* **180**:4711-4717.
300. **Smelt, M. J., B. J. de Haan, P. A. Bron, I. van Swam, M. Meijerink, J. M. Wells, M. M. Faas, and P. de Vos.** 2012. *L. plantarum*, *L. salivarius*, and *L. lactis* attenuate Th2 responses and increase Treg frequencies in healthy mice in a strain dependent manner. *PloS one* **7**:e47244.
301. **Sokol, H., B. Pigneur, L. Watterlot, O. Lakhdari, L. G. Bermudez-Humaran, J. J. Gratadoux, S. Blugeon, C. Bridonneau, J. P. Furet, G. Corthier, C. Grangette, N. Vasquez, P. Pochart, G. Trugnan, G. Thomas, et al.** 2008. *Faecalibacterium prausnitzii* is an anti-

- inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**:16731-16736.
302. **Sommer, F., and F. Backhed.** 2013. The gut microbiota - masters of host development and physiology. *Nat Rev Microbiol.* In press.
303. **Sonnenburg, J. L., C. T. Chen, and J. I. Gordon.** 2006. Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. *PLoS Biol* **4**:e413.
304. **Sorek, R., and P. Cossart.** 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**:9-16.
305. **Sorensen, S. R., J. Rasmussen, C. S. Jacobsen, O. S. Jacobsen, R. K. Juhler, and J. Aamand.** 2005. Elucidating the key member of a linuron-mineralizing bacterial community by PCR and reverse transcription-PCR denaturing gradient gel electrophoresis 16S rRNA gene fingerprinting and cultivation. *Appl Environ Microbiol* **71**:4144-4148.
306. **Stappenbeck, T. S., L. V. Hooper, and J. I. Gordon.** 2002. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A* **99**:15451-15455.
307. **Stock, A. M., V. L. Robinson, and P. N. Goudreau.** 2000. Two-component signal transduction. *Annu Rev Biochem* **69**:183-215.
308. **Stolaki, M., W. M. De Vos, M. Kleerebezem, and E. Zoetendal.** 2012. Lactic Acid Bacteria in the Gut. In S. Lahtinen, A. C. Ouwehand, S. Salminen, and A. Von Wright (ed.), *Lactic Acid Bacteria, Microbiological and functional aspects.* CRC Press, Taylor & Francis Group.
309. **Suau, A., R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins, and J. Dore.** 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* **65**:4799-4807.
310. **Sun, Y., Y. Cai, L. Liu, F. Yu, M. L. Farrell, W. McKendree, and W. Farmerie.** 2009. ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res* **37**:e76.
311. **Sundquist, A., S. Bigdeli, R. Jalili, M. L. Druzin, S. Waller, K. M. Pullen, Y. Y. El-Sayed, M. M. Taslimi, S. Batzoglou, and M. Ronaghi.** 2007. Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol* **7**:108.
312. **Surana, N. K., and D. L. Kasper.** 2012. The yin yang of bacterial polysaccharides: lessons learned from *B. fragilis* PSA. *Immunol Rev* **245**:13-26.
313. **Suzuki, M. T., L. T. Taylor, and E. F. DeLong.** 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. *Appl Environ Microbiol* **66**:4605-4614.
314. **Swann, J., Y. Wang, L. Abecia, A. Costabile, K. Tuohy, G. Gibson, D. Roberts, J. Sidaway, H. Jones, I. D. Wilson, J. Nicholson, and E. Holmes.** 2009. Gut microbiome modulates the toxicity of hydrazine: a metabonomic study. *Mol Biosyst* **5**:351-355.
315. **Tanaka, K., and H. Ishikawa.** 2004. Role of intestinal bacterial flora in oral tolerance induction. *Histol Histopathol* **19**:907-914.
316. **Tannock, G. W., K. Munro, R. Bibiloni, M. A. Simon, P. Hargreaves, P. Gopal, H. Harmsen, and G. Welling.** 2004. Impact of consumption of oligosaccharide-containing biscuits on the fecal microbiota of humans. *Appl Environ Microbiol* **70**:2129-2136.
317. **Tap, J., S. Mondot, F. Levenez, E. Pelletier, C. Caron, J. P. Furet, E. Ugarte, R. Munoz-Tamayo, D. L. Paslier, R. Nalin, J. Dore, and M. Leclerc.** 2009. Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* **11**:2574-2584.
318. **Taroncher-Oldenburg, G.** 2009. RNA direct. *SciBX* **2**:1-2.
319. **Tartar, A., M. M. Wheeler, X. Zhou, M. R. Coy, D. G. Boucias, and M. E. Scharf.** 2009. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuels* **2**:25.
320. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**:33-36.
321. **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**:22-28.
322. **Tims, S., C. Derom, D. M. Jonkers, R. Vlietinck, W. H. Saris, M. Kleerebezem, W. M. de Vos, and E. G. Zoetendal.** 2013. Microbiota conservation and BMI signatures in adult monozygotic twins. *Isme J* **7**:707-717.

