Stellingen

1. Streptococcen zijn overvloedig aanwezig in ileostoma effluent en bovendien metabool actief.

   Dit proefschrift (hoofstuk 3 en 6)

2. De microbiota van ileostoma effluent fluctueert rondom een stabiele kerngroep.

   Dit proefschrift (hoofdstuk 3 en 8)

3. De huidige bioinformatica tools zijn ontoereikend voor de analyse van volgende generatie sequentie data.

   Snyder et al. *(2009) Next-generation sequencing-the pearls and perils of charting the great microbial unknown. Microb. Ecol. 57:1-3*
   Pop et al. *(2008) Bioinformatics challenges of next generation sequencing. Trends Genet. 24:142-149*

4. Het publiceren van een nieuwe hypothese en het falsificeren ervan zouden een vergelijkbare impact op moeten leveren.

   Keppler et al. *(2006) Methane emissions from terrestrial plants under aerobic conditions. Nature 439:187-191*
   Dueck et al. *(2007) No evidence for substantial aerobic methane emission by terrestrial plants: a 13C-labelling approach. New Phytol. 175:29-35*

5. Hoe hoger de opleiding, hoe lager het didactische niveau van de docent.

6. De Nederlandse infrastructuur rechtvaardigt een verbod op SUVs.

7. Tukkers praten platter dan mensen uit de Achterhoek.

8. Omwille van de huiselijke vrede zouden de makers van gezelschapsspellen de spelregels zorgvuldiger moeten formuleren.

Stellingen behorende bij het proefschrift:
"Analysis of Diversity and Function of the Human Small Intestinal Microbiota"

Carien Booijink
Wageningen, 14 april 2009

Propositions

1. Streptococci are abundantly present and moreover metabolically active in ileostomy effluent.

   This thesis (Chapters 3 and 6)

2. The microbiota of ileostomy effluent fluctuates around a stable core.

   This thesis (Chapters 3 and 8)

3. The current bioinformatics tools are incapable for analysis of next-generation sequencing results.

   Snyder et al. *(2009) Next-generation sequencing-the pearls and perils of charting the great microbial unknown. Microb. Ecol. 57:1-3*
   Pop et al. *(2008) Bioinformatics challenges of next generation sequencing. Trends Genet. 24:142-149*

4. The publication of a novel hypothesis and the subsequent falsification of it should result in an equal impact factor.

   Keppler et al. *(2006) Methane emissions from terrestrial plants under aerobic conditions. Nature 439:187-191*
   Dueck et al. *(2007) No evidence for substantial aerobic methane emission by terrestrial plants: a 13C-labelling approach. New Phytol. 175:29-35*

5. The higher the education, the lower the didactic level of the teacher.

6. The Dutch infrastructure justifies a ban on SUVs.

7. People from "Twente" speak in a flatter tongue compared to people from "de Achterhoek".

8. For the sake of cosiness at home, the producers of parlour games should be more precise in formulating the rules.

Propositions accompanying the thesis:
"Analysis of Diversity and Function of the Human Small Intestinal Microbiota"

Carien Booijink
Wageningen, April 14th 2009

Analysis of Diversity and Function

of the

Human Small Intestinal Microbiota

Promotoren	**Prof. dr. W.M. de Vos** Hoogleraar Microbiologie Wageningen Universiteit
	Prof. dr. M. Kleerebezem Hoogleraar Bacterial Metagenomics Wageningen Universiteit
Co-promotor	**Dr. E.G. Zoetendal** Wetenschappelijk onderzoeker Laboratorium voor Microbiologie Wageningen Universiteit
Promotiecommissie	**Prof. dr. L.M.A. Akkermans** Universiteit Utrecht
	Prof. dr. J.M. Wells Wageningen Universiteit
	Dr. M. Leclerc INRA, Jouy-en-Josas, France
	Dr. A. Nauta Friesland-Coberco, Deventer

Dit onderzoek is uitgevoerd binnen de onderzoeksschool VLAG.

Analysis of Diversity and Function

of the

Human Small Intestinal Microbiota

Carien C. G. M. Booijink

Proefschrift
Ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. M. J. Kropff,
in het openbaar te verdedigen
op dinsdag 14 april 2009
des namiddags te vier uur in de Aula

Booijink, C.C.G.M. Analysis of Diversity and Function
of the Human Small Intestinal Microbiota

In Dutch
Analyse van de diversiteit en functie van de microbiota in de dunne darm van de mens

PhD Thesis Wageningen University, Wageningen, the Netherlands (2009)
144 p. - with summary in Dutch

ISBN **978-90-8585-362-6**

TABLE OF CONTENTS

	Abstract	1
CHAPTER 1	General introduction	3
CHAPTER 2	Microbial communities in the human small intestine-coupling diversity to metagenomics	11
CHAPTER 3	Population dynamics and diversity of the luminal microbiota in human ileum samples	23
CHAPTER 4	RNA isolation and mRNA enrichment from samples of the human gastrointestinal tract	39
CHAPTER 5	Gene expression profiling of the human GI tract microbiota	51
CHAPTER 6	Metatranscriptomics approach to study the functionality of the human ileostomy effluent microbiota	65
CHAPTER 7	Novel mucin degradation genes identified in a human ileal metagenome library	77
CHAPTER 8	Summary, concluding remarks & future perspectives	89
	Literature cited	100
	Supplementary & colour figures	112
	Co-author affiliations	129
	Dutch summary - Nederlandse samenvatting	130
	Acknowledgements - Dankwoord	134
	VLAG PhD education form	136
	Curriculum Vitae	137
	Publications	138

ABSTRACT

The gastrointestinal (GI) tract is the main site where the conversion and absorption of food components takes place in humans. As the small intestine is the first site of interaction between the microbiota and ingested food, knowledge about the microbial composition as well as functionality is essential for a complete understanding of the symbiotic interactions and to the potential modulation of metabolically important groups. Subjects carrying an ileostomy were chosen as model system and ileostomy effluent samples were collected over time. The diversity as well as activity of the inhabiting microbiota was analysed in ileostomy effluent samples of five healthy individuals, collected in the morning and afternoon over a period of 28 days. This revealed that the diversity of the ileostomy effluent microbiota was different from that in the faeces, mainly concerning the lower complexity and stability over time. In terms of composition the relative abundance of species belonging to the genera *Streptococcus* and *Veillonella* was higher, whereas a lower relative abundance of species related to the *Ruminococcus obeum*, *R. gnavus* and *Bacteroides plebeius*-like organisms was observed in ileostomy effluent samples. Marked differences in microbiota composition between the five subjects with an ileostomy were found, indicative for a highly personal ileal microbiota profile. Differences in microbiota composition profiles were observed over time, even visible within one day, although the overall fluctuations were around a relatively large stable core group, consisting of species belonging to three streptococci-related groups (*S. bovis*, *S. intermedius* and *S. smitis*), *Clostridium* cluster I, *Enterococcus*, *Veillonella* and *Oxalobacter formigenes*.

For gene expression analysis of the small intestinal microbiota, procedures for extraction of enriched mRNA as well as transcript profiling suitable for complex microbial populations are essential. First, an RNA extraction procedure based on mechanical disruption of the cells followed by mRNA enrichment based on selective removal of rRNA molecules was optimised for GI tract samples. Enriched mRNA from faecal origin was subsequently used as starting material for RNA fingerprint analysis as proof of principle. This revealed that especially transcripts of the metabolism cluster were abundantly found, underlining that even at the very end of the intestinal tract the microbes are still active. The same transcript analysis was subsequently applied to enriched mRNA extracted from ileostomy effluent samples. The gene expression profile of ileostomy effluent microbiota proved to be especially enriched in carbohydrate uptake and degradation. In addition, 70% of the retrieved transcripts showed highest similarity with genes of the genus *Streptococcus*. The majority of retrieved transcript sequences seem to be commonly present in the microbiota of the small intestine of humans, since almost all analysed transcript sequences detected in one person, could also be detected in DNA samples of ileostomy effluent of at least one other ileostomy subject.

Metagenomics analysis offer another means to investigate potential functions exerted by complex microbial ecosystem. A metagenomic library constructed with DNA from the ileal mucosa microbiota was used to analyse the mucin-degrading capacity of the human small intestinal microbiota. Sequence analysis of the inserts of one metagenomic clone that was able to grow on mucin-derived substrates revealed two putative genes with sequence similarity to a glycosyltransferase and an inner-membrane sugar transporter, probably originating from

Bacteroides. In addition, the insert of a fosmid that grew on mucin encoded an uncharacterised protein from *Enterococcus faecalis* which possibly functions as a soluble lytic murein transglycosylase. The results displayed here showed that the functional screening of a metagenome can reveal functional genes involved in mucin utilisation, even from Gram-positive species.

Overall, the data presented in this thesis indicated that the genus *Streptococcus* is not only numerically abundant, but also predominates randomly generated metabolic activity profiles of the microbial ecosystem of the ileostomy effluent microbiota. Predominant functions exerted were related to metabolism, especially carbohydrate metabolism and transport. The fast transit of the ileal contents appears to generate an environment in which the capacity to rapidly metabolise the available carbohydrates is an important selective advantage.

CHAPTER 1

General introduction

INTRODUCTION

The gastrointestinal (GI) tract is the most densely populated collection of organs in the human body. The community populating the GI tract is termed the microbiota and consists of 10^{14} microbes from which the fast majority are the bacteria (but also archaea and fungi can be found), collectively termed the microbiota. This chapter will briefly summarise the so-far gathered knowledge related to the human gastrointestinal tract microbiota with a focus on the terminal part of the small intestine, which is called the ileum. It is at this location that food and microbes have the first intimate interactions with each other and the host, as the intestinal surface area is here the largest of the GI tract. However, the number of studies that investigated the *in vivo* ileal microbiota is very limited as the ileum is particularly difficult to reach and sample in a non-invasive way due to its location in the healthy human body. The work described in this thesis and outlined at the end of this chapter aimed at improving our understanding of the diversity and the functional role played by the microbes in the human small intestine.

GI tract architecture and the dynamics of its inhabitants

The human gastrointestinal (GI) tract can be regarded as a complex and dynamic ecosystem, which consists of several distinct anatomical regions. Each region has its own physico-chemical characteristics, including variation in transit rate of the luminal content, local pH, and availability of host as well as diet-derived components. These changing conditions along the intestinal tube exert a selective pressure on the microbiota, resulting in a strong increase in microbial density as well as diversity going from stomach to colon [115]. In addition, within each anatomical region of the intestine different microhabitats can be distinguished; the intestinal lumen, the mucus layer, the intestinal crypts filled with mucus, and the surface of mucosal epithelial cells [14, 100]. For example, it was demonstrated that the predominant bacterial species associated with the colonic mucosa differed significantly from the predominant faecal community in several healthy adults [236] and subjects with inflammatory bowel disease [101].

The different parts of the human GI tract are generally accepted as being sterile before birth [47, 115], although a recent study detected DNA of bifidobacteria or lactobacilli in all studied placentas obtained by caesarean section [178], while another study indicated that the placentas of women who delivered by elective caesarean were colonised [147]. However, these cases can also be considered as intra-amniotic infections as the presence and quantity of microbes or their DNA in amniotic fluid were found to be positively correlated with elevated levels of interleukin-6 (IL-6) and preterm labour [39]. Upon initial colonisation by a few different species, including bifidobacteria and lactobacilli, the microbiota of infants succeeds towards a complex and highly diverse community similar to that of adults [115]. Based on characterisation of faecal samples from a wide variety of adults, this climax community in the GI tract shows a relatively stable and host-specific community in which host-microbe and microbe-microbe interactions [76, 229], as well as dietary influences [37] play a crucial role.

It has been proposed that the stability of the microbiota of healthy adults is formed by a significant fraction of the microbes that are continuously present over time, combined with transient visitors [155]. In addition, it was hypothesised that besides this individual-specific core

of microbes a common core of microbes formed by stable colonisers is shared by all people [41, 234] and hence, the microbiota maintains its capacity to carry out a basic set of biochemical reactions, including fermentation, and synthesis of vitamins that are crucial for the host [76]. However, as these results were obtained from colonic samples, which reflect the end situation of the colon, extrapolation of the common core hypothesis to other parts of the GI tract, such as the small intestine is as yet unfounded, since studies describing the population dynamics of the microbiota in these parts are lacking. The few reported studies describing the microbiota in the small intestine made use of single biopsy samples of the luminal or mucosal ileum [216, 217], or ileum samples collected from sudden death victims [72]. These studies indicated that species belonging to the *Clostridium* cluster XI and XIVa, and the phylum Bacteroidetes were most numerously present in the mucosa, whereas Proteobacteria and streptococci dominated the lumen of ileum (more extensively described in Chapter two). These studies provided usefull information about the diversity of the ileal microbiota. However, due to the invasive procedures these individuals could only be sampled at a single time point and hence, the population dynamics of the ileal microbiota cannot not be studied in this way.

Functionality of model microorganisms

Current knowledge about the functionality of the microbiota is derived from transcriptome, proteome or metabolome analysis of either one (or a group of) model microorganism(s), or from the coherent study of the complete microbiota. From a reductionist point of view, groups or even single species of the GI tract microbiota are studied to get insight in the microbial functionality. When these species belong to the predominant members of a community within a particular niche of the human GI tract, they can be used as model organisms for understanding the functioning of the microbiome as a whole. For several inhabitants commonly found in the human colon, this approach already led to essential knowledge regarding key roles fulfilled by the microbiota. For example, one abundant coloniser of the colon, *Akkermansia muciniphila* was found to be a specialist in mucin degradation and could be retrieved from mucosal biopsy samples in higher numbers compared to faeces [34, 35]. Moreover, this species was found in mucosa biopsies from terminal ileum, ascending colon and rectum, comprising 5%, 6% and 9%, respectively, of the total microbiota based on random clone libraries [216]. With mucus concentration about ten-fold higher in the small intestine relative to the large intestine, species capable of mucin degradation are likely to be found in higher relative number in the ileum compared to the colon. An overview of the range of enzymes encoded by species capable of mucin degradation is given in Table 1-1.

Another abundant model species of the human GI tract, *Bacteroides thetaiotaomicron*, has been studied extensively in germ-free mouse model systems to investigate its interaction with the host at the molecular level [77, 192, 223]. This Gram-negative anaerobe is estimated to make up about 6% of the colonic microbiota [42] and was also detected in 16S rRNA gene clone libraries of ileal and colonic mucosa samples [216, 217]. At the mucosa it might make use of the mucin as energy source, since this bacterium encodes several enzymes involved in mucin degradation (Table 1-1). In addition, this microorganism is very well equipped for the acquisition and hydrolysis of otherwise indigestible dietary polysaccharides [223]. Partly due to this extensive glycobiome,

B. thetaiotaomicron exerts an influence on several aspects of host physiology such as mucosal barrier reinforcement, immune system modulation, and the metabolism of nutrients as was recently reviewed [228].

TABLE 1-1 References describing the mucin-degrading enzymes of GI tract bacteria (adapted from Derrien [34]). Abbreviations used for the GI tract bacteria are: Am = *Akkermansia muninciphila*; Bf = *Bacteroides fragilis*; Bt = *Bacteroides thetaiotaomicron*; Bv = *Bacteroides vulgatus*; B = *Bifidobacterium* sp.; Bb = *Bifidobacterium bifidum*; Bi = *Bifidobacterium infantis*; Bl = *Bifidobacterium longum*; Cc = *Clostridium cocleatum*; Cp = *Clostridium perfringens*; Cs = *Clostridum septicum*; P = *Prevotella* sp. RS2; Rt = *Ruminococcus torques*; S = *Streptomyces* sp.; Vc = *Vibrio cholerae*.

		Microorganisms													
Enzymes		**Am**	**Bf**	**Bt**	**Bv**	**Bb**	**Bi**	**Bl**	**Cc**	**Cp**	**Cs**	**P**	**Rt**	**S**	**Vc**
Glycosidases	α-N(Ac)-galactosaminidase	[34]	[110]	[223]	[169]	[88] [168]		[59]	[20]	[104]			[78] [80]	[84]	
	β-N(Ac)-galactosaminidase	[34]			[169]	[78]							[78]		
	β-N(Ac)-glucosaminidase	[34]	[110]	[223]	[169]	[78]	[78]	[78]	[20]		[113]		[78]		
	α-fucosidase	[34]	[110]	[223]	[169]	[88] [168]								[67] [174]	
	β-fucosidase	[34]													
	α-galactosidase	[34]			[169]										
	β-galactosidase	[34]	[110]	[223]	[169]	[78]	[78]		[20]		[113]		[78]		
	β- glucosidase	[34]							[20]						
	α-mannosidase	[34]													
	endo-β-N(Ac)-hexosaminidase														[196]
	galactosyl-N(Ac)-hexosamine phosphorylase					[33]									
Protease			[112]												[196]
Sialidase (neuraminidase)			[13]	[223]	[169]	[78]	[78]	[29]	[20]	[104]	[113]		[29] [78]		[196]
Sulphatase		[34]		[203] [223]							[113]	[221]	[29]		

Microorganisms that exhibit specific functions related to beneficial effects on the host, the so-called probiotics, have been studied quite extensively regarding their activity in the GI tract. However, molecular studies of the *in vivo* activities of probiotic strains in the human host are very limited. A recent transcriptomics approach demonstrated that *Lactobacillus plantarum* expressed a variety of genes involved in its adaptation to the intestinal environment, reflected by a high degree of metabolic activity [31]. For example, in the human ileum, the adaptive response of *L. plantarum* was related to carbohydrate transport and metabolism as well as energy metabolism.

An advantage in studying GI tract bacteria like *A. muciniphila, L. plantarum* and *B. thetaiotaomicron* is that their genome sequences are available. Furthermore, in case of *L. plantarum,* this enables the convenient construction of gene deletion and over-expression mutants, to subsequently study the effect of a single bacterial gene on the host [23-25].

Functionality of the complete microbiota

Although the reductionist approaches have revolutionised the perception of the impact of microbes on the metabolism of the host, the properties of the community as a whole cannot be explained by the isolated features of its component parts [160]. Therefore, the microbiota should be studied by holistic approaches, to gain insight into all exerted functions and activities. For this purpose, the so-called meta-"omics" approaches like metatranscriptomics, metaproteomics and metabonomics are very suitable as they focus on the profiling of microbial activity [234] (Fig. 1-1). In addition, metagenomics provides researchers with genome catalogues that are representative for the microbial diversity and the functional blueprint of the ecosystem under study [68]. Following construction of a metagenomic library, screening can be realised by sequence- or function-driven approaches, or a combination of both [49]. Such analyses make predictions of potential functions possible and have been applied to several diverse microbial communities besides the human GI tract (as recently reviewed [212]). This already resulted in retrieval and characterisation of many metagenome-derived enzymes, as recently reviewed [195], including enzymes originating from the human GI tract microbiota [18]. Additionally, high-throughput sequencing projects of GI tract metagenomes and subsequent comparative analysis gained a lot of information about the microbial genome diversity as well as potential functionality. For example, comparative sequence analysis of faecal metagenomes of 13 healthy children and adults has revealed overrepresentation of functional genes within the carbohydrate transport and metabolism cluster, and underrepresentation of the lipid transport and metabolism cluster compared to a reference database [95]. It is expected that further application of the next generation sequencing approaches will provide an explosion in the metagenome sequence information of GI tract microbes [191, 234]. However, metagenomics does not reveal whether the predicted genes are originating from viable cells, nor if they are expressed, and if so under what conditions and to what extent. By directly assessing transcripts, proteins, or metabolites these limitations can be overcome because the genes must at least have been transcribed [211].

The metatranscriptome of microbial communities can so far be assessed in a limited number of ways. During the last decade, microarrays have progressed to a robust and standardised method for transcriptome profiling for single, free-living microbes. Whereas classical DNA arrays target genes of the single bacterial species under study, functional gene arrays (FGA) have been developed to investigate genes that encode key enzymes in metabolic routes in a bacterial population [106, 199, 222]. Unfortunately, the application of FGAs is limited to communities for which the microbial diversity is largely known [214]. Furthermore, application of such a mixed genomes platform to a complex ecosystem such as the human GI tract has the drawback that sequence variation in closely related genes causes artefacts in the quantitative interpretation of the gene specific signals obtained after array analysis. Next generation sequence approaches applied to cDNAs circumvent the limitations related to hybridisation-based

techniques and can provide a wealth of qualitative and quantitative information about the microbial population under study [6]. So far, sequencing-based transcription profiling has not yet been applied to the human microbiota. However, with the rapid development of new sequencing technologies a higher throughput can be reached [6] enabling application to the diverse range of transcript sequences likely to be found in the microbiota. In contrast, studies aimed at investigating the proteins expressed by the GI tract microbiota have already been performed. In a pioneering study with babies, the faecal proteome was found to show temporal variation and subsequent mass spectrometry of specifc proteins allowed for the identification of a key enzyme in bifidobacterial metabolism [93]. Recently, shotgun metaproteomics was applied on the extracted proteins from faecal samples and revealed that more than half of the detected proteins were involved in translation, carbohydrate metabolism or energy production [211].

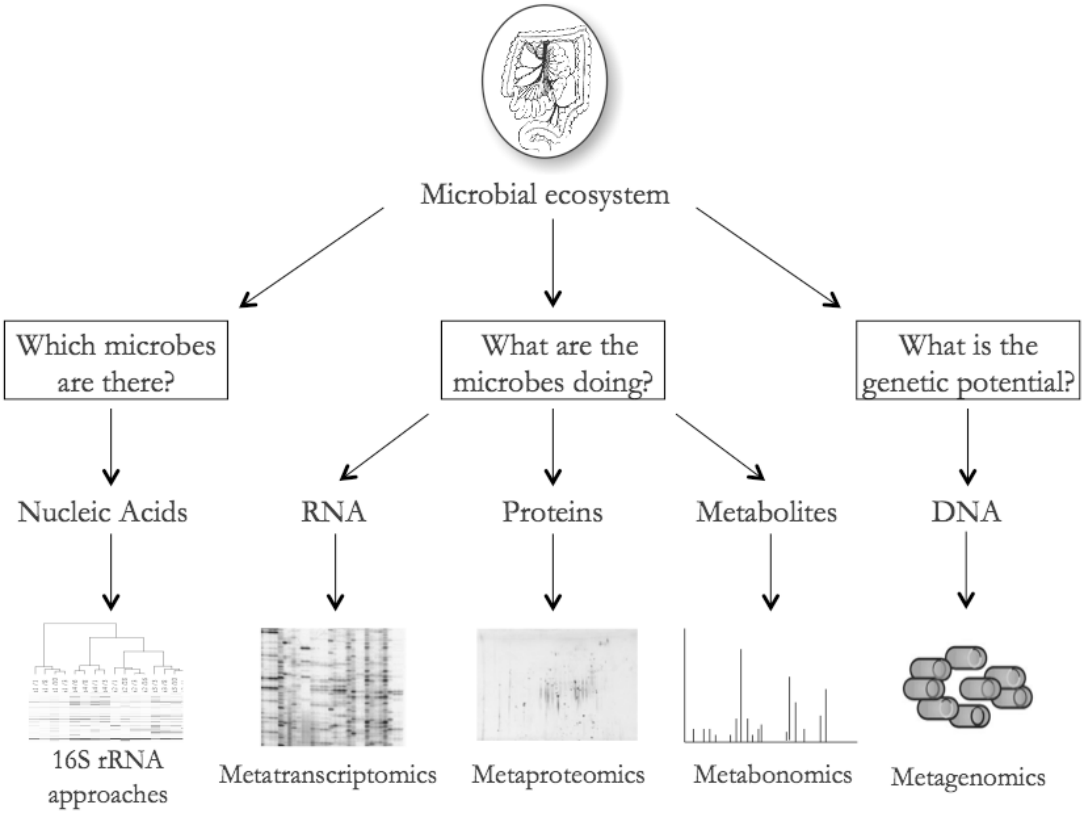


FIGURE 1-1 Overview of the meta-"omics" approaches that can be employed to answer key questions regarding the GI tract microbiota (modified after Zoetendal *et al.* [234]).

Overall, the few meta-"omics" approaches that were employed so far on the human GI tract microbiota have provided us insight into its genetic potential and some activities expressed by the microbiota. However, all these studies focussed on the distal part of the human GI tract, mainly colonic or faecal samples. The research described in this thesis anticipated that these meta"-omics" approaches will contribute to broaden the knowledge about the GI tract diversity, stability and functionality of the small intestine. The microbiota of the human small intestine was chosen as study target for several reasons. First of all, the composition, diversity and functionality

of the small intestinal microbiota is largely unexplored. Secondly, as the small intestine is basically the first location where host, microbes and food components have close interaction, this is predicted to be the most suitable site for dietary modulation of the microbiota in health and disease. Considering the awareness that the human microbiota plays an important role in health and diseases (such as inflammation [128], allergies [109], and obesity [204]), a better understanding of the small intestinal microbial ecology is essential to manage and maintain our health.

Outline of this thesis

The work described here aims to provide an increased insight into the diversity, population dynamics and functionality of the human small intestinal microbiota. To explore the luminal fraction of the small intestinal microbiota, ileostomy subjects were used as a model system and the microbiota composition, as well as the metatranscriptome was investigated at different time points. In addition, the ileal mucosa microbiota was examined with a functional metagenomics approach. The specific objectives that are addressed in each of the chapters are listed below.

Chapter 2 provides an overview of our knowledge concerning the description of the microbial diversity of the human small intestinal microbiota, including the few molecular studies that were performed so far to investigate the microbial composition of the human small intestine. The second part describes how metagenomics approaches can be applied to allow the functional description of complex microbial communities, such as the human small intestinal microbiota. **Chapter 3** describes the diversity of the ileal luminal microbiota of seven individuals with an ileostomy of which two were medicated, and one individual with an ileorectal pouch. Molecular analysis of samples collected at different time points showed that the microbial community is less diverse and less stable compared to the faecal microbiota. Studies investigating the functionality of the human microbiota at the metatranscriptome level have not been performed so far. In order to make analysis of the metatranscriptome of the human GI tract samples possible, a RNA isolation procedure followed by enrichment of mRNA was optimised as described in **Chapter 4**. This RNA isolation procedure was subsequently applied in the following two chapters. In **Chapter 5** the transcription profile of the faecal microbiota of two healthy adults was investigated by cDNA-AFLP as proof of principle. Subsequent sequence analysis revealed a wide range of transcripts and showed that this metatranscriptomic profiling method was applicable to human GI tract samples. In **Chapter 6** the gene expression of the ileostomy effluent microbiota was investigated to increase the knowledge regarding core functions exerted by the microbiota. To gain insight into the genetic potential of the microbiota residing in the ileal mucosa, a metagenomic library was constructed and screened for its mucin degradation capacity as described in **Chapter 7**. Based on enzyme activity screens two fosmids were selected for full-insert sequence analysis and one as yet unknown pathway for mucin degradation was detected. Finally, the general discussion (**Chapter 8**) summarises and discusses the contribution of this work to the current knowledge and understanding of the human small intestinal microbiota. In addition, an overview of further developments and future directions is provided.

CHAPTER 2

Microbial communities in the human small intestine – coupling diversity to metagenomics

Carien C.G.M. Booijink, Erwin G. Zoetendal,
Michiel Kleerebezem & Willem M. de Vos

*Reproduced from Future Microbiol. (2007) **2** (3), 285-295 with permission of Future Medicine Ltd.*

ABSTRACT

The gastrointestinal tract is the main site where the conversion and absorption of food components takes place. The host-derived physiological processes as well as the residing microorganisms, especially in the small intestine, are contributing to this nutrient supply. To circumvent sampling problems of the small intestine, several model systems have been developed to study the small-intestinal microbial diversity and functionality. In addition, metagenomics offers novel possibilities to gain insight into the genetic potential and functional properties of these microbial communities. Here, an overview is presented of the insights into the diversity and functionality of the microorganisms in the human GI tract, with a focus on the small intestine.

INTRODUCTION

The human gastrointestinal tract (GI tract) is an organ that is of critical importance in the processes involved in food conversion and absorption, and hence essential for supplying the body with the appropriate energy and essential nutrients. In addition, it plays a prominent role in defending the body against ingested and intruding microorganisms by the low pH in the stomach and the coverage of the surfaces of the tract with a mucus layer. This host defence as well as food digestion is further aided by myriads of specialised residential microbes that colonise the surfaces as well as the lumen of the human GI tract. These commensals, together with the food-ingested, or transient microorganisms, are called microbiota.

The approximately seven meter long GI tract can be divided in several anatomically distinct regions, starting from the mouth, via the oesophagus to the stomach, proceeding to the small intestine (divided into duodenum, jejunum and ileum, respectively) and the large intestine (divided into caecum, colon (with an ascending, transverse, descending and sigmoid part) and rectum, respectively) and ending at the anus (Fig. 2-1). Classically, the microbiota within the different regions was studied by culture-dependent techniques, leading to underestimated and incomplete views of the microbial diversity in the GI tract. With the development of ribosomal RNA (rRNA)-based and other culture-independent approaches, the knowledge of the total microbial diversity within the human gastrointestinal tract has increased considerably [235]. This increase in knowledge holds especially true for the microbiota residing in the lower regions of the large intestine, since sampling from the terminal colon is relatively easy. In contrast, very little is known about the microbial communities residing in the lower parts of the small intestine due to sampling difficulties caused by the inaccessibility of the ileum and the need to process samples anaerobically. Since by far the largest part of food digestion and absorption takes place in the small intestine it is essential to overcome sampling difficulties in order to obtain a full understanding of the processes and interactions that go on in this part of the gastrointestinal tract. Also the fact that this is after the stomach, the first region of the intestine where food and commensal microbes meet, a further understanding is necessary to possibly influence or even modulate the microbiota by dietary alterations. In addition, the microbiota composition of the small intestine might influence events in the large intestine and will play an even more important role in individuals that had part or all of their large intestine removed.

Physiology of the small intestine

The interactions between the microbiota and the host cells are largely influenced by several physiological factors. These can be divided into host factors such as transit time [187], immune response [156, 193] and secretory products (gastric acid [145], bile [92], pancreatic enzymes (amylase, trypsin) [219] and hormones (for instance insulin, glucagon) [220]), and environmental factors such as pH [46, 133] and dietary components [70]). Transit of the liquid contents in the small intestine is an understudied area [21], but with estimated transit times between 30 minutes and 3,5 hours [81] transit is faster than further down the GI tract, reducing the growth of microorganisms in the lumen of the small intestine [51, 179]. This limited growth is also caused by the fact that of the dietary carbohydrates passing through the GI tract, only the easy digestible

components can be taken up by the host and the microbiota. Here, the host-derived mucus forms an important carbon-source for the ileal microbiota, creating habitats for mucosal populations that might be important for colonising luminal material [36]. Towards the terminal part of the small intestine, transit-rate reduces and the number of microbial cells increases rapidly, while the proportion of anaerobic species increases as well [72].