323. **Tjaden, B., R. M. Saxena, S. Stolyar, D. R. Haynor, E. Kolker, and C. Rosenow.** 2002. Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* **30**:3732-3738.
324. **Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, et al.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539-547.
325. **Tran-Winkler, H. J., J. F. Love, I. Gryllos, and M. R. Wessels.** 2011. Signal transduction through CsrRS confers an invasive phenotype in group A *Streptococcus*. *PLoS Pathog* **7**:e1002361.
326. **Tremaroli, V., and F. Backhed.** 2012. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**:242-249.
327. **Troy, E. B., and D. L. Kasper.** 2010. Beneficial effects of *Bacteroides fragilis* polysaccharides on the immune system. *Front Biosci* **15**:25-34.
328. **Turnbaugh, P. J., and J. I. Gordon.** 2008. An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**:708-713.
329. **Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, and J. I. Gordon.** 2009. A core gut microbiome in obese and lean twins. *Nature* **457**:480-484.
330. **Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon.** 2007. The human microbiome project. *Nature* **449**:804-810.
331. **Turnbaugh, P. J., C. Quince, J. J. Faith, A. C. McHardy, T. Yatsunenko, F. Niazi, J. Affourtit, M. Egholm, B. Henrissat, R. Knight, and J. I. Gordon.** 2010. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci U S A* **107**:7503-7508.
332. **Ulluwishewa, D., R. C. Anderson, W. C. McNabb, P. J. Moughan, J. M. Wells, and N. C. Roy.** 2011. Regulation of tight junction permeability by intestinal bacteria and dietary components. *J Nutr* **141**:769-776.
333. **Urich, T., A. Lanzen, J. Qi, D. H. Huson, C. Schleper, and S. C. Schuster.** 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS one* **3**:e2527.
334. **Utada, S., P. M. Dunman, D. Macapagal, E. Murphy, S. J. Projan, V. K. Singh, R. K. Jayaswal, and B. J. Wilkinson.** 2003. Genome-wide transcriptional profiling of the response of *Staphylococcus aureus* to cell-wall-active antibiotics reveals a cell-wall-stress stimulon. *Microbiology* **149**:2719-2732.
335. **van Baarlen, P., F. Troost, C. van der Meer, G. Hooiveld, M. Boekschoten, R. J. Brummer, and M. Kleerebezem.** 2011. Human mucosal *in vivo* transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways. *Proc Natl Acad Sci U S A* **108 Suppl 1**:4562-4569.
336. **van Baarlen, P., F. J. Troost, S. van Hemert, C. van der Meer, W. M. de Vos, P. J. de Groot, G. J. Hooiveld, R. J. Brummer, and M. Kleerebezem.** 2009. Differential NF-kappaB pathways induction by *Lactobacillus plantarum* in the duodenum of healthy humans correlating with immune tolerance. *Proc Natl Acad Sci U S A* **106**:2371-2376.
337. **van Bokhorst-van de Veen, H., I. van Swam, M. Wels, P. A. Bron, and M. Kleerebezem.** 2012. Congruent strain specific intestinal persistence of *Lactobacillus plantarum* in an intestine-mimicking *in vitro* system and in human volunteers. *PLoS one* **7**:e44588.
338. **Van den Abbeele, P., S. Roos, V. Eeckhaut, D. A. MacKenzie, M. Derde, W. Verstraete, M. Marzorati, S. Possemiers, B. Vanhoecke, F. Van Immerseel, and T. Van de Wiele.** 2012. Incorporating a mucosal environment in a dynamic gut model results in a more representative colonization by lactobacilli. *Microb Biotechnol* **5**:106-115.
339. **van den Bogert, B., W. M. de Vos, E. G. Zoetendal, and M. Kleerebezem.** 2011. Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. *Appl Environ Microbiol* **77**:2071-2080.
340. **van den Bogert, B., O. Erkus, J. Boekhorst, M. de Goffau, E. J. Smid, E. G. Zoetendal, and M. Kleerebezem.** 2013. Diversity of human small intestinal *Streptococcus* and *Veillonella* populations. *FEMS Microbiol Ecol* **85**:376-388.
341. **Van den Bogert, B., M. M. Leimena, W. M. De Vos, E. G. Zoetendal, and M. Kleerebezem.** 2011. Functional Intestinal Metagenomics, p. 170-190. *In* F. J. De Bruin (ed.), *Handbook of Molecular Microbial Ecology*, vol. II: Metagenomics in Different Habitats. Wiley-Blackwell.