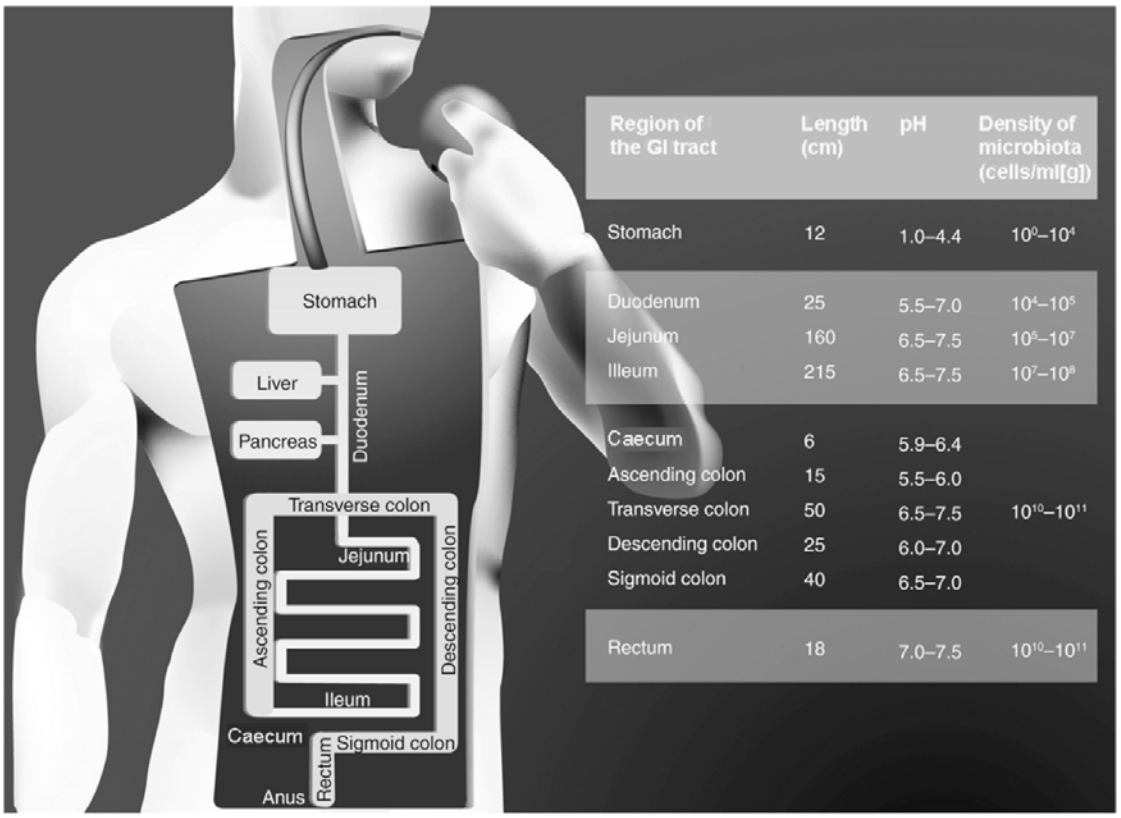


FIGURE 2-1 Schematic representation of the anatomy of the human gastrointestinal tract with the average length, pH and density of microbiota per segment.

Secretory components and enzymes formed by the human body aid in digestion of specific dietary components and are mainly secreted in the small intestine. For instance, bile acids, synthesised by the liver from cholesterol, are secreted into the small intestine to facilitate a more effective fat degradation by gastric and pancreatic lipases [123]. Different types of nutrient dose-responsive mechanisms regulate the release of enzymes in the intestine but also the passage through the intestine. For instance, gastric emptying is regulated by negative feedback control by the small intestine, which in its turn is modulated by the presence of specific food components in the lumen of the small intestine [201]. Moreover, food exposure, like chewing, already stimulates salivary flow [124] and the secretion of digestive enzymes by the pancreas [142, 176]. Also gastric lipase synthesis and activity are induced upon oral fat exposure [125].

One of the environmental factors that has been studied most is the pH change during food transit through the GI tract. Measurements of gastrointestinal luminal pH data in normal healthy

volunteers showed pH's ranging from 5.5 to 7.0 in the proximal small intestine (Fig. 2-1). The pH rises gradually to 6.5-7.5 in the distal ileum [46] caused by the buffering of the acidic gastric contents by alkaline pancreatic secretions in the proximal ileum. While human physiology is well equipped to absorb simple sugars (i.e. glucose and galactose) and to release monosaccharides from several disaccharides (e.g. sucrose, lactose and maltose) [48], its capacity to utilise other, more complex polysaccharides is very limited [76]. This material includes polysaccharides from components of plant cell walls (including cellulose, xylan, and pectin) as well as undigested starch [43]. The availability of these polysaccharides for the colonic microbiota in a normal Western diet varies between 16-58 g/day [17]. Due to the fermentation of these undigested dietary carbohydrates to short chain fatty acids by the colonic microbiota [111], the pH drops in the caecum to increase gradually in the descending colon and rectum up to 7.5 [46, 144].

Model systems of the human small intestine

In vitro model systems

To circumvent the sampling problems associated with the location of the small intestine in the human body, several *in vitro* model systems have been set up to follow digestive processes and the microbial composition in time. These include the computer controlled model developed by Molly and co-workers [131], the SHIME-model (Simulator of the Human Intestinal bacteria) and the TNO Intestinal Model (TIM) [130]. These systems enable fast reproducible experiments under controlled conditions, though validation of these systems is difficult because it requires comparison with the largely inaccessible small intestine. In addition, there are important limitations including the representative character of the microbiota inoculum used to colonise the *in vitro* systems, and the absence of host cells and mucosal layer [43]. Studies using host cell lines, like Caco-2 and the mucus producing HT29-MTX, have been used as *in vitro* systems to study the adherence of probiotic strains to the colonic epithelial cells [185] or intestinal cell responses upon exposure to mucosal bacteria [60]. To date, no *in vitro* co-culture studies for the small intestine have been conducted due to absence of suitable cell lines. However, as long as validation of the predictive value of *in vitro* models for the *in vivo* situation remains unproven, these studies will not really contribute to our understanding of the host-microbe interactions.

In vivo model systems

Several animal model systems are well appreciated due to practical reasons: multiple sites of the GI tract, as well as both the attached (e.g. mucosal) and luminal microbiota can be sampled and compared. The best *in vivo* animal model system for the human GI tract in terms of physiology, anatomy, immunology and microbial composition is the pig [103, 170]. At present though, the translation from the animal to the human situation remains difficult and not very well validated. Once this problem is bridged, the combination of germ-free technology (gnotobiotics) and genetically defined animal models might create future opportunities to define microbial ecology within the different regions of the small intestine, under rigorously controlled environmental conditions [76].

Gastroenterologists have explored various methods to sample the different regions of the inaccessible small intestine. The development of advanced endoscopic tools has made sampling

of the caecum via nasally administered intestinal tubes possible [120], offering possibilities to sample both luminal and mucosal sites. Alternatively, colonoscopy allows the collection of samples from the ileum in a relatively easy, but quite invasive manner [132]. However, using this approach sample collection is restricted to the distal part of the ileum. Biopsy sampling during a (preventive) surgical operation provides another means to collect samples from the mucosal part of the small intestine [101, 184].

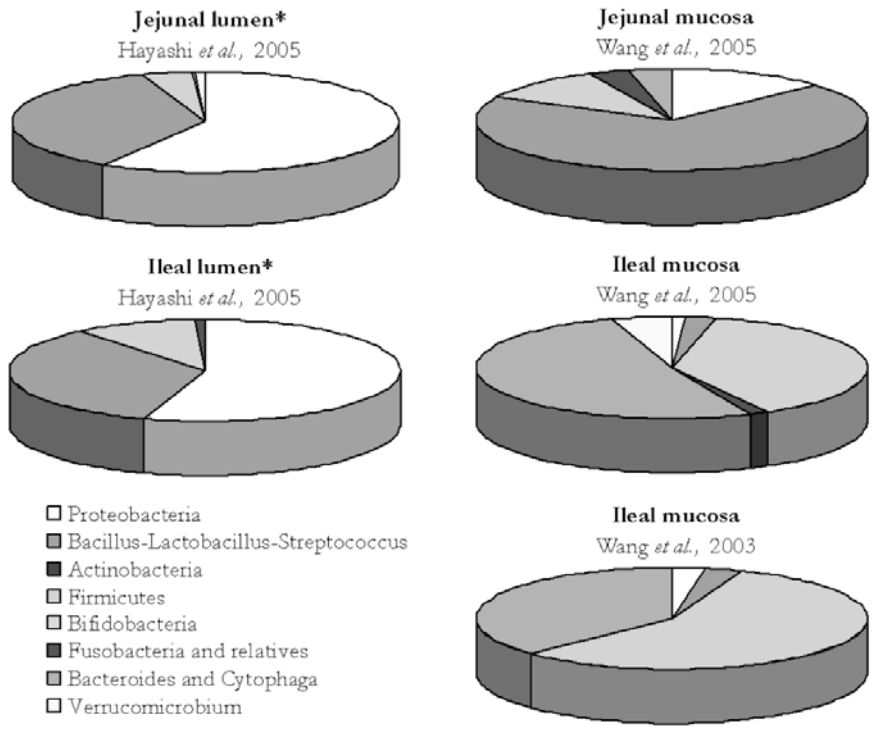
Several researchers have circumvented the inaccessibility problem of the small intestine by using *in vivo* model systems. One that has been used in several studies is the ileostomy patient [89, 140, 175]: a person that underwent surgical removal of his or her colon due to a disease occurring in this part of the GI tract (e.g. Crohn's disease, ulcerative colitis, cancer or polyposis). During this operation, the terminal part of the ileum was connected to either the anus (called ileorectal pouch) or to an ileostomy in the abdominal wall of the patient (called Brooke ileostomy) [89]. As a consequence, collection of the ileostomy effluent offers unique and non-invasive access to the luminal microbiota of the distal ileum. Moreover, sampling at different time points of a single ileostomy patient offers opportunities to investigate the stability of the microbiota as such or in relation to dietary changes or interventions [65, 74, 99, 187].

Microbial diversity

Over the past decade, the knowledge of human microbial diversity in the GI tract has been expanded by bypassing cultivation strategies with molecular techniques. This explosion of information is mainly fuelled by the use of small subunit ribosomal RNA (16S rRNA) gene sequences as culture-independent markers of microbial taxa. Different 16S rRNA-based techniques are suitable for different investigations, including identification of microbial species, quantification of microbial taxa, or broad comparisons of microbial communities [232]. Studies based on these techniques revealed that the majority of intestinal microbes belong to previously unknown (or uncultivated) lineages [42, 75, 217, 232]. In addition, it showed that the mucosa-associated bacterial community is host specific and uniformly distributed along the GI tract [101, 236], although there is also evidence that the bacteria do not come into contact with the mucosa [207]. In terms of microbial numbers the stomach and proximal small intestine contain relatively low amounts because of low pH and rapid luminal flux in this region [16]. Acid-tolerant lactobacilli and streptococci predominate in the upper small intestine. Towards the distal ileum microbial numbers as well as microbial diversity increases [70] (Fig. 2-1).

Only few studies applied molecular techniques to identify the microbial population present in regions of the small intestine (Fig. 2-2). Wang and co-workers showed that Bacteroidetes and *Clostridium* clusters XIVa and XI belonged to the dominant species present in biopsy samples from the ileal mucosa of one individual [217]. In a subsequent study, it was shown that species of the phylum Bacteroidetes and *Clostridium* were dominant in biopsies of the distal ileum whereas the *Bacillus-Lactobacillus-Streptococcus* group (mostly streptococci) and representatives of the phylum Proteobacteria were most abundant in jejunal biopsy samples [216]. In another study molecular techniques were applied to ileal and jejunal contents collected within several hours after the decease of three individuals (causes of death were aortic disorders or uterine cervical cancer). The results showed that streptococci, enterococci, lactobacilli, γ-Proteobacteria, and Bacteroidetes

were among the dominant groups present [72]. The studies described above already gave indications for the microbiota composition in the different regions of the small intestine. However, far more culture-independent studies, including on small intestinal material from a wide variety of individuals are essential to get a complete perspective of these host-specific communities.



*Results based on sequence analysis of gut contents from 3 elderly individuals at autopsy.

FIGURE 2-2 Overview of molecular studies performed to study the phylogenetic distribution of 16S rRNA gene sequences in the human jejunum and distal ileum.

Metagenome analysis

As the above mentioned studies are largely based on 16S rRNA gene sequences, the results cannot be interpreted beyond the description of the microbial diversity. In contrast, metagenomics can provide a survey of the microbial diversity, as well as an inventory of potential functions, thereby offering a means to link diversity to functionality. However, the effects that sample handling and cloning bias might have on oxygen-sensitive genera like *Bacteroides* and GC-rich organisms like *Bifidobacterium*, respectively, should be taken into account. The term metagenomics refers to genomic analysis of a community by combining the comprehensive analysis of an organism's genetic material (genomics) with the separate statistical analyses of a set of related data (meta-analysis) [165, 181]. A metagenomic library can be constructed by extracting DNA from the microbial community under study followed by cloning of small or large DNA fragments in a suitable host (usually *E. coli*) vector (Fig. 2-3). Provided that the cloned DNA fragments are large enough to encode complete operons they may direct the synthesis of complex molecules or enzymes in the heterologous host used for cloning [181]. The current screening

approaches of metagenomes can be sequence-driven or function-driven [68] and the possibilities and limitations of these approaches are summarised in Table 2-1 and will be discussed below.

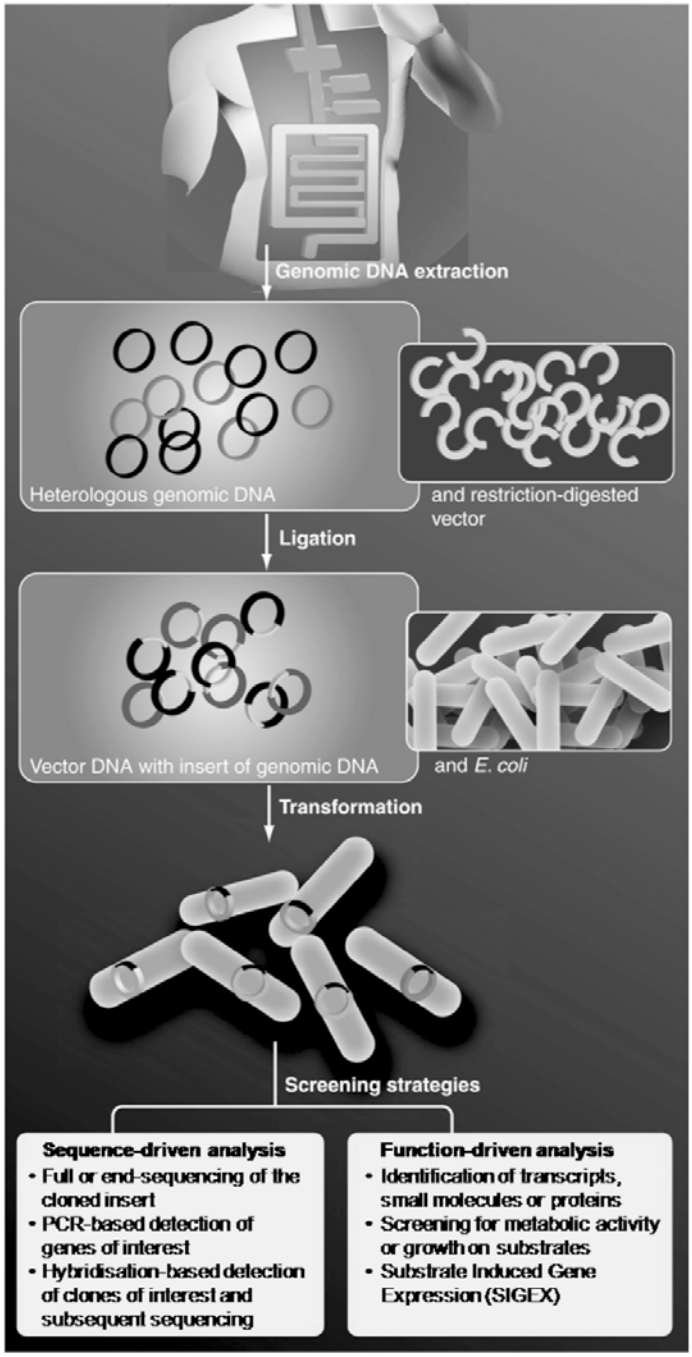


FIGURE 2-3 Schematic overview of construction and screening strategies of metagenomic libraries from GI tract samples.

TABLE 2-1 Comparison of available screening methods of metagenomes.

Screening method	Screening principle	Advantages	Disadvantages	
PCR or hybridisation	DNA sequence consensus	No problem with heterologous gene expression	Requires a database and analyses of the DNA sequence consensus No guarantee to acquire complete forms of genes or gene clusters	**Sequence-driven**
Random sequencing	Sequencing by primer walking	No problem with heterologous gene expression	High redundancy of genes of minor interest Relatively expensive	
Transposon-Aided Capture of plasmids (TRACA)	Plasmid tagging with transposons	Efficient way to study functions encoded by large range of plamids	Genes of interest may be inactivated by transposons Original host species cannot be identified	
Organism-targeted sequencing	Parallel DNA pyrosequencing	Functional genes can directly be linked to species	Only applicable to isolated species	
Ezyme-activity screening	Detectable changes due to enzyme activity	Complete form of genes or gene cluster can be analysed Completely novel genes might be detected	Heterologous expression and secretion essential for detection Labour-intensive	**Function-driven**
Substrate Induced Gene Expression (SIGEX)	Trapping the operon induced by a substrate and sorting by FACS	Substrates taken up by the cytoplasm can be used in native form Fast and cheap	Dependent on recognition of the promotor by host transcription machinery Uptake of substrate essential for detection Sensitive to orientation of functional genes	

Sequence-driven screening strategies

These strategies make either use of PCR or hybridisation approaches to screen for clones that contain genes of interest, or random sequencing of metagenomic clones. The first approach applied to the human GI tract microbiota was used to compare two large-insert (~40kb) metagenomic libraries constructed from DNA extracted from faecal samples of healthy individuals and patients with Crohn's disease (CD). This study revealed that the diversity of the *Firmicutes* was greatly reduced in CD patients [118]. A random sequencing approach was used in the analysis of the human GI tract viral metagenome library constructed from a faecal sample and revealed about 1200 viral genotypes [22]. Another recently published study made use of Transposon-Aided Capture (TRACA) to investigate plasmid-encoded traits in the human gut ecosystem, and genes encoding for phosphoesterase or phosphohydrolase enzymes could be detected [85]. In addition, large-scale sequencing of a small insert library (2-10 kb) of the faecal microbiota was recently reported and a dedicated comparison with the human genome revealed enrichment of genes encoding functions related to the metabolism of for instance glycans and amino acids [64]. In this study, *Bacteroides* (one of the predominant groups in the microbiota of healthy individuals) could not be detected demonstrating the big influence of DNA isolation as well as choice of cloning system on the construction of the metagenomic library.

Tools for comparative and post-metagenomic analyses are indispensable to uncover the wealth of information encoded by the largely scattered pieces of DNA of unknown species present in any metagenomic library. For the re-assembling of the sequenced genomes, software packages have been developed, but due to the immense genomic complexity of microbial communities several scientists prefer to circumvent this computationally intensive step [202]. Comparative metagenome analysis can be performed on raw data-sets to reveal for instance niche-specific gene-function enrichment or to predict differences in site-specific effective genome sizes [53, 154].

As the field of metagenomics is still in its infancy, construction and analyses of metagenomes from digestive ecosystems, especially the small intestine, are still limited. With the continuous development of new screening methodologies and new sequencing techniques, like 454 pyrosequencing [119], and linked to that the decrease in costs, the wealth of information from metagenomic sequence-based analysis will increase in the near future. A recently initiated organism-targeted approach that makes use of pyrosequencing technique aims to deliver the full-genome sequences of 100 representative species from the distal gut microbiota within three years (the Human Gut Microbiome Initiative (HGMI); http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/HGMISeq.pdf). For the retrieval of selected non-cultured-representatives, the consortium proposed the use of germ-free mice [90] combined with mucus isolation by laser-capture microdissection [77, 194]. The sequence information released from this project can provide a key reference for metagenome projects and will further catalyse the development of comparative and post-genomics tools.

Function-driven screening strategies

Whereas sequence-driven analysis is necessarily limited to the discovery of genes in previously described families, function-driven screening of large-insert libraries enables the discovery of new

classes of genes with specific functions. It requires functional transcription and translation of the metagenome clone in the heterologous host used for library construction. In addition, for at least a large part of the functional screening efforts secretion of the encoded product(s) is required for positive clone identification. Given the flexibility of the endogenous transcription and translation machinery of *E. coli*, it is predicted that up to 40% of enzymatic potential from environmental DNA may successfully be expressed in *E. coli* by random cloning [61]. Metagenomic studies that used enzyme activity-based screening led to the discovery of novel hydrolase genes in rumen bacteria [50], β-glucanase genes in the large bowel of mice [215], antibiotic resistance genes in the oral cavity [38], and genes encoding lipolytic enzymes and phosphatases in compost [97]. Screening for growth on several carbon-sources of an ileal mucosa library identified several clones capable of xylan and cellobiose degradation (M. Leclerc, personal communication). As the above described results require the mining of relatively large genomic libraries on indicator agar media or liquid media to obtain 1-5 positives per enzymatic screen, they are quite labour intensive [50].

Enrichment steps for microorganisms harbouring desired traits have recently been employed before library construction to reduce the number of tested clones and to increase the amount of hits in a screen [45]. Arguments against enrichments are the favouring of the fast-growing and culturable part of a microbial community, and loss of microbial diversity [30]. Enrichment cultures combined with metagenomic analysis have not yet been applied to libraries of intestinal origin as most researchers are more interested in the overall functionality of the microbiota rather than retrieving microorganisms with desired functions.

Recently, a high throughput screening strategy, termed substrate-induced gene-expression screening (SIGEX), based on the activation of catabolic genes by various substrates in concert with fluorescence activated cell sorting (FACS) was published [205]. This screening tool makes semi-automated high throughput selection of positive clones possible [224]. However, important constraints of this approach are presented by the requirements related to recognition (and transport) of the inducing substrate as well as the regulated promotor by the corresponding *trans*-acting factor that can either be encoded by the cloned insert or the expression host. As this technique is rather novel it has not been applied to many metagenomic libraries yet, though attempts to screen a metagenome from intestinal origin are ongoing (J.R. Marchesi, personal communication).

An alternative function-driven approach is followed when *in vivo* expressed genes are used as markers for clone selection prior to sequencing. As an example, we have employed sequences retrieved from RNA fingerprinting profiles of the ileal microbiota to trace the encoding metagenomic clones in an ileal effluent metagenomic fosmid library. Several *in situ* expressed genes with either unknown functions or involved in carbohydrate metabolism could successfully be detected in the library and subsequent analysis of insert sequences of the encoding large-insert was performed (see Chapter 7).

CONCLUSION

The microbial ecosystem residing in the human gut is a quite complex one to study, as many factors such as microbiota, dietary components, host genotype and digestive enzymes interplay. In addition, the challenges of small intestinal microbiota sampling provide hurdles in research aiming to study this region of the human GI tract by molecular techniques. The few studies available showed that species belonging to the phylogenetic phyla of Firmicutes, Fusobacteria, Bacteroidetes or γ-Proteobacteria were among the most abundant. With the continuous increase of culture-independent methods as well as availability and validation of model systems researchers will be provided with an increasing toolbox to reveal more details of the diversity and functionality of the microbiota residing in the human small intestine. This will result in an improved understanding of the microbial processes that occur in this largely inaccessible organ.

Analyses of metagenomic libraries have so far shown that most detected genes belong to novel species and gene classes. As the field of metagenomics is hardly ten years old, and the application to intestinal microbial ecosystems is even younger, this high number of detected novel genes is indicative for the enormous genetic potential and metabolic diversity of these ecosystems. As decreasing sequencing costs combined with increasing tools for advanced and effective bio-informatic analysis will enhance sequence-driven analysis of metagenomic libraries, insight into the microbial diversity within the small intestine will develop in parallel. Especially for functional analysis there is a need for further method development that circumvents the laborious screening of individual clones on media. Genes expressed *in situ* by the microbiota might serve as starting point for further metagenome analysis, thereby linking the sequence information to *in situ* activity and functionality.

CHAPTER 3

Population dynamics and diversity of the luminal microbiota in human ileum samples

Carien C.G.M. Booijink, Sahar El-Aidy, Mirjana Rajilić-Stojanović, Hans G.H.J. Heilig, Freddy J. Troost, Hauke Smidt, Michiel Kleerebezem, Willem M. de Vos & Erwin G. Zoetendal

Submitted for publication

ABSTRACT

The diversity and temporal stability of the predominant bacteria in human ileostomy effluent samples was studied with the use of small subunit ribosomal RNA (16S rRNA) gene based approaches. Ileal effluent samples were collected over a period of 28 days from five healthy subjects who had received an ileostomy by total colectomy more than five years ago. The total number of bacteria within the ileostomy effluent was in the range of 10^7-10^8 bacteria per gram. The diversity of the bacteria in the ileostomy effluent showed marked differences compared to that in faecal samples from age-matched healthy adults. The ileostomy effluent had a higher abundance of species within the orders Lactobacillales and Clostridiales, mainly *Streptococcus bovis*-related species, and the *Veillonella* group, and lower abundance of species related to the *Ruminococcus gnavus*, *R. obeum* and *Bacteroides plebeius*-like organisms. In addition, inter-individual differences were found, indicative for a highly personal ileal microbiota profile. Differences in microbial profiles between healthy and medicated subjects with an ileostomy included a higher microbial diversity and an increase in species related to the genera *Bacteroides* and *Eubacterium* in medicated subjects carrying an ileostomy. Furthermore, temporal profiles showed large fluctuations per individual over a period of 8 to 28 days, and differences between morning and afternoon profiles were observed. Parallel cloning and sequencing efforts revealed several phylotypes that were not identified in previous studies (12 out of 65 showed less than 97% sequence similarity with previously reported sequences. Overall, the results indicate that the microbiota of the human ileum is relatively unstable, less complex and consisting of different dominating phylotypes when compared to the colonic microbiota.

INTRODUCTION

The human gastrointestinal (GI) tract harbours a complex community of microbes, which is termed the GI tract microbiota and is crucial for our health [126]. Our knowledge of the diversity of the human GI tract microbiota has expanded considerably during the last decade [234, 235], which is especially true for the microbiota residing within the oral cavity and colon [36]. In contrast, the microbiota in the enclosed regions of the GI tract, including the small intestine, remains largely unknown. The lack of information concerning the small intestinal microbiota is largely due to the accessibility of this part of the intestine. Our current knowledge about the small intestinal microbiota mostly derives from either tissues that were taken during surgical operations [1, 216, 217] or samples from sudden death victims [72]. The latter study was the only to address the ileal luminal microbiota and showed based on 16S clone library sequence analysis that phylotypes within the phylum Proteobacteria and the genera *Bacillus*, *Lactobacillus* and *Streptococcus* were the most abundant in biopsies from three diseased individuals [72]. To date, this is the only reported study that aimed at a description of the ileal luminal microbiota. In contrast, the other studies that were published so far focussed on the ileal mucosa and reported high level representation of numbers of the Bacteroidetes, and *Clostridium* clusters XIVa and IV in 16S clone libraries constructed from biopsy samples [216, 217], or showed that the bacterial cell densities were higher in mucosa samples of the terminal ileum compared to colonic mucosa samples as determined with real-time PCR [1]. Since the collection of biopsy samples from living volunteers is extremely invasive and often requires a pretreatment to remove the bowel contents, these samples are not suitable to study the luminal microbiota over time. Still, there is considerable interest in studying the ileal luminal microbiota as it reflects the microbial communities that have the largest contacts with the intestinal mucosa while being involved in the first conversions of the daily diet.

To circumvent these limitations, subjects with an ileostomy were used as model system since to date these individuals provide the closest possible accessible source to luminal samples of the human ileum in a non-invasive way. Ileostomy subjects are persons who underwent surgery to remove the complete colon because of a disease, mostly colon cancer or inflammatory bowel disease (IBD). The terminal part of the ileum of these subjects is connected to an ileostomy and as a result the ileostomy effluent is leaving the body into an appliance, allowing non-invasive collection of ileostomy effluent samples [66]. In general, ileostomy subjects recover within months following the operation, are healthy and enjoy an active life. A recent study compared the ileal colonisation before and after the surgical removal of the colon in humans and showed no significant changes in bacterial counts and species composition [140]. However, since classical plating approaches were employed this cannot be extrapolated to the complete microbiota. As expected, ileostomy subjects were found to be very suitable to study uptake and digestion of food components [4]. Moreover, multiple time-range sampling can be realised in these subjects in an easy way, especially since the subjects can collect the samples themselves at home.

The aim of this study was to describe the microbial diversity and stability of the ileal effluent microbiota of subjects carrying an ileostomy using 16S rRNA-based approaches. The overall diversity present in ileostomy effluent samples collected over time was lower than that of

the faecal microbiota of age-matched healthy subjects, and temporal fluctuations were larger. Furthermore, phylotypes within the genera *Streptococcus* and *Veillonella* were relatively abundant.

MATERIALS & METHODS

Ileostomy effluent samples

Five healthy subjects with an ileostomy, two medicated subjects with an ileostomy and one individual with an ileorectal pouch participated in this study. The study was approved by the Universtiy Hospital Maastricht Ethical Committee, and conducted in full accordance with the principles of the 'Declaration of Helsinki' (52nd WMA General Assembly, Edinburgh, Scotland, Oct. 2000). Subject characteristics are listed in Table 3-1. The five healthy ileostomy subjects were considered as the primary interest group and samples from the medicated individuals were used for comparison purposes only. The subject with an ileorectal pouch was included to investigate whether the orientation of the terminal part of the ileum (horizontally in case of ileostomy, vertically in case of an ileorectal pouch) affected the microbiota composition. In all volunteers, a total colectomy (removal of the colon) was performed due to IBD complications at least five years before participation, enabling the restoration of steady-state conditions after the initial adaptation phase to the colectomy. Apart from absence of the colon, the volunteers had no known abnormalities of the digestive system and had not been subjected to any feeding trial, specific diet, or antibiotic treatment for the last year.

Five out of the seven selected subjects with an ileostomy did not take any medication at least one month prior to and during the study, and were considered as the healthy group. The other two volunteers were taking mesalamine or prednisone due to relapse of Crohn's disease during the study, which may have affected the bacterial composition. These volunteers collected fresh ileostomy effluent samples by emptying the ileostomy effluent in freezer baskets as soon as the bulk of ileostomy effluent was collected in the morning and afternoon. The time samples were collected at least three hours apart and stored on dry ice at approximately -80°C. Samples were processed within three days after collection.

TABLE 3-1 Overview of the characteristics of the subjects included in this study.

Person	Gender	Age	Ileostomy/pouch	Medication
i1	female	55	ileostomy	none
i2	female	60	ileostomy	none
i3	male	57	ileostomy	none
i4	male	74	ileostomy	none
i5	male	55	ileostomy	none
m1	male	41	ileostomy	prednisone
m2	female	52	ileostomy	mesalamine
p1	female	48	ileorectal pouch	none

DNA isolation

DNA was isolated from 0.2 g of fresh ileostomy effluent samples as described before [233] by usage of the Stool DNA Isolation Kit (Qiagen, Leiden, the Netherlands). DNA was quantified spectrophotometrically (Nanodrop ND-1000 spectrophotometer, NanoDrop® Technologies, Wilmington, USA).

Real-time PCR detection

Real-time PCRs were performed in a reaction volume of 25 µl containing 1× SYBR Green PCR Master Mix (Applied Biosystems, Foster City, USA), 100 nM each of the primers from MWG (Ebersberg, Germany), and 5 µl of 10 times diluted DNA, extracted from the ileostomy effluent samples. For quantification of total bacteria and the *Lactobacillus*-cluster the following primers were used: Bact-1369f with Prok-1492r [198] and LactoF with LactoR [27] (Table 3-2). The archaea were quantified with the use of primers A-751F and A-976R [9]. Real-time PCR amplification and detection of DNA were performed with the iQ5 (Bio-Rad Laboratories B.V., Veenendaal, the Netherlands). Data analysis was conducted with iQ5 Optical System Software Version 1.1.

TABLE 3-2 Primers used in this study.