342. **van Kraaij, C., W. M. de Vos, R. J. Siezen, and O. P. Kuipers.** 1999. Lantibiotics: biosynthesis, mode of action and applications. *Natural product reports* **16**:575-587.
343. **van Passel, M. W., R. Kant, E. G. Zoetendal, C. M. Plugge, M. Derrien, S. A. Malfatti, P. S. Chain, T. Woyke, A. Palva, W. M. de Vos, and H. Smidt.** 2011. The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PloS one* **6**:e16876.
344. **van Schaik, W., and R. J. Willems.** 2010. Genome-based insights into the evolution of enterococci. *Clin Microbiol Infect.*
345. **van Veen, S. Q., E. C. Claas, and E. J. Kuijper.** 2010. High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. *J Clin Microbiol* **48**:900-907.
346. **Vaughan, E. E., F. Schut, H. G. Heilig, E. G. Zoetendal, W. M. de Vos, and A. D. Akkermans.** 2000. A molecular view of the intestinal ecosystem. *Curr Issues Intest Microbiol* **1**:1-12.
347. **Veloo, A. C., M. Knoester, J. E. Degener, and E. J. Kuijper.** 2011. Comparison of two matrix-assisted laser desorption ionisation-time of flight mass spectrometry methods for the identification of clinically relevant anaerobic bacteria. *Clin Microbiol Infect* **17**:1501-1506.
348. **Verberkmoes, N. C., A. L. Russell, M. Shah, A. Godzik, M. Rosenquist, J. Halfvarson, M. G. Lefsrud, J. Apajalahti, C. Tysk, R. L. Hettich, and J. K. Jansson.** 2009. Shotgun metaproteomics of the human distal gut microbiota. *Isme J* **3**:179-189.
349. **Vermes, I., C. Haanen, H. Steffens-Nakken, and C. Reutelingsperger.** 1995. A novel assay for apoptosis. Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V. *J Immunol Methods* **184**:39-51.
350. **Von Wright, A., and L. Axelsson.** 2012. Lactic Acid Bacteria: An Introduction. *In* S. Lahtinen, A. C. Ouwehand, S. Salminen, and A. Von Wright (ed.), *Lactic Acid Bacteria, Microbiological and functional aspects*. CRC Press, Taylor & Francis Group.
351. **Vonk, R. J., R. E. Hagedoorn, R. de Graaff, H. Elzinga, S. Tabak, Y. X. Yang, and F. Stellaard.** 2000. Digestion of so-called resistant starch sources in the human small intestine. *Am J Clin Nutr* **72**:432-438.
352. **Vrieze, A., E. Van Nood, F. Holleman, J. Salojarvi, R. S. Kootte, J. F. Bartelsman, G. M. Dallinga-Thie, M. T. Ackermans, M. J. Serlie, R. Oozeer, M. Derrien, A. Druesne, J. E. Van Hylckama Vlieg, V. W. Bloks, A. K. Groen, et al.** 2012. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143**:913-916 e917.
353. **Walter, J., and R. Ley.** 2011. The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol* **65**:411-429.
354. **Wang, M., S. Ahrne, B. Jeppsson, and G. Molin.** 2005. Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiol Ecol* **54**:219-231.
355. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**:5261-5267.
356. **Warnecke, F., and M. Hess.** 2009. A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J Biotechnol* **142**:91-95.
357. **Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton.** 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189-1191.
358. **Weinstock, G. M.** 2012. Genomic approaches to studying the human microbiota. *Nature* **489**:250-256.
359. **Wen, L., R. E. Ley, P. Y. Volchkov, P. B. Stranges, L. Avanesyan, A. C. Stonebraker, C. Hu, F. S. Wong, G. L. Szot, J. A. Bluestone, J. I. Gordon, and A. V. Chervonsky.** 2008. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**:1109-1113.
360. **Wexler, D. E., D. E. Chenoweth, and P. P. Cleary.** 1985. Mechanism of action of the group A streptococcal C5a inactivator. *Proc Natl Acad Sci U S A* **82**:8144-8148.
361. **Whitehead, T. R., and M. A. Cotta.** 1995. Identification of intracellular amylase activity in *Streptococcus bovis* and *Streptococcus salivarius*. *Curr Microbiol* **30**:143-148.
362. **Whiteley, A. S., S. Jenkins, I. Waite, N. Kresoje, H. Payne, B. Mullan, R. Allcock, and A. O'Donnell.** 2012. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* **91**:80-88.

363. **Wikoff, W. R., A. T. Anfora, J. Liu, P. G. Schultz, S. A. Lesley, E. C. Peters, and G. Siuzdak.** 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* **106**:3698-3703.
364. **Wilmes, P., and P. L. Bond.** 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* **14**:92-97.
365. **Wommack, K. E., J. Bhavsar, and J. Ravel.** 2008. Metagenomics: read length matters. *Appl Environ Microbiol* **74**:1453-1463.
366. **Wong, J. M., R. de Souza, C. W. Kendall, A. Emam, and D. J. Jenkins.** 2006. Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol* **40**:235-243.
367. **Woyke, T., D. Tighe, K. Mavromatis, A. Clum, A. Copeland, W. Schackwitz, A. Lapidus, D. Wu, J. P. McCutcheon, B. R. McDonald, N. A. Moran, J. Bristow, and J. F. Cheng.** 2010. One bacterial cell, one complete genome. *PLoS one* **5**:e10314.
368. **Yamada, T., I. Letunic, S. Okuda, M. Kanehisa, and P. Bork.** 2011. iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* **39**:W412-415.
369. **Yamazaki, S., K. Machii, S. Tsuyuki, H. Momose, T. Kawashima, and K. Ueda.** 1985. Immunological responses to monoassociated *Bifidobacterium longum* and their relation to prevention of bacterial invasion. *Immunology* **56**:43-50.
370. **Yatsunenkov, T., F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, A. C. Heath, B. Warner, J. Reeder, J. Kuczynski, J. G. Caporaso, et al.** 2012. Human gut microbiome viewed across age and geography. *Nature* **486**:222-227.
371. **Yoder-Himes, D. R., P. S. Chain, Y. Zhu, O. Wurtzel, E. M. Rubin, J. M. Tiedje, and R. Sorek.** 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci U S A* **106**:3976-3981.
372. **Zaura, E., B. J. Keijsers, S. M. Huse, and W. Crielaard.** 2009. Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* **9**:259.
373. **Zeng, L., S. Das, and R. A. Burne.** 2011. Genetic analysis of the functions and interactions of components of the LevQRST signal transduction complex of *Streptococcus mutans*. *PLoS one* **6**:e17335.
374. **Zengler, K., G. Toledo, M. Rappe, J. Elkins, E. J. Mathur, J. M. Short, and M. Keller.** 2002. Cultivating the uncultured. *Proc Natl Acad Sci U S A* **99**:15681-15686.
375. **Zhou, Q., X. Qin, M. Qin, and L. Ge.** 2011. Genotypic diversity of *Streptococcus mutans* and *Streptococcus sobrinus* in 3-4-year-old children with severe caries or without caries. *Int J Paediatr Dent* **21**:422-431.
376. **Zoetendal, E. G., A. D. Akkermans, and W. M. de Vos.** 1998. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol* **64**:3854-3859.
377. **Zoetendal, E. G., C. C. Booiijink, E. S. Klaassens, H. G. Heilig, M. Kleerebezem, H. Smidt, and W. M. de Vos.** 2006. Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat Protoc* **1**:954-959.
378. **Zoetendal, E. G., C. T. Collier, S. Koike, R. I. Mackie, and H. R. Gaskins.** 2004. Molecular ecological analysis of the gastrointestinal microbiota: a review. *J Nutr* **134**:465-472.
379. **Zoetendal, E. G., H. G. Heilig, E. S. Klaassens, C. C. Booiijink, M. Kleerebezem, H. Smidt, and W. M. de Vos.** 2006. Isolation of DNA from bacterial samples of the human gastrointestinal tract. *Nat Protoc* **1**:870-873.
380. **Zoetendal, E. G., C. M. Plugge, A. D. Akkermans, and W. M. de Vos.** 2003. *Victivallis vadensis* gen. nov., sp. nov., a sugar-fermenting anaerobe from human faeces. *Int J Syst Evol Microbiol* **53**:211-215.
381. **Zoetendal, E. G., J. Raes, B. van den Bogert, M. Arumugam, C. C. Booiijink, F. J. Troost, P. Bork, M. Wels, W. M. de Vos, and M. Kleerebezem.** 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *Isme J* **6**:1415-1426.
382. **Zoetendal, E. G., M. Rajilić-Stojanović, and W. M. de Vos.** 2008. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* **57**:1605-1615.
383. **Zoetendal, E. G., A. von Wright, T. Vilpponen-Salmela, K. Ben-Amor, A. D. Akkermans, and W. M. de Vos.** 2002. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* **68**:3401-3407.

List of publications

Leimena, MM*, **van den Bogert B***, Boekhorst J, Smid EJ, Zoetendal EG, Kleerebezem M. Congruency of phylogenetic composition and activity patterns in the small intestine. Manuscript in preparation.

Van den Bogert B, Boekhorst J, Smid EJ, Zoetendal EZ, Kleerebezem M. Draft genome sequence of *Veillonella parvula* HSIVP1, isolated from the human small intestine. Manuscript in preparation.

Van den Bogert B, Boekhorst J, Smid EJ, Zoetendal EZ, Kleerebezem M. Draft genome sequence of *Enterococcus gallinarum* HSIEG1, isolated from the human small intestine. Manuscript in preparation.

Pérez Gutiérrez O, **van den Bogert B**, Derrien M, Koopmans S-J, Molenaar D, de Vos WM, Smidt H. Design of a high throughput diagnostic microarray for the characterization of pig gastrointestinal tract microbiota. Manuscript in preparation.

de Goffau M, Fuentes S, **van den Bogert B**, Honkanen H, de Vos WM, Welling G, Hyöty H, Harmsen H. Intestinal microbial development in toddlers with type 1 diabetes differs from healthy age-matched controls. Manuscript in preparation.

Brugman S, Witte M, Schneeberger K, Klein MR, **van den Bogert B**, Boekhorst J, Timmerman HM, Boes M, Kleerebezem M, Nieuwenhuis EES. T lymphocytes control microbial composition by preventing overgrowth of pathogenic bacteria in zebrafish. Manuscript in preparation.

van den Bogert B*, Meijerink, M*, Zoetendal EG, Wells JM, Kleerebezem M. Immunomodulatory properties of *Streptococcus* and *Veillonella* isolates from the human small intestine microbiota. Submitted for publication.

Van den Bogert B, Boekhorst J, Herrmann R, Smid EJ, Zoetendal EZ, Kleerebezem M. Comparative genomics analysis of *Streptococcus* isolates from the human small intestine reveals their adaptation to a highly dynamic ecosystem. Submitted for publication.

Leimena MM*, Ramiro-Garcia J*, Davids M, **van den Bogert B**, Smidt H, Smidt EJ, Boekhorst J, Zoetendal EZ, Schaap PJ, Kleerebezem M. 2013. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC Genomics 14: 530.

van den Bogert B, Erkus O, Boekhorst J, de Goffau M, Smid EJ, Zoetendal EG, Kleerebezem M. 2013. Diversity of human small intestinal *Streptococcus* and *Veillonella* populations. FEMS Microbiol Ecol 85: 376-388.

Ijssennagger N*. Derrien M*, van Doorn GM, Rijnierse A, **van den Bogert B**, Müller M, Dekker J, Kleerebezem M, van der Meer R. 2012. Dietary Heme Alters Microbiota and Mucosa of Mouse Colon without Functional Changes in Host-Microbe Cross-Talk. PLoS One 7:e49868.

Zoetendal EG*, Raes J*, **van den Bogert B**, Arumugam M, Booijink CC, Troost FJ, Bork P, Wels M, de Vos WM, Kleerebezem M. 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. Isme J 6:1415-1426.***

van den Bogert B, de Vos WM, Zoetendal EG, Kleerebezem M. 2011. Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. Appl Environ Microbiol 77:2071-2080.**

van den Bogert, B*, Leimena MM*, de Vos WM, Zoetendal EG, Kleerebezem M. 2011. Functional Intestinal Metagenomics, p. 170-190. In F. J. De Bruin (ed.), Handbook of Molecular Microbial Ecology, vol. Vol II: Metagenomics in Different Habitats. Wiley-Blackwell.