Primer	Sequence (5'-3')	Experiment	Reference
T7prom-Bact-27f	TGAATTGTAATACGACTCACTATAGGGGTTTGATCCTGGCTCAG	HITChip	Rajilić-Stojanović, 2007
Uni-1492r	CGGCTACCTTGTTACGAC	HITChip	Rajilić-Stojanović, 2007
A-751F	CCGACGGTGAGRGRYGAA	q-PCR	Baker *et al.*, 2003
A-976R	YCCGGCGTTGAMTCCAATT	q-PCR	Reysenbach & Pace, 1995
Bact-1369f	CGGTGAATACGTTCYCGG	q-PCR	Suzuki *et al.*, 2000
Prok-1492r	GGWTACCTTGTTACGACTT	q-PCR	Suzuki *et al.*, 2000
LactoF	TGGAAACAGRTGCTAATACCG	q-PCR	Byun *et al.*, 2004
LactoR	GTCCATTGTGGAAGATTCCC	q-PCR	Byun *et al.*, 2004
U-968-GC-r	**GC-clamp**-AACGCGAAGAACCTTAC	DGGE	Nubel *et al.*, 1996
L-1401r	CGGTGTGTACAAGACCC	DGGE	Nubel *et al.*, 1996
Bact-27f	AGAGTTTGATYMTGGCTCAG	DGGE	Kane *et al.*, 1993
Lab-667r	CACCGCTACACATGGAG	DGGE	Heilig *et al.*, 2002
Lab-159f	GGAAACAGRTGCTAATACCG	DGGE	Heilig *et al.*, 2002
Uni-515-GC-r	**GC-clamp**-ATCGTATTACCGCGGCTGCTGGCAC	DGGE	Lane, 1991
GC-clamp	CGCCGGGGGCGCGCCCCGGGCGGGGCGGGGGCACGGGGGG	DGGE	Muyzer *et al.*, 1993
T7	TAATACGACTCACTATAGG	Sequencing	Promega
Sp6	GATTTAGGTGACACTATAG	Sequencing	Promega
907r	TACCCGTCAATTCCTTTGAGTTT	Sequencing	Lane, 1991
1100r	TACAAGCTTAGGGTTGCGCTCGTTG	Sequencing	Lane, 1991

HITChip analysis

For phylogenetic analysis of the ileostomy effluent samples, the HITChip was used. This phylogenetic microarray consists of more than 4,800 oligonucleotides based on 16S rRNA gene sequences targeting over 1,100 intestinal microbial phylotypes. The procedure for hybridisation and analysis was performed as described before [155]. In short, the 16S rRNA gene was amplified

from 10 ng DNA extracted from ileostomy effluent samples with the *T7prom*-Bact-27f and Uni-1492r primers (Table 3-2). After purification of the PCR products (DNA Clean & Concentrator, Zymo Research, Orange, USA) the DNA concentration was measured using a NanoDrop spectrophotometer.

The *T7*-promotor and subsequent 16S rRNA gene were transcribed *in vitro* to RNA using the Riboprobe System (Promega, La Jolla, USA) according to the manufacturers' instructions with a nucleotide mix that also contained aminoallyl-rUTP (Ambion Inc., Austin, Texas, USA). The template DNA was digested with RNAse-free DNAse (Qiagen, Hilden, Germany), and RNA was purified (RNeasy Mini-Elute Kit, Qiagen, Hilden, Germany). Subsequently, the amino-allyl-modified nucleotides were labelled with CyDye (Post-Labeling Reactive Dye, GE Healthcare, Little Chalfont, UK). After stopping the reaction, the labelled RNA was purified and quantified as described before.

Microarrays synthesised by Agilent Technologies (Agilent Technologies, Palo Alto, USA) in format 8×15K were used for hybridisation with two samples, labelled with Cy3 and Cy5, respectively. The Cy3/Cy5 labelled target mixes were fragmented with 10× fragmentation reagent (Ambion Inc., Austin, Texas, USA). Hybridisation on the arrays was performed at 62.5°C for 16 h in a rotation oven (Agilent Technologies). Slides were washed at room temperature in 2× SSC with 0.3% SDS for 10 min, followed by 0.1× SSC with 0.3% SDS at 38°C for 10 min and 0.06× SSPE for 5 min [171].

Data extraction from the microarray images was performed using the Agilent Feature Extraction software, versions 7.5-9.1 (http://www.agilent.com). Subsequently, the microarray data normalisation and the further microarray analysis were performed using a set of R-based scripts (http://r-project.org) in combination with a custom designed relational database which runs under the MySQL database management system (http://www.mysql.com) [155]. After the normalisation hierarchical clustering of probe profiles was carried out using a distance matrix based on the squared difference between each pair of profiles (squared Euclidian distance) with the use of Ward's minimum variance method.

PCR amplification

DNA isolated from total ileostomy effluent community was used as a template to amplify the V6 to V8 regions of 16S rRNA gene with primers U-968-GC-f and L-1401r [143] (Table 3-2). Amplification of 16S rRNA gene fragments from the *Lactobacillus* group was performed by a nested-PCR approach. First, partial amplification of the 16S rRNA gene was performed with the primers Bact-27f and Lab-677r. This PCR product was 10-fold diluted and used for subsequent amplification with primers Lab-159f and Univ-515-GCr as described before [73] (Table 3-2). The latter two primers target, besides species belonging to the genus *Lactobacillus* (except *L. maltaromicus*, *L. catenaformis* and *L. vitulinus*), also *Leuconostoc* spp., *Weisella* spp., *Aerococcus* spp. and *Pediococcus* spp. [73].

PCRs were performed using a *Taq* DNA polymerase kit from Invitrogen Life Technology (Breda, The Netherlands). The reaction mixture consisted of 20 mM Tris HCl [pH 8.4], 3 mM $MgCl_2$, 0.2 mM deoxynucleoside triphosphate (dNTPs), a 0.2 μM concentration of each primer, 1.25 U of *Taq* polymerase, and 1 μl of appropriately diluted template DNA in a final volume of

50 µl. Samples were amplified in a PE Applied Biosystems GenAmp PCR system 9700 (Foster City, California, USA) using the following program: pre-denaturation for 2 min at 94˚C and 35 cycles of 94°C for 30 sec, 56˚C for 40 sec, 72˚C for 1 min, and final extension at 72˚C for 5 min. The integrity of the nucleic acids was checked visually after electrophoresis of 5 µl of PCR products on a 1.2 % agarose gel containing ethidium bromide.

DGGE analysis and calculation of similarity indices

PCR amplicons were separated by DGGE based on the protocol as described by Heilig *et al.* [73] using the Dcode system (Bio-Rad Laboratories, Hercules, California, USA) with some minor modifications. In short, the polyacrylamide gels (8 % [vol/vol] polyacrylamide (ratio of acrylamide-bisacrylamide, 37.5:1)) were made with denaturing gradients ranging from 30 to 60 %, and 30 to 55 % to separate the generated amplicons of the total bacteria and the *Lactobacillus* communities, respectively. Electrophoresis was performed first for 5 min at 200V followed by 16 h at a fixed voltage of 85 V in 0.5× TAE buffer at constant temperature of 60˚C. The gels were stained after completion of electrophoresis with $AgNO_3$ and developed [173]. After overnight drying at 60˚C, DGGE gels were scanned at 400 dpi and the statistical software package BioNumerics 3.5 (Applied-Maths, Sint-Martens-Latum, Belgium) was used to compare the DGGE profiles. All temporal samples from one subject were loaded on the same gel to facilitate intra-individual profile comparison.

Cloning of the PCR-amplified products and sequence analysis

PCR amplicons generated with primers for the total bacterial (U-968-GC-f and L-1401r) and the lactobacilli community (Lab-159f and Lab-515-GC-r) were purified with the Qiaquick PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturers' instructions. Purified PCR products were quantified by loading 1 µl on a NanoDrop and cloned in *E. coli* JM109 using the Promega pGEM-T vector system (Promega, Madison, USA). Colonies of ampicillin-resistant transformants were transferred with a sterile toothpick to 50 µl of TE and were lysed at 95˚C for 15 min. PCR was performed on cell lysates using pGEM-T-specific primers T7 and SP6 to check the size of the inserts. Plasmids containing inserts with the right size were used to screen with the corresponding V6 to V8 primers for total bacteria or lactobacilli, as mentioned above.

Clones for subsequent sequence analysis were selected according to the migration position of the cloned PCR fragment in the DGGE compared to the original DGGE profile. Insert PCR amplicons of selected transformants were purified by the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and were subjected to DNA sequence analysis (BaseClear, the Netherlands). Sequence similarity was analysed using BLAST tool [3] at the NCBI database. The phylogenetic position of the sequences was also scored based on the ARB database release of 2002 [108] with the addition of all 16S rRNA sequences retrieved from the human gastrointestinal tract as recently performed by Rajilić-Stojanović [155].

Statistical analysis

The diversity of the microbial community assessed by HITChip analysis was expressed as Simpson's reciprocal index of diversity (1/D) [189]. This diversity measure was calculated using

the equation $\lambda = 1 / \Sigma Pi^2$ where Pi is the proportion of ith taxon. The proportion of each taxon was calculated as the proportion of each probe signal compared to the sum of all probe signals [155]. Simpson's reciprocal index of diversity takes the number of taxons present in the sample, as well as their abundance in the community into account. Therefore, a higher value of this index corresponds to a more diverse community.

Principal component analysis (PCA) was used as implemented in the software package Canoco for Windows 4.5 [102]. The average signal intensities for 129 bacterial groups defined on the bases of the 16S rRNA gene sequence similarity, or single probe hybridisation intensity per sample were used as species data. PCA was performed focussing on inter-species correlation and diagrams were plotted by using the CanoDraw program.

Similarity of temporal profiles of the predominant microorganisms as profiled by DGGE was assessed by calculating the similarity indices (SI's) of the compared profiles based on densitometric curves [69] of the scanned DGGE profiles by using the Pearson's correlation [58]. Similarity between the temporal in-depth microbiota compositions based on HITChip profiles of ileal effluent samples of four healthy subjects with an ileostomy was also calculated based on the Pearson's correlation, which reflects the degree of linear relationship between analysed datasets. The Student's t-test was used to calculate the P-values, based on two-tailed hypothesis testing.

RESULTS

Analysis of the total bacterial diversity in ileostomy effluent

To quantify the total number of bacteria and archaea present in ileostomy effluent samples, real-time PCR was performed using universal bacterial and archaeal primers on DNA isolated from all collected time samples from the seven ileostomy subjects and one subject with an ileorectal pouch as template. The amount of archaeal DNA was below the detection limit in all ileostomy effluent samples tested, while the total copy number of bacteria ranged from 4.8×10^7 to 3.4×10^8 per gram ileostomy effluent for all five healthy subjects with an ileostomy (Fig. 3-1). Copy numbers retrieved for the subjects with an ileostomy under medication and the person with an ileorectal pouch were in the same range (Fig. 3-1).

The microbial composition of the ileostomy effluents was profiled using the HITChip, a phylogenetic microarray. Subsequent cluster analysis of phylogenetic fingerprints obtained by the HITChip resulted in grouping of the profiles of five healthy ileostomy subjects with the profile of the person with an ileorectal pouch, whereas the two medicated ileostomy subjects clustered separately (Fig. 3-2). The sub-grouping of the healthy subject profiles from the medicated subjects with an ileostomy was caused by a relatively high abundance of bacteria belonging to the genera *Veillonella*, *Streptococcus*, and *Clostridium* cluster I in healthy subjects. In addition, within the phylum Bacteroidetes and *Clostridium* cluster XIVa hardly any species, except *Clostridium nexile*-like species were detected in the healthy individuals, whereas bacteria belonging to these groups were abundant in the medicated subjects with an ileostomy.

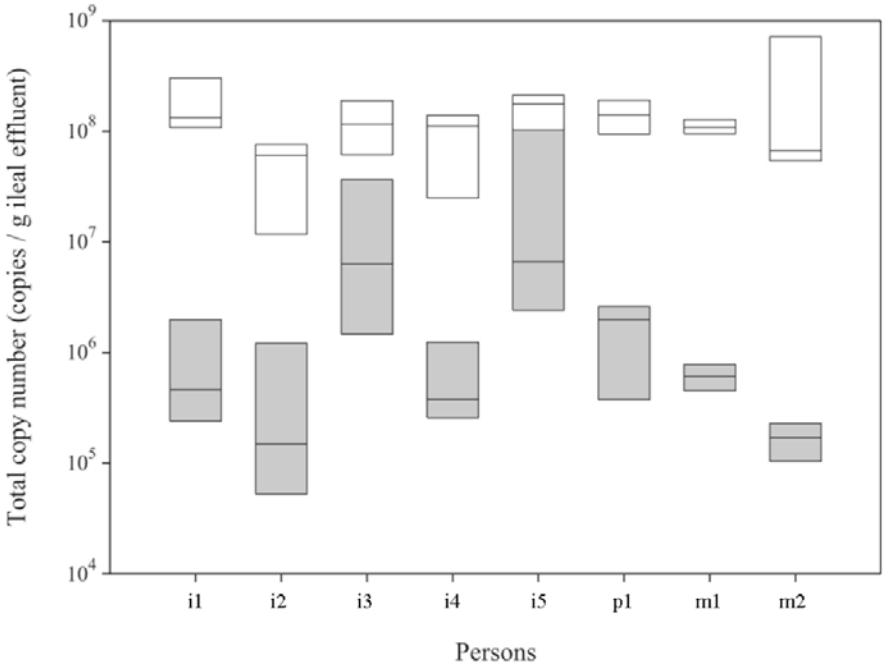


FIGURE 3-1 Average copy number of total bacteria and lactobacilli per gram of ileostomy effluent samples determined with real-time PCR with universal or *Lactobacillus*-specific primers, respectively, expressed in boxplots. Samples used were from of five healthy ileostomy subjects (i1 to i5), two medicated ileostomy subjects (m1 and m2) and one person with an ileorectal pouch (p1). The box extends from the 25th percentile to the 75th percentile, with a line at the median. Copy numbers for lactobacilli are depicted as gray boxes, the total copy number as white boxes.

Analysis of the temporal stability of the ileal microbiota of healthy subjects with an ileostomy

For subsequent in-depth analysis of the dynamics of the ileal effluent microbiota, samples collected over time from four of the healthy subjects carrying an ileostomy were profiled by HITChip. All samples profiled were collected at the same time of the day over a period ranging from 10 days for three persons and 28 days for one person. Subsequent Pearson correlation analysis, revealed an individual-based clustering of the temporal microbiota profiles (Fig. 3-3a). As a measure for temporal stability, the similarity index based on Pearson correlation was calculated. For the healthy ileostomy subjects the average similarity index over a period of nine days was 44% based on HITChip profile comparison. In addition, for one person (i4) both morning and afternoon samples were profiled to get insight in the fluctuations within a day. Subsequent cluster analysis resulted in grouping of the morning profiles separately from the afternoon profiles (Fig. 3-3b). This result indicates that the ileostomy effluent microbiota shows short-term fluctuations (within a day) that exceed the fluctuations observed in effluents sampled over extended periods of time. Though the dynamics of the ileostomy effluent microbiota composition is high, it possibly fluctuates around a stable core. To investigate this further, the probes responding in all ileostomy effluent samples collected over time were deduced from the complete dataset. This revealed that 40% of responding probes per subject were common for all analysed samples, corresponding to a hypothetical ileal effluent common core of eight phylogenetically related groups, including species belonging to the genera *Veillonella*, *Streptococcus*, *Clostridium* cluster I and *Enterococcus* (Fig. 3-S1).