* These authors contributed equally to this work

** Nominated for TIFN publication prize 2011

*** TIFN publication prize 2012

Co-author affiliations

Milkha M. Leimena^{1,2}

Marjolein Meijerink³

Jos Boekhorst^{4,5}

Oylum Erkus^{1,2}

Ruth Herrmann^{1,2}

Marcus de Goffau⁶

Jerry M. Wells³

Willem M. de Vos^{2,7}

Eddy J. Smid^{1,8}

Erwin G. Zoetendal^{1,2}

Michiel Kleerebezem^{1,2,3,5}

¹TI Food and Nutrition, P.O. Box 557, 6700 AN Wageningen, the Netherlands

²Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

³Host-Microbe Interactomics Group, Wageningen University, P.O. box 338, 6700 AH

⁴Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, Nijmegen, Netherlands

⁵NIZO Food Research B.V., P.O. Box 20, 6710 BA Ede

⁶Department of Medical Microbiology, University Medical Center Groningen, Hanzeplein 1, 9700 RB Groningen

⁷Department of Basic Veterinary Medicine, Division of Microbiology and Epidemiology, University of Helsinki, Helsinki, Finland

⁸Laboratory of Food Microbiology, Wageningen University, P.O. Box 8129, 6700 EV Wageningen

Samenvatting

In onze darmen leven veel verschillende micro-organismen, voornamelijk bacteriën, die gezamenlijk 'microbiota' worden genoemd. Door onderzoek weten we dat darmbacteriën bijdragen aan onze gezondheid, wat ertoe heeft geleid dat er veel interesse is om erachter te komen hoe de bacteriën dit doen. De studies hebben zich hoofdzakelijk gericht op de bacteriën die leven in de dikke darm. De bacterie populaties in de dikke darm verschillen echter van de bacteriën die leven in andere plekken van het darmkanaal, zoals de dunne darm. Hoewel de dunne darm de eerste plek is waar ons voedsel in contact komt met darmbacteriën, weten we nog weinig over de bacteriën in onze dunne darm. Dit komt doordat het moeilijk is om monsters te verkrijgen uit de dunne darm, door de ligging tussen de maag en de dikke darm.

Het onderzoek beschreven in dit proefschrift heeft zich gericht op het verkrijgen van een beter inzicht in de samenstelling en dynamiek van de bacteriën in de dunne darm van volwassenen. Ook is onderzocht wat de functionaliteit is van bepaalde dunne darmbacteriën, die bij veel mensen voorkomen, en hoe ze ons immuunsysteem kunnen beïnvloeden. Om de problemen met het verzamelen van dunne darm monsters te omzeilen, zijn er mensen met een ileostoma benaderd. De dikke darm van deze mensen is operatief verwijderd als gevolg van darmontsteking of kanker, waarna het einde van de dunne darm is bevestigd aan een stoma in de buikwand. De inhoud van de dunne darm komt terecht in een stoma zakje. Dit 'ileostoma effluent' wordt vervolgens gebruikt als onderzoeksmateriaal.

De samenstelling van bacteriën in het ileostoma effluent van het stoma zakje kan bepaald worden met kweekmethoden, maar tegenwoordig worden hiervoor veelal moleculaire technieken gebruikt. Hierbij wordt er gebruik gemaakt van genetisch materiaal (DNA) geïsoleerd uit de monsters. DNA is voor te stellen als twee om elkaar heen gedraaide ketens die elk bestaan uit een opeenvolging van 4 verschillende moleculen (nucleotiden). Zoals met letters van een alfabet woorden gevormd worden, wordt met de volgorde (sequentie) van nucleotiden genen samengesteld waarin is vastgelegd welke eiwitten de cel kan maken.

In hoofdstuk 2 zijn er allereerst twee moleculaire technieken, namelijk pyrosequencing en een fylogenetische microarray vergeleken met betrekking tot hun capaciteit om de bacteriesamenstelling in fecale en dunne darm monsters te bepalen. Beide technieken berusten op het detecteren van een specifiek bacterieel gen (16S rDNA) dat codeert voor één component van de ribosomen: de eiwitfabriekjes van de cellen. Hoewel alle bacteriën het 16S rDNA gen dragen, verschilt het van soort tot soort. In de voorbereiding van de monsters voor beide technieken wordt allereerst het 16S rDNA uit de monster vermeerderd door middel van Polymerase Chain Reaction (PCR), zodat er genoeg materiaal is voor de vervolgstappen. Hiervoor worden 'primers' gebruikt die kleven aan de uiteinden van het 16S rDNA en aangeven dat dit deel van het DNA door een enzym moet worden vermeerderd. Met pyrosequencing wordt de sequentie van het 16S rDNA bepaald, welke vervolgens wordt gebruikt voor de identificatie van bacteriën in elk monster. Het aantal specifieke sequenties is indicatief voor de bacteriesamenstelling per

monster. De fylogenetische microarray doet dit op een andere manier. Op een glazen plaatje staan duizenden stukjes DNA geprint, ook wel probes genoemd, en zijn specifiek voor het 16S rDNA van een bepaalde groep bacteriën. Ze zijn ontworpen op basis van bekende sequenties van 16S rDNA genen, waarvan we weten dat de bacteriën die deze genen dragen voorkomen in ons darmkanaal. Als het vermeerderde 16S rDNA uit de monsters over de glazen plaatjes worden verspreid blijven ze plakken aan de probes. Vervolgens kan met een nauwkeurige scanner bepaald worden welke probes gebonden zijn, wat aangeeft welke en hoeveel bacteriën er in het oorspronkelijke monster zaten. Aangezien PCR een cruciale stap is in de voorbereiding van de monsters, zijn verschillende primers getest om te bepalen hoe primers de uitkomst van de bacteriesamenstelling beïnvloeden. De analyses toonden aan dat de verschillende primers, en beide technieken, vergelijkbare uitkomsten gaven wat betreft de bacteriesamenstelling in fecale monsters en monsters verkregen uit het einde van de dunne darm. De bacteriesamenstelling in ileostoma effluent verschilde echter wel voor pyrosequencing en de fylogenetische microarray. Dit komt (hoogstwaarschijnlijk) doordat de probes op de fylogenetische microarray gebaseerd zijn op het 16S rDNA van bacteriën die voorkomen in de dikke darm (en dus niet alle bacteriën uit de dunne darm kunnen detecteren). Pyrosequencing kan nu juist deze bacteriën identificeren waarvan we van tevoren niet wisten dat ze in monsters voorkomen.

In hoofdstuk 3 hebben we daarom pyrosequencing toegepast om de bacteriesamenstelling te bepalen in ileostoma effluent monsters van vier volwassenen. Daarnaast gaan we in hoofdstuk 3 dieper in op welke bacteriën actief zijn in de dunne darm en wat ze daar doen. Dit is onderzocht op zowel ribosomaal RNA (rRNA) niveau als messenger RNA (mRNA) niveau. rRNA en mRNA worden beide afgelezen van het DNA, maar verschillen in functie. Het rRNA dient als component van ribosomen en wordt aangemaakt wanneer de bacterie meer eiwitten gaat produceren. Het mRNA wordt aangemaakt wanneer een specifiek gen in de bacterie wordt aangezet. Het mRNA wordt vervolgens door ribosomen afgelezen als instructie voor de productie van een specifiek eiwit. Net zoals bij het rDNA kan de sequentie van rRNA en het mRNA worden bepaald. Met de sequenties van het rRNA krijgen we inzicht in welke bacteriën actief zijn en met de mRNA sequenties krijgen we een idee wat voor genen zijn aangezet en wat de bacteriën dus aan het doen zijn. De analyses in hoofdstuk 3 toonden aan dat de samenstelling van de bacteriën op rDNA, rRNA en mRNA niveau vergelijkbaar was, wat aangeeft dat de dominante bacteriën in de dunne darm actief zijn. Hoewel *Streptococcus* bacteriën in elk ileostoma monster van alle personen werden aangetroffen, was er een sterk verschil in hoeveelheid van deze bacteriën per monster. Naast streptococci werd in het merendeel van de monsters *Veillonella* bacteriën gedetecteerd wat aangeeft dat ze een prominente rol spelen in de dunne darm van volwassenen. Dit geeft aan dat er een potentiële metabole interactie plaatsvindt tussen deze bacteriën, wat inhoudt dat de streptococci bijvoorbeeld een suiker kunnen gebruiken als energiebron en omzetten in een product waar *Veillonella* van kan leven.

In hoofdstuk 4 zijn *Streptococcus* en *Veillonella* bacteriën uit ileostoma effluent verder onderzocht door ze te kweken en te karakteriseren. *Streptococcus* bacteriën uit ileostoma effluent, zijn op basis van hun 16S rDNA in te delen in drie groepen, namelijk: de *S. mitis*, *S. bovis*, en de *S. salivarius* groep. Deze groepen kunnen verder ingedeeld worden in 7 stammen aan de hand van verdere typeringsmethoden. Dit geeft aan dat er meer verschillende streptococcus bacteriën in de dunne darm leven dan dat er op basis van 16S rDNA analyses aangetoond kan worden. De bacteriën behorende tot de stammen kunnen verschillende suikers gebruiken om te groeien, waaruit is af te leiden dat de hoeveelheid bacteriën per *Streptococcus* stam in de dunne darm waarschijnlijk afhangt van ons voedsel. Dit komt overeen met de observatie dat er maar 1 van de 7 stammen in een ileostoma effluent monster van een jaar later werd teruggevonden. Er werd maar 1 *Veillonella* stam gekweekt uit ileostoma effluent. Deze *Veillonella* stam en 3 van de 7 *Streptococcus* stammen werden ook verkregen uit mond monsters van dezelfde persoon, waar ook het ileostoma effluent van was afgenomen. Dit geeft aan dat de *Streptococcus* en *Veillonella* bacteriën uit de mond en de dunne darm verwant zijn, en dat ingeslikte bacteriën het eerste gedeelte van ons darm kanaal kunnen bewonen.

Het verschil in metabole capaciteit is in hoofdstuk 5 onderzocht door de genomen van bacterieisolaten behorende tot 6 *Streptococcus* stammen, te onderzoeken. Elk genoom codeerde voor verschillende enzymen, de bacteriën nodig hebben om suikers op te nemen en te gebruiken als energiebron. Deze resultaten komen goed overeen met observaties uit hoofdstuk 4, welke laten zien dat de verschillende streptococcus stammen kunnen groeien op verschillende suikers.

In hoofdstuk 6 proberen we een beter inzicht te krijgen hoe bacteriën van de verschillende *Streptococcus* stammen ons immuunsysteem konden beïnvloeden, en werden dendritische cellen blootgesteld aan de *Streptococcus* stammen. Dendritische cellen zijn normaal gezien gelokaliseerd onder de cellen van de dunne darm wand en kunnen met lange tentakels (dendrieten) door de darmwand heen prikken en kijken welke bacteriën er in de dunne darm aanwezig zijn. Als de dendritische cellen bacteriën zien die daar niet horen, kunnen ze met signaal moleculen (interleukinen) bepalen hoe het immuunsysteem wordt geactiveerd. De hoeveelheid, en welke interleukinen geproduceerd worden door de dendritische cellen is afhankelijk van aan welke *Streptococcus* stam ze werden blootgesteld. Aangezien *Streptococcus* en *Veillonella* vaak samen voorkomen in de dunne darm, werden dendritische cellen ook blootgesteld aan een combinatie van deze bacteriën. De hoeveelheid interleukinen die bij deze test geproduceerd werden, verschilt van wat voorspeld kon worden op basis van tests met alleen *Streptococcus* of *Veillonella*. Dit geeft aan dat het ontzettend moeilijk is om te bepalen hoe ons immuunsysteem reageert op de totale samenstelling van de bacteriën in de dunne darm.