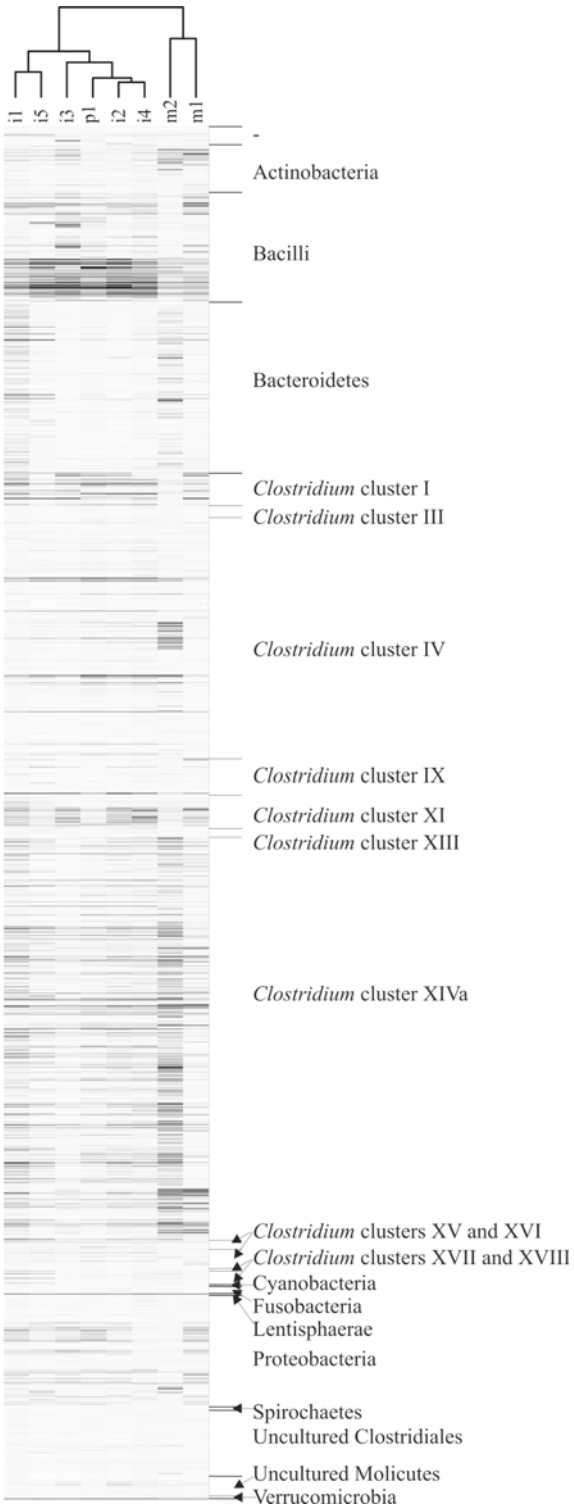


FIGURE 3-2 Phylogenetic fingerprints of seven ileostomy subjects (five healthy and two medicated) and one person with an ileorectal pouch based on HITChip. Abbreviations used are: i1 till i5 for the healthy ileostomy subjects, m1 and m2 for ileostomy subjects taking medication and p1 for the person with an ileorectal pouch. The clustering of the eight ileal effluent profiles is depicted above the fingerprints, whereas the highest phylogenetic level of specificity of probes is indicated on the right panel of the figure.

To gain insight in the fluctuations in relative abundance of phylogenetic groups present in the ileostomy effluent samples over time, changes in the average signal intensities for 129 bacterial groups represented on HITChip were studied. This revealed that person i2 shows a remarkably stable profile over time, with high abundances of bacterial species related to the genera *Streptococcus* and *Veillonella*, which contributed on average 47% and 26%, respectively to the total microbiota (Fig. 3-3c depicted in different shades of grass green and orange). For the other three ileostomy subjects the *Streptococcus*- and *Veillonella*-related species also appeared to be the predominating ones. However, person i1 deviates from the other three persons in the relatively high abundance of *Clostridium*-like organisms present in all time samples (Fig. 3-3c depicted in different shades of yellow), and the increase of organisms belonging to the Bacteroidetes up to a level of 30% on day 10 (depicted in different shades of aqua-blue). Another remarkable difference between the subjects with an ileostomy is the presence of *Lactobacillus*-like groups (depicted in different shades of gray); in person i3 this group contributes approximately 13% to the total profile in most time points with the exception of day 3 with a level as high as 47% of *Lactobacillus*-like organisms. Overall, remarkable differences between the ileostomy subjects were found, in composition as well as temporal stability of the microbiota, indicative for a host-specific microbiota.

Analysis of the lactobacilli in the ileal microbiota of ileostomy subjects

To quantify the *Lactobacillus*-like population, real-time PCR using *Lactobacillus*-specific primers was applied. The copy number corresponding to the lactobacilli population varied between 6.1×10^5 and 4.3×10^7 for the healthy subjects carrying an ileostomy (Fig. 3-1). For two subjects (i3 and i5) the lactobacilli were relatively abundant, averaging on 10.9% and 3.4% of the total community and notably for subject i3 peaking at relative population sizes as high as 43% in one sample (t = 3m), comparable to the 47% relative contribution of lactobacilli detected by HITChip. In contrast, significantly lower average relative abundance of lactobacilli was detected in the other three healthy subjects with an ileostomy (0.16 to 0.90%), the ileostomy subjects under medication (0.04 to 0.14%) as well as the person with an ileorectal pouch (0.01 to 0.40%). The correlation between the results obtained by the *Lactobacillus* specific real-time quantification and the sum of all *Lactobacillus*-related HITChip hybridisation signals was analysed. Results of both techniques correlated with strong positive correlation with Pearson's index of 0.97, indicating that the HITChip hybridisation signals can be used to assess quantitative information on the ileostomy effluent microbiota.

To get insight into the species diversity of the *Lactobacillus*-like phylogenetic group within the ileal effluent microbiota, *Lactobacillus*-specific primers were used for construction of a 16S clone library and subsequent sequence analysis was performed. A total of 33 sequences were obtained from the eight subjects and these were used to construct a phylogenetic tree (Fig. 3-4). Of these sequences, 24 showed 97-99% similarity with 16S rRNA database entries. The majority of these originated from species within the genera *Lactobacillus* and *Leuconostoc*. The specificity of the used primers (Lab-159f and Uni-515r) also includes the genera *Pediococcus* and *Weissella* [73]; only one representative of these two genera were found once in the clone library. Furthermore, nine sequences that did not show significant similarity to sequences present in GenBank were detected in six out of eight volunteers.

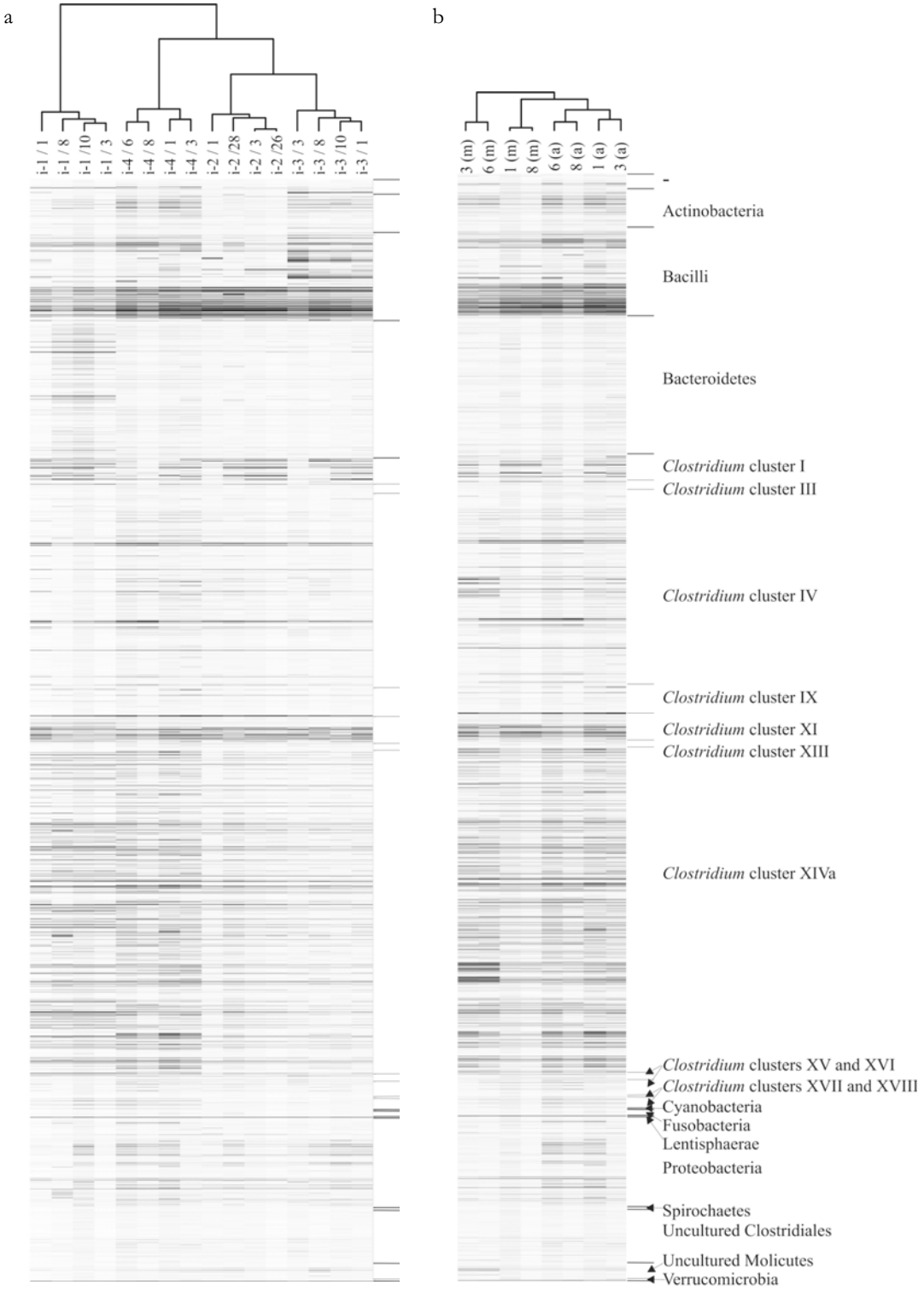


FIGURE 3-3 (**a**) Phylogenetic fingerprints of the human ileostomy effluent microbiota of four healthy ileostomy subjects on four time points. The healthy subjects are encoded by i1 to i4, while the numbers after the forward slash indicate the day of sampling. The highest phylogenetic level of specificity of probes is depicted on the right side of the panel. (**b**) Temporal variation of one ileostomy subject (i4) in time. The number indicates the day of sampling, whereas m = morning and a = afternoon sample.

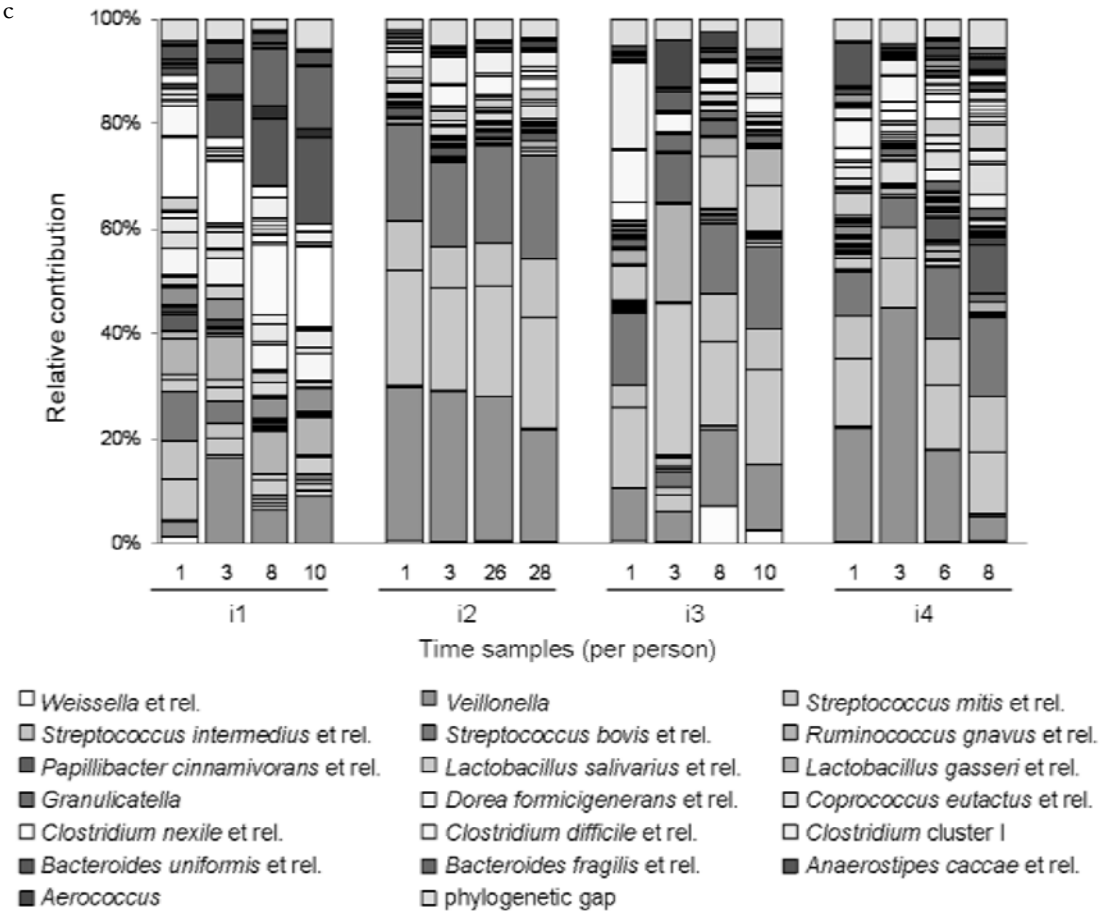


FIGURE 3-3 (**c**) Relative contribution of the detected phylotypes with HITChip in four ileostomy effluent samples collected over time of four healthy ileostomy subjects. The day of sampling per person is indicated under the bars. In the legend, phylotypes that contribute at least 5% to one of the profiles are indicated. The added spike is indicated as phylogenetic gap on top of the columns. A colour version of this picture can be found in de supplementary and colour figures section at the end of this thesis.

The closest relatives of these sequences were found to belong to the genera *Lactobacillus* and *Leuconostoc* (Fig. 3-4). The relatively high retrieval of unknown phylotypes (27%) probably can be attributed to the relatively unexplored nature of the ileal *Lactobacillus* population.

Analysis of novel phylotypes in the ileal microbiota of ileostomy subjects

Since sequence analysis of a *Lactobacilli*-related 16S clone library revealed several as yet unknown phylotypes, the abundant microbiota was analysed for new bacterial phylotypes. Use was made of DGGE analysis of abundant 16S amplicons followed by cloning and sequencing. Differences between the eight subjects were observed in terms of number and position of the detected bands in the universal bacterial DGGE profiles between individuals, whereas the average number of ten bands was observed (±3.6; Fig. 3-5a). Those were targeted for cloning as described previously [230] and plasmid DNA from the corresponding clone was purified and its nucleotide sequence was subsequently determined. Comparison to sequences available in GenBank showed that out

of the 72 sequences, 12 showed less than 97% identity with any deposited GenBank sequence, indicative that these sequences were derived from new not (yet) described bacterial species. Furthermore, a relative dominance of species belonging to the phylum of the *Firmicutes*, mainly genera *Clostridium* and *Streptococcus*, became obvious (Fig. 3-5b), as is in accordance with the HITChip results. The relatively low number of detected sequences within the genus *Veillonella*, in contrast to the HITChip results, can possibly be explained by cloning bias or by missed clones.

To estimate the coverage of HITChip, the 72 sequences retrieved from the ileostomy effluent clone libraries were added to the same ARB tree as used for probe construction of HITChip (Fig. 3-S2). This phylogenetic analysis showed that about 77% of the retrieved sequences were covered by HITChip (≥97% sequence similarity with phylotypes targeted by HITChip) (Fig. 3-S1).

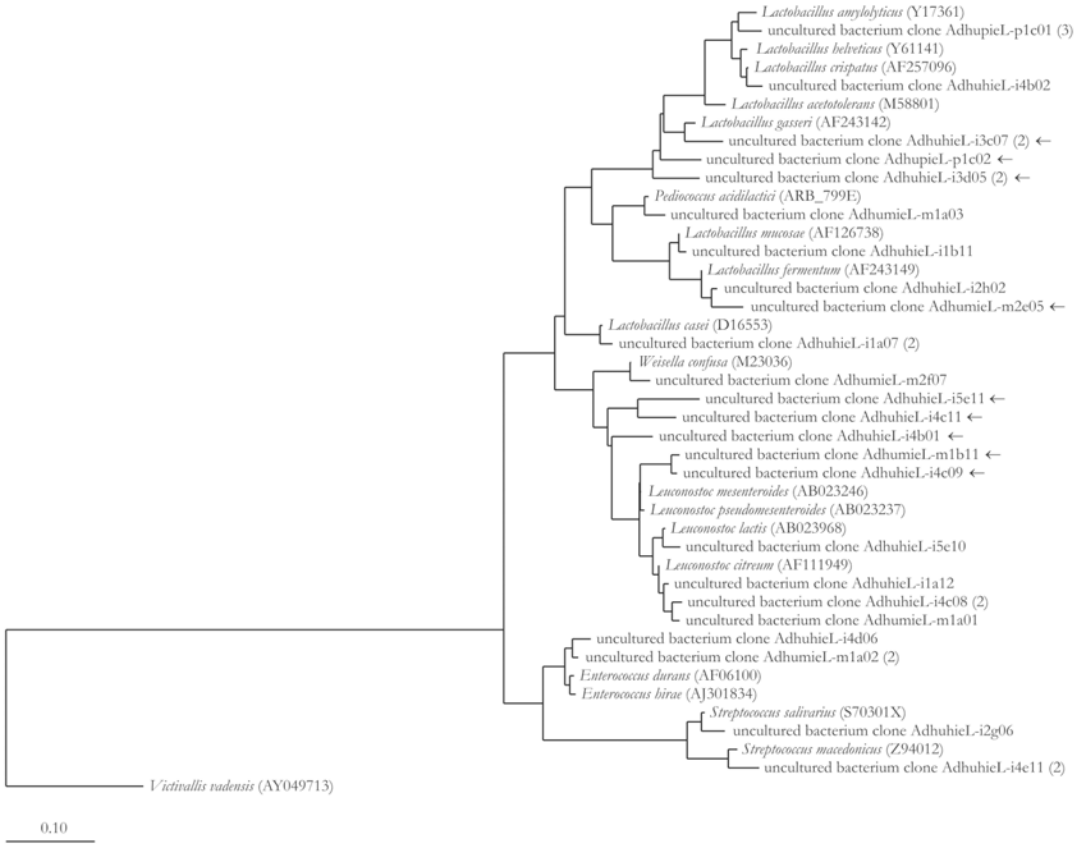
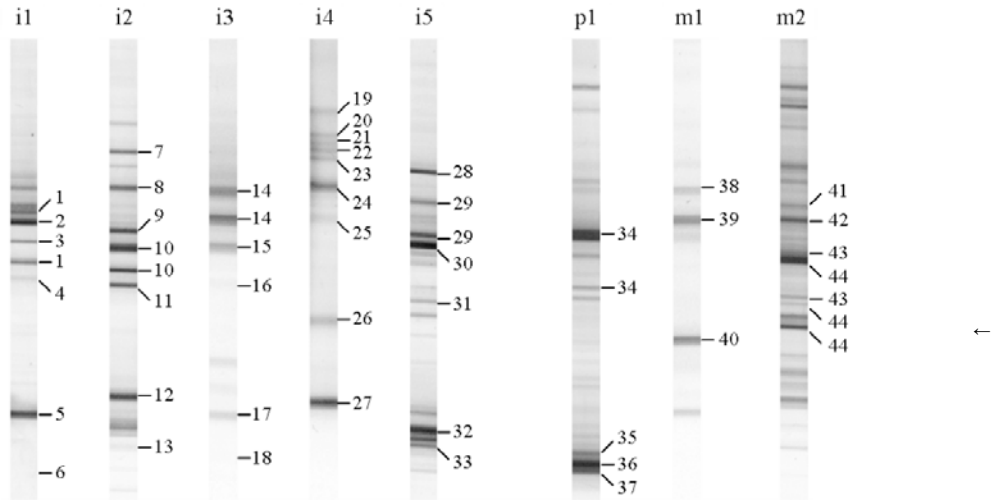


FIGURE 3-4 Neighbour-joining tree of the *Lactobacillus* species and phylotypes based on 33 sequences of the *Lactobacillus*-specific 16S rRNA amplicon (± 400 bp amplified with Bact-27f and Lab-667r followed by Lab159r and Uni515r) and closest related cultured representatives in GenBank (accession numbers in parentheses) identified in ileostomy effluent samples collected in time. The eight individuals are identified as five healthy ileostomy subjects (i1 to i5), two medicated ileostomy subjects (m1 and m2) and one person with an ileorectal pouch (p1). Numbers between parentheses behind the clone reference indicate the frequency of the sequence within the clone library. Arrows indicate sequences with <97% similarity to its closest relative. Sequences were aligned with Pearson correlation within ARB, and the distance matrix was calculated with the neighbour joining algorithm. The phylogenetic tree was rooted with the appropriate amplicon region of the 16S rRNA gene sequence of *Victivallis vadensis*. The scale bar represents the genetic distance.

a



b.

Band	Subject	Closest relative	# clones	Clones (% similarity)
1	i1	*Ruminococcus gnavus*	1	i1b10 (98)
2		*Yersinia frederiksenii*	1	i1a04 (96)
3		*Clostridium glycolicum*	1	i1a03 (99)
4		*Clostridium celatum*	1	i1a01 (98)
5		*Escherichia coli*	1	i1a06 (98)
6		*Bacteroides fragilis*	1	i1b09 (98)
7	i2	*Proteus mirabilis*	1	i2d08 (99)
8		*Clostridium perfringens*	1	i2d09 (99)
9		*Streptococcus salivarius*	1	i2b06 (99)
10		*Clostridium celatum*	3	i2a8 (98), i2a10 (99)
				i2a11(98)
11		*Haemophilus parainfluenzae*	1	i2b04 (98)
12		*Streptococcus bovis*	1	i2a09 (98)
13		*Eubacterium tenue*	1	i2b01 (98)
14	i3	*Clostridium perfringens*	2	i3f05 (99), i3f04 (99)
15		*Clostridium butyricum*	1	i3f06 (99)
16		*Clostridium glycolicum*	2	i3f01 (99), i3h01 (99)
17		*Escherichia coli*	1	i3h04 (97)
18		*Bifidobacterium dentium*	1	i3f02 (97)
19	i4	*Eubacterium tenue*	2	i4a07 (93), i4a08 (93)
20		*Clostridium bartletti*	1	i4h10 (95)
21		*Streptococcus oralis*	1	i4b07 (98)
22		*Clostridium nexile*	2	i4a02 (99), i4b08 (98)
23		*Clostridium celatum*	2	i4a04 (99), i4a03 (98)
24		*Streptococcus bovis*	1	i4g01 (99)
25	i4	*Clostridium glycolicum*	4	i4b03 (98), i4a03 (98)
				i4b08 (97), i4b11 (93)
26		*Veillonella dispar*	5	i4b01 (99), i4b06 (97)
				i4b10 (97), i4a11 (90)
				i4a06 (92)
27		*Streptococcus salivarius*	3	i4e07 (92), i4a05 (91)
				i4f07 (90)
28	i5	*Clostridium ramosum*	1	i5c04 (99)
29		*Clostridium butyricum*	2	i5a11 (99), i5c02 (98)
30		*Ruminococcus gnavus*	1	i5c03 (99)
31		*Fusobacterium nucleatum*	1	i5a10 (92)
32		*Streptococcus bovis*	2	i5a07 (99), i5c01 (98)
33		*Clostridium bartlettii*	1	i5a08 (94)
34	m1	*Proteus mirabilis*	1	m1e08 (99)
35		*Yersinia frederiksenii*	1	m1e09 (96)
36		*Ruminococcus obeum*	1	m1d10 (94)
37	m2	*Ruminococcus gnavus*	1	m2g08 (98)
38		*Clostridium nexile*	3	m2g10 (99), m2h05 (98)
				m2h06 (98)
39		*Ruminococcus obeum*	1	m2h02 (97)
40		*Ruminococcus gnavus*	2	m2g07 (98), m2h03 (97)
41	p1	*Clostridium perfringens*	1	p1d06 (99)
42		*Escherichia coli*	1	p1c11 (99)
43		*Escherichia coli*	2	pc09 (98), p1d03 (97)
44		*Streptococcus infantarius*	2	pd01 (99), pd04 (99)

FIGURE 3-5 (**a**) Distribution of dominant bands in the DGGE pattern of the V6 to V8 regions of ileostomy effluent samples of eight individuals. Healthy ileostomy subjects are indicated with i1 to i5; medicated ileostomy subjects are indicated with m1 and m2 and the individual with an ileorectal pouch is indicated with p1. The bands identified from the 16S rRNA gene clone libraries are indicated with numbers 1 to 44. (**b**) Closest relatives as determined by comparative sequence analysis, level of identity with this relative and clone designation for each band indicated in (**a**).

DISCUSSION

In this study we have used several culture-independent techniques to investigate the human luminal ileum microbiota diversity and the temporal stability. As a model for the human ileum we have used ileal effluent samples from a group of subjects with an ileostomy, which were used for in-depth diversity analysis by HITChip and DGGE profiling. In addition, specific bands from the

DGGE profiling gels were subjected to sequence analysis and real-time PCR, and targeted DGGE was employed to zoom in on the group of lactobacilli, which were shown to display drastic variation in abundance between individuals. Inter-individual differences in the diversity and temporal stability of the microbiota HITChip profiles were observed (Figs. 3-2 and 3-3a). This indicated that each individual has his own ileal microbiota with few groups in common. In the present study, diet seems to play an important role as different microbiota profiles were observed between morning and afternoon samples collected on the same day (Fig. 3-3b). Overall, these results support the generally accepted view that each healthy person has his or her own unique microbiota.

The phylogenetic profiles of the eight included subjects (five healthy ileostomy subjects, two medicated ileostomy subjects and one subject with an ileorectal pouch) were compared with eight randomly chosen faecal profiles from age-matched healthy adults investigated previously by HITChip. This analysis showed similarities between the medicated ileostomy subjects and faecal profiles mainly concerning the relative abundance of *Bacteroides vulgatus*, *Eubacterium hallii*, *E. rectale* and *Ruminococcus obeum*-like organisms (Fig. 3-S3). This was confirmed by principal componentanalysis (PCA) that showed a clear separation between the ileostomy effluent and faecal microbiota (Fig. 3-S4). This statistically significant distinction ($p = 0.007$) was mainly due to higher relative abundance of eight groups in ileostomy effluent samples, including *Veillonella*, Fusobacteria, and *Streptococcus*-related phylotypes. In addition, the PCA triplot showed that the intra-individual differences in microbiota composition were larger in case of ileostomy effluent samples compared to the faecal samples (Fig. 3-S4). The average diversity of the microbiota of effluent samples of healthy subjects with an ileostomy appeared to be comparable to the microbiota diversity of ileostomy subjects under medication (average Simpson's indices of 39.3 ± 7.5 and 58.0 ± 17, respectively). However, this was found to be significantly lower compared to the microbial diversity in faeces (average Simpson's index of 136.0 ± 64.5; $p = 0.013$).

Comparison of ileostomy effluent samples collected over time revealed fluctuations in microbial composition that were very large within a day, indicative for a highly dynamic ecosystem, but seemed to return to a stable core over longer time intervals. For the healthy subjects with an ileostomy the average similarity index over a period of nine days was 44% based on HITChip profile comparison. In contrast, a similarity index as high as 76% was found for DGGE profiles of faecal samples with two year time difference [230]. This difference is illustrative for the relative instability of the ileostomy effluent microbiota compared to the faecal microbiota. A hypothetical ileostomy effluent common core of eight phylogenetically related groups was deduced based on probe signals detected in all ileostomy effluent samples collected over time by HITChip (Fig. 3-S1). Not surprisingly, probes targeting species that belong to the genera *Veillonella* and *Streptococcus* belonged to this common group. The very fast sugar metabolism and transport systems of streptococci [129] make them well equipped competitors for substrates during the fast transit through the small intestine. Carbohydrate fermentation by the streptococci leads to the formation of lactate, which can be further fermented by *Veillonella*-species to acetate and propionate [55]. Possibly these groups fulfill key roles in the ileal microbiota, dictated by the environmental factors of this microbial ecosystem.

The single reported study that aimed to describe the microbiota residing in the luminal part of the human ileum, used mucosal tissue samples collected from sudden death victims [72] and revealed a relative abundance of species within the phylum Proteobacteria, and the genera *Bacillus*, *Streptococcus* and *Lactobacillus*. Comparison to the ileostomy effluent microbiota described in this study indicated a marked difference in lower relative abundance of species with the Proteobacteria phylum, as well as higher relative abundance of species within the different *Clostridium* clusters (Fig. 3-S5). Here the age difference between the subjects should be taken into account, since the average age of the sudden death victims was 82 years, compared to an average age of 59 years for the healthy ileostomy subjects. Aging has an effect on the microbiota composition, and higher retrieval of species within the γ-Proteobacteria in faecal samples of elderly was reported by the same researchers before [71]. Furthermore, the time interval between the time of death and sampling is crucial for the representativeness of the microbiota composition for the *in vivo* situation, since due to relapse of ileal transit and complete oxygen depletion, the microbiota composition is likely to change completely.

In conclusion, this study showed that the ileostomy effluent microbiota is less diverse and less stable compared to the faecal microbiota. The total number of bacteria in ileostomy effluent samples was in the range of 10^7 to 10^8 bacteria per gram. Main differences in species distribution were a higher relative abundance of species within the class Bacilli, more specifically within the genera *Streptococcus* and *Veillonella*. Also *Clostridium* cluster I was detected in high levels, whereas a lower relative abundance of species belonging to the phylum Bacteroidetes, and *Clostridium* clusters IV and XIVa was observed. Differences observed between healthy and medicated ileostomy subjects included the higher relative abundance of species belonging to the genera *Eubacterium* and *Bacteroides* in the medicated subjects.

ACKNOWLEDGEMENTS

We are very grateful for the participation of all eight volunteers and their provision of samples.

CHAPTER 4

RNA isolation and mRNA enrichment from samples of the human gastrointestinal tract

Carien C.G.M. Booijink, Erwin G. Zoetendal, Eline S. Klaassens,
Michiel Kleerebezem, Hauke Smidt & Willem M. de Vos

Part of this manuscript is published in Nature Protocols 1(2); 954-959

ABSTRACT

The human gastrointestinal (GI) tract contains a complex microbial community that consists of numerous uncultured microbes. Therefore, nucleic-acid-based approaches have been introduced to study microbial diversity and activity, and these depend on the proper isolation of DNA, rRNA and mRNA. Here, we present an RNA isolation protocol followed by an mRNA enrichment method that is suitable for a wide variety of GI tract samples. One of the benefits of our overall sample processing, storage and RNA isolation protocol is that it is compatible with sampling outside the laboratory, which offers possibilities for implementation in large intervention studies. The RNA isolation method is based on mechanical disruption, followed by isolation of nucleic acids using phenol:chloroform:isoamylalcohol extraction and removal of DNA. For mRNA enrichment, two different methods were evaluated; one based on several enzymatic treatments whereas the other made use of rRNA-probes to selectively fish out the 16S and 23S rRNA molecules. Of these, the most suitable method proved to be the probe-based removal of rRNA molecules from the total RNA pool. In our laboratory, this protocol has resulted in the isolation of total RNA and mRNA of sufficient quality and quantity for microbial diversity and activity studies. Depending on the number of samples, the sample type, and the quenching procedure chosen, the whole procedure can be performed within 4.5-6 hours.