De studies van de dunne darm microbiota gepresenteerd in dit proefschrift dienen als aanvulling op de kennis van de microbiota in ons darmkanaal. De verkregen data kan gebruikt worden als basis en richtlijn voor vervolgstudies die zich richten op de interactie tussen zowel ons voedsel en immuunsysteem, als de microbiota in de dunne darm.

Acknowledgements - Dankwoord

Na een uitdagend maar vooral leuk project is het dan nu echt klaar. Dit proefschrift is niet geworden zoals het nu is zonder de bijdrage van vele personen. Daarvoor wil ik graag iedereen bedanken, maar een aantal mensen in het bijzonder.

Michiel, dank je wel voor de toegewijde begeleiding als promotor en de nuttige discussies tijdens mijn promotie. Ik ben je ook dankbaar voor de steun en het vertrouwen in mij om dit project tot een goed einde te volbrengen, vooral in tijden als ik zelf het licht aan het einde van de tunnel niet meer kon zien. Dankzij je vele connecties heb ik ook in andere 'toko's' mogen kijken, met name NIZO Food Research en de Host-Microbes Interactomics (HMI) groep. Die samenwerkingen waren cruciaal voor het tot stand komen van een aantal mooie papers.

Erwin, dank je wel voor alle praktische adviezen tijdens mijn promotie, vooral in het begin bij de opzet van de PCRs, clonering, sequencing en analyse van de vele 16S rRNA genen. Deze kennis was van groot belang tijdens mijn promotie en zal goed van pas komen in mijn verdere loopbaan. Dank je wel voor alles!

Willem en Hauke, ik wil jullie beiden bedanken dat ik onderzoek heb mogen doen binnen de vakgroep en MolEco. Ook ben ik dankbaar voor jullie interesse in mijn project en de kritische feedback op een aantal papers. Ik vind het fijn dat ik kan blijven om als postdoc onderzoek te doen binnen de vakgroep.

Sebastian, dank je wel voor je inzicht met data analyses, het schrijven van scriptjes en relativerende gesprekken. Ik vind het fijn dat, ondanks je eigen drukke bezigheden, je mijn paranimf wil zijn. Ik wens je veel succes je de afronding van je eigen proefschrift.

Janneke, al vanaf het begin dat je bij microbiologie kwam ben ik onder de indruk van je zelfverzekerdheid en je vermogen om het een-en-ander goed te regelen. Dit zijn waardevolle eigenschappen wat je maken tot een bijzondere paranimf en die goed van pas komen tijdens je eigen promotie. Dank je wel voor alles!

During my PhD project I learned that good science goes hand in hand with a well-balanced social life. Great examples are congress visits (... Copenhagen), We-days as well as the many dinners, city visits (... Amsterdam), and parties. I thank everyone at microbiology for making this, and much more, possible as well as for their contribution to this thesis. A special thanks goes out to Anja, Carien, Carmen, Carolien, Caroline, Clara, Coline, Corina, Detmer, Floor, Gerben, Hans, Heissa, Ineke, Jannie, Javi, Jing(eling...), Jorn, Jueeli, Kyle, Leo, Maria, Mark, Martha, Mauricio, Milkha, Mirjana, Muriel, Naïm, Noora, Odette, Oylum, Peter, Petia, Philippe, Sven, Steven, Teresita, Tian, Susana, Vicente, Vincent, Wilma, and Wim.

Kun and Ruth I am very happy that I was your supervisor during your internships at microbiology. Thank you so much for your contribution to the chapters in this thesis.

The research in this thesis could not have been completed without the help of collaborators to which I am very grateful.

Allereerst wil ik graag de mensen van NIZO Food Research bedanken. Jos, dank je wel voor de ongeëvenaarde uitleg en uitvoering van bio-informatische analyses van pyrosequencing en genoom data. Zonder dit werk was het niet mogelijk om 3 hoofdstukken van dit boekje te voltooien. Iris en Saskia, bedankt voor al het praktische advies en alle gezelligheid. Laten we vooral deze samenwerking behouden in toekomstige projecten. Oylum, it was very nice to work with you, thank you for your valuable contribution to chapter 4.

Marjolein, het werk bij HMI stond ver van mijn eigen bedje. Ik heb veel geleerd van onze samenwerking en hoop dat de resulterende paper (hoofdstuk 6) niet de laatste zal zijn. Dank je wel voor de open gesprekken om soms even wat frustraties eruit te gooien. Jerry, thank you for your interest in my work, the introduction into the complex world of immunology, and facilitating the work at HMI. Anja, Jurgen, Linda en Nico, dank jullie wel voor de gezellige sfeer in het lab bij HMI en jullie hulp bij het uitvoeren de experimenten.

Gjalt en Hermie, ik wil jullie graag bedanken voor de waardevolle adviezen en de mogelijkheid om FISH te leren tijdens de cursus in Groningen. Marcus, dank je wel voor jou bijdrage aan hoofdstuk 4.

The work presented in this thesis was part of the Top Institute Food and Nutrition project c1001 and I thank everyone for the useful feedback during project meetings: Johan (voor het faciliteren van de procedures in het begin van het project), Eddy (bedankt voor de leiding tijdens het project, je waardevolle suggesties en input bij de papers), Greer, Ingrid, Lucie, Maria, Milkha, Michiel W, Onur, Oylum, Patrick, Roland (ook voor je kritische input op hoofdstuk 5), Sacha, Sander en Victor.

Verder wil ik ook Corine, Luc, Marion, Michiel S en Rianne bedanken voor alle ondersteuning bij administratieve zaken en de snelle behandeling bij het printen van posters en publiceren van artikelen.

Apart from the many friends that I made at microbiology I wish to thank Angie, Aimy, Oswaldo, and Quinta for the fun times and the support during my PhD. I am thankful to have such great friends.

Tenslotte wil ik mijn familie bedanken. Lieve Pa en Ma, tante Annelien en oom Jan, opa en oma, bedankt voor alle zorg en support tijdens mijn opleiding en promotie, vooral als de frustraties even te veel werden. Ik hoop dat dit boekje goed laat zien waar ik de afgelopen jaren zo druk mee bezig ben geweest. Zonder 'thuis' had ik dit allemaal niet kunnen bereiken en daar ik dan ook zeer dankbaar voor.

Bartholomeus/Tommycoccus/Tommy/Tompie... Tom

About the author

Bartholomeus van den Bogert was born on October 27th, 1984, in Delft, The Netherlands. After completing secondary school at Interconfessionele Scholengemeenschap Westland (ISW) in Naaldwijk in 2002, he continued with the study “Biology and Medical Laboratory Research” at Rotterdam University (Hogeschool Rotterdam) to obtain his BSc. The BSc thesis, titled “Characterization and dynamics of the bacterial flora in Barrett’s esophagus” was defended at the Erasmus MC, Department of Gastro-enterology and hepatology in Rotterdam under the supervision of Dr. V. Menke and Dr. J.G. Kusters. He graduated (cum laude) in 2006 with a specialization in Medical Microbiology. In the same year he started the MSc study “Molecular and Cellular Biology” at Leiden University, which he finished in 2008. His final MSc thesis was performed in the field of Molecular Microbiology at Unilever R&D in Vlaardingen, during which he was supervised by Dr. R.A. Kemperman and Dr. E.E. Vaughan.

In September 2008 he started his PhD-research at the Laboratory of Microbiology at Wageningen University under the supervision of Dr. E.G. Zoetendal and Prof. Dr. M. Kleerebezem. His research focused on community and genomic analysis of the human small intestine microbiota in a TI Food and Nutrition funded project entitled “Complex fermentations”.

Currently, he is employed as a Postdoc at the Laboratory of Microbiology.

Overview of completed training activities

Discipline specific activities

Meetings

- Gut day, Utrecht, NL 2008
- TIFN Autumn We-day, Wageningen, NL 2008
- TIFN Spring We-day, Groningen, NL 2009
- Second Next-Generation Sequencing Users Meeting, Utrecht, NL 2009
- TIFN Autumn We-day, Wageningen, NL 2009
- Gut day (**poster presentation**), Vlaardingen, NL 2009
- TIFN Food Summit 2009 (**invited speaker**), Wageningen, NL 2009
- TIFN Spring We-day, Renesse, NL 2010
- 1st Annual Conference TI Food and Nutrition, Ede, NL 2010
- ROWETT/INRA meeting-Conference 2010 (Gut Microbiology: New insights into gut microbial ecosystems; (**oral presentation**), Aberdeen, UK 2010
- Gut day (**poster presentation**), Gent, BE 2010
- TIFN Autumn We-day, Wageningen, NL 2010
- 100th anniversary NVvM meeting (**poster presentation**; 1st prize), Papendal, NL 2011
- TIFN Spring We-day, Wageningen, NL 2011
- 2nd Annual Conference TI Food and Nutrition (**poster Presentation**; 1st prize); Groenekan, NL 2011
- 10th symposium on Lactic Acid Bacteria (**poster Presentation**), Egmond aan Zee, NL 2011
- Gut day (**oral presentation**), Wageningen, NL 2011
- Mini symposium TIFN program 3 projects (**oral presentation**), Wageningen, NL 2012
- 3rd Annual Conference TI Food and Nutrition, Papendal, NL 2012
- NVvM meeting (**poster presentation**), Papendal, NL 2012
- Mini symposium on the ecology and pathogenomics of *Streptococcus suis*; (**invited speaker**) Wageningen, NL 2012
- ISME14, 14th International ISME Symposium on Microbial Ecology (**poster presentation**), Copenhagen, DK 2012

Courses

- Clinical anaerobe bacteriology, Groningen, NL 2009
- Light in the intestinal tract Tunnel (poster presentation), Helsinki, FI 2009
- Extended basic ARB workshop, Wageningen, NL 2009
- Next generation sequencing (NGS) data analysis, Leiden, NL 2009
- Systems Biology: "Statistical analysis of -omics data", Wageningen, NL 2010
- Functional Metagenomics of the Intestinal Tract and Food-Related Microbes (poster presentation), Helsinki, FI 2011

General courses

- VLAG PhD week, Bilthoven, NL 2009
- PhD Competence assessment, Wageningen, NL 2009
- TIFN presentation course, Wageningen, NL 2009
- Project and Time Management, Wageningen, NL 2009
- Teaching and Supervision Thesis Students, Wageningen, NL 2009
- Techniques for writing and presenting scientific papers, Wageningen, NL 2011
- Dealing with the media and the general public, Wageningen, NL 2011

Optional activities

- Preparing PhD proposal 2008
- Laboratory of Microbiology PhD trip, Northeast Coast, USA 2009
- Laboratory of Microbiology PhD/postdoc meetings 2008-2012
- Molecular ecology group meetings (weekly) 2008-2012
- TIFN colloquia and meetings 2008-2012

The studies presented in this thesis were performed within the framework of TI Food and Nutrition.

Thesis layout: Tom van den Bogert
Cover design: Tom van den Bogert
Printed by: Gildeprint Drukkerijen - The Netherlands