INTRODUCTION

Microbial ecosystems are found all over the world and one of the more complex communities is found in our gastrointestinal (GI) tract. This community is composed of numerous uncultured microbes, and therefore culture-independent nucleic-acid-based approaches were introduced to study microbial diversity and functionality [231, 233]. The first and most critical step in these approaches is a reliable isolation of nucleic acids from GI tract samples. In particular, the isolation of RNA is complicated, as it is easily degraded. In addition, RNA isolation procedures require immediate quenching of RNA metabolism in ecological samples to avoid *ex situ* adaptation of the RNA patterns, which is anticipated to occur very rapidly based on the high adaptation rates known in microorganisms [158, 218]. RNA is present in different forms in the cell, including ribosomal RNA (rRNA), messenger RNA (mRNA) and transfer RNA (tRNA). When total RNA is isolated, 95-99% of the RNA consists of rRNA. Bacterial rRNA is present as 16S, 23S and 5S rRNA and these molecules form the ribosomes, together with the ribosomal proteins. As rRNA is present in every cell and has a low mutation and horizontal transfer rate, it has been used as a phylogenetic marker for detection, identification and quantification of uncultured microbes from a variety of ecosystems including the human GI tract. In particular, 16S rRNA and its corresponding gene have been used, and at present more than 800,000 sequences are available in various databases [28, 152]. Compared to rRNA, mRNA is unstable: in many cases, the half-life is no longer than a few minutes. As mRNA is the result of transcription, and therefore reflects the activity of the microbes, rapid quenching of the RNA turnover in complex biological samples, combined with high quality RNA isolation procedures is of utmost importance.

An additional difficulty in studying the human GI tract is that intestinal samples cannot always be processed freshly, as sampling often occurs in the homes of volunteers or in hospitals. Therefore, we have developed and validated protocols for the isolation of RNA and subsequent mRNA enrichment from intestinal samples that are ready-to-use for RT-PCR-based diversity analysis or transcript analysis, and that can be applied to samples collected outside the laboratory. The procedure is based on quenching of the samples in RNA*later* and storage at 4°C or in methanol/HEPES buffer on dry ice [148], followed by the isolation of RNA (adapted from Fitzsimons *et al.* [52]) and the removal of DNA via DNase I digestion (adapted from the RNeasy mini kit, Qiagen). This protocol has been used successfully for faeces [11, 73, 229], ileostomy effluent [18] and tissue samples [31, 236], and the nucleic acids have served as targets for 16S rRNA- and mRNA-based approaches, which argues strongly for their use in the functional analysis of GI tract communities.

For mRNA transcript profiling purposes, total RNA is not suitable as starting material as it consists for 95-97% of rRNA [166], making an additional enrichment for mRNA necessary. For example, for transcript profiling using the RNA fingerprinting technique cDNA-AFLP (as described in chapters five and six), mRNA enrichment resulted in strongly increased recovery of mRNA derived sequences as compared to total RNA analyses. Therefore, two mRNA enrichment procedures were tested and validated on total prokaryotic RNA. The first mRNA enrichment procedure tested was the RNA Ligation Mediated Rapid Amplification of cDNA

Ends (RLM-RACE [107]), originally developed to map the 3' and 5' ends of eukaryotic (m)RNA. Additions to the procedure made use of enzymatic removal of the 5'-cap and replaced it with an anchor or r-oligo that was used to selectively amplify the molecules originating from mRNA after cDNA-synthesis [122, 180]. In these cases, the poly-A tail present at the 5'-end of eukaryotic mRNA was used for cDNA synthesis with a poly-T primer. As prokaryotic mRNA does not contain a poly-A tail, modifications to the original protocol were essential to make it suitable for prokaryotic mRNA. In short, enzymatic discrimination between bacterial mRNAs on one side, and rRNA and tRNAs on the other is possible based on differences in phosphorylation at the 5' site of these molecules [2]. Ligation of a RNA probe to the mRNA molecules makes subsequent cDNA synthesis and retrieval of fragments originating from mRNA transcripts possible by selective capture, at least in theory. A schematic overview of the procedure is shown in Fig. 4-1.

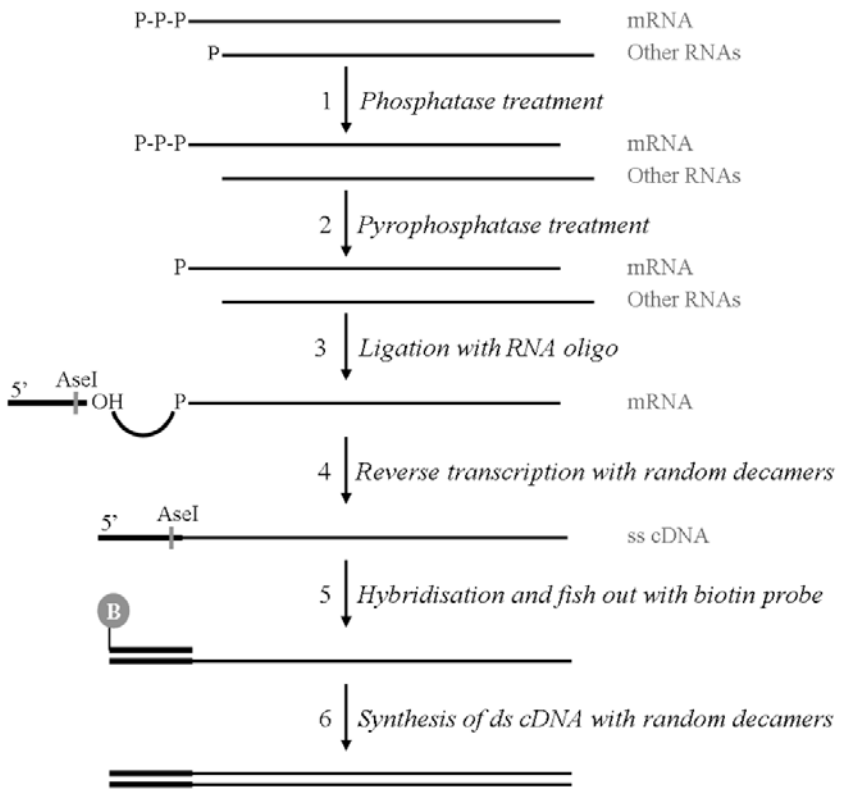


FIGURE 4-1 Outline of the mRNA enrichment procedure from total prokaryotic RNA by RLM-RACE. After phosphatase and pyrophospatase treatment, the RNA oligonucleotide with an AseI digestion site is, in theory, only ligated to mRNA molecules. Reverse transcription, followed by hybridisation with a biotin-labelled probe and fishing out the biotin with streptavidin will ultimately result in catching the cDNA of mRNA origin.

The second mRNA enrichment approach has been developed to enrich bacterial mRNA from purified total RNA by removing the 16S and 23S rRNA based on complementary probes, the so-called capture oligonucleotides. This commercial kit called MICROB*Express* (Ambion, Applied Biosystems, Niewerkerk a/d IJssel, the Netherlands) makes use of a novel modification to the conventional sandwich capture hybridisation protocols developed for the capture and detection

of specific nucleic acids [83, 134, 135]. As shown in Figure 4-2, a hybridisation sandwich between an rRNA molecule, capture oligonucleotide and oligo magnetic bead is formed. The magnetic beads allow dragging of the rRNA sandwich to one side of the tube with aid of a magnet. In the remaining liquid the mRNAs, tRNAs, 5S rRNA and other small RNAs are retained.

The result of the first method, including several enzymatic steps, is mainly depending on the efficiency of the selective ligation of the RNA oligo to only the mRNA molecules. In contrast, the enrichment efficiency of the second, probe-based, mRNA enrichment method is largely dependent on the target range of the probes used for selective removal. The success rates of both mRNA enrichment procedures are described in this chapter.

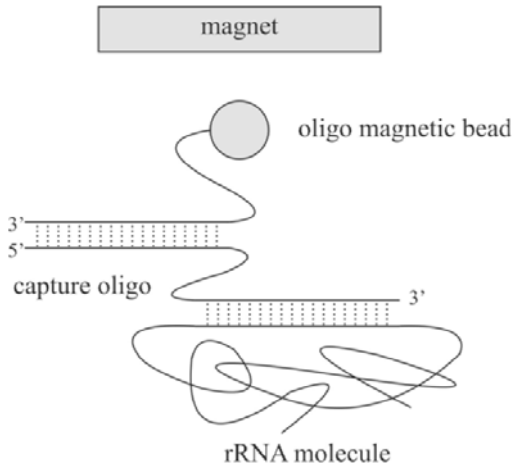


FIGURE 4-2 The hybridisation "sandwich" of oligo magnetic bead - capture oligonucleotide - rRNA molecule that is formed during the MICROB*Express* mRNA enrichment procedure. Picture adapted from MICROB*Express* instruction manual, p1.

MATERIALS & METHODS

General procedural comments

RNA is highly sensitive to degradation by RNase. Therefore, we performed all steps at 0˚C or on ice, and several steps in the hood and wore gloves all the time. RNase-free micro-centrifuge tubes and pipette filter tips were used, glassware was baked at 180˚C for 4 h prior to use, reagents were prepared with DEPC-treated water, and pipettes and benches were cleaned using RNase AWAY™ (Invitrogen, Breda, the Netherlands) prior to use.

Quenching of samples

Samples used for RNA isolation were quenched as quickly as possible after sampling, in order to stop the cellular processes that are ongoing in the bacterial cells, and to preserve the RNA that is present at the moment of sampling. For faecal samples, at least 15 g was taken, whereas for ileostomy effluent 150 g was taken as starting material. For faecal samples, the commercial quenching buffer RNA*later* was used in a ratio of 1 ml per 1 gram of faecal sample, and samples were stored at 4°C. To quench the cells in ileostomy effluent samples, four volumes of pre-

cooled methanol/HEPES (-80°C) buffer were added to one volume of sample [148]. The samples were homogenised by vigorous shaking and the samples were stored on dry ice or at -80°C until further processing with the RNA isolation.

RNA isolation procedure

RNA was isolated within one week of sample collection. To detach bacterial cells from the organic matter present in ileostomy effluent or faecal samples, the samples were sonicated in a water-based sonifier with pre-cooled water at 40 kHz for 10 min, after addition of ~15 glass beads. To separate bacterial cells from the organic matter, the samples were treated as follows: first the samples were centrifuged at 300× g at -20°C for 5 min, the supernatant was transferred to a new tube and subsequently centrifuged at 9,000× g at 4°C for 1 min. The supernatant was removed and the pellet was resuspended into 0.5 ml ice-cold TE buffer. The cell suspensions were added to a microfuge tube containing 0.18 g macaloid suspension (Kronos Titan GmbH, Leverkusen, Germany; pretreated as described before; [171]), 0.8 g zirconia beads and 50 µl SDS (10%). 500 µl acid phenol (4°C; pH 3.75) was added before treating the samples three time in a FastPrep (speed of 5.5 m s^{-1} for 45 sec) with cooling on ice for 90 sec in between. The samples were centrifuged at 13,400× g in a microcentrifuge at 4°C for 15 min. The aqueous phase was transferred in a new microfuge tube and a 25:24:1 mixture of acid phenol (pH 3.75), chloroform and isoamylalcohol was added before vortexing for a few seconds. The samples were centrifuged at 13,400× g in a microcentrifuge at 4°C for 5 min. The washing with acid phenol:chloroform:isoamylalcohol (25:24:1) was repeated followed by the centrifugation step. A final washing step with 500 µl choloroform:isoamylacohol (24:1) was performed. After vigorous shaking the samples were centrifuged at 13,400× g in a microcentrifuge at 4°C for 5 min.

DNase treatment

To remove the DNA that is co-extracted with the RNA, an on-column DNase digestion step was performed using the RNeasy mini kit (Qiagen, Venlo, the Netherlands) as described by the manufacturer with some minor modifications. To improve the efficiency of the DNA degradation, a 10-fold concentrated DNase solution was used, i.e., 10 µl of RNase-free DNase solution of 10 U μl^{-1} (Roche, Almere, the Netherlands) was added to 15 µl DNase buffer and subsequently used for on-column DNA digestion. To elute the RNA from the column, 30 µl of RNase-free milliQ water was added to the column and after incubation at room temperature for 1 min. the filter tube placed inside a clean microfuge tube was centrifuged at 9,300× g for 1 min. Another portion of 30 µl RNase-free milliQ was added onto the column and the RNA was retrieved by the same centrifugation step.

RNA quantification and qualification

To quantify the RNA yield, the spectrophotometer NanoDrop 1000 (NanoDrop Technologies, Isogen Life Science, De Meern, the Netherlands) was used. For qualitative RNA analysis, the 2100 BioAnalyzer (Agilent, Amstelveen, the Netherlands) was used. A 23S/16S rRNA ratio of 1.7 or higher was considered as indicative for sufficient quality for use in subsequent analysis. Unless direct continuation with a subsequent experiment, RNA samples were stored at -80°C

after addition of 1/10th volume 3M NaAc (4°C, pH 5.2), 1/100th volume of glycogen (5 mg ml^{-1}; Ambion) and 3 volumes of 96% ethanol (-20°C).

mRNA enrichment by RLM-RACE

The procedure followed for prokaryotic mRNA enrichment based on RLM-RACE is an adaptation of the original GeneRacer Kit (Invitrogen). An overview of the procedure is given in Figure 4-1. To evaluate the efficiency of the approach, it was applied to total RNA isolated from a culture of *Lactobacillus plantarum* WCFS1 in exponential growth phase. The different enzymatic steps, including phosphatase, pyrophosphatase and ligase treatment, were evaluated separately (as described below) as well as sequentially.

Efficiency of phosphatase treatment

The analysis of the phosphatase efficiency included the following steps: vector pUC18 was digested with HindIII (Invitrogen) and treated with different concentrations of Calf Intestinal Alkaline Phosphatase (CIAP; Fermentas Life Sciences, St. Leon-Rot, Germany). The vector was religated with T4 DNA ligase (2 U for 1½ h at 37°C; Invitrogen) and transformed to XL1-Blue competent *Escherichia coli* cells (Stratagene, Amsterdam, the Netherlands). The reduction of transformation efficiency was a measure for the efficiency of the phosphatase treatment.

Efficiency of pyrophosphatase treatment

To test the efficiency of the pyrophosphatase, an DNA oligo of 50 nucleotides was labelled with [γ^{32}P]-ATP by kinase treatment. After purification on a Qiagen Column to wash away unincorporated ATP, the labelled oligo was subjected to different concentrations of Tobacco Acid Pyrophosphatase (TAP; Epicentre, Landgraaf, the Netherlands). The product was again purified on a column and radioactivity was measured. The decrease in radioactivity was a measure for the activity of the pyrophosphatase.

mRNA enrichment by MICROBExpress (Ambion)

The procedure as described by the supplier was tested on both total RNA extracted from exponentially grown *L. plantarum* cells, as well as on faecal samples. The efficiency was analysed by analysis on a BioAnalyzer nanochip. On the website of the provider, an up-to-date list is published, indicating the compatibility of MicrobExpress with the 16S and 23S rRNA sequences of microorganisms with genome sequences and other databases (http://www.ambion.com/techlib/misc/microbe.html). Here, also the microorgansisms that are not fully compatible (removal of 50-100% of the 16S or 23S rRNA molecules) with the probes present in MicrobExpress, are indicated.

RESULTS AND DISCUSSION

RNA isolation procedure

In our laboratory, the RNA isolation procedure has been used successfully for GI tract samples. Typical concentrations and RNA are listed in Table 4-1. The RNA isolation method resulted in high-quality RNA from intestinal samples with 23S/16S rRNA ratios of approximately 2 in case of faecal samples (Fig. 4-3). This indicated that the RNA degradation was limited during the isolation procedure. Based on these quality and quantity scores, the isolated RNA can be used for subsequent 16S rRNA-based approaches but also for mRNA-based approaches.

TABLE 4-1. Range of RNA concentrations obtained during the isolation procedures.

Sample	RNA (µg g^{-1} sample)
Faeces	2.0-20.0*
Ileostomy effluent	0.1-0.2

*Data from Klaassens, 2007.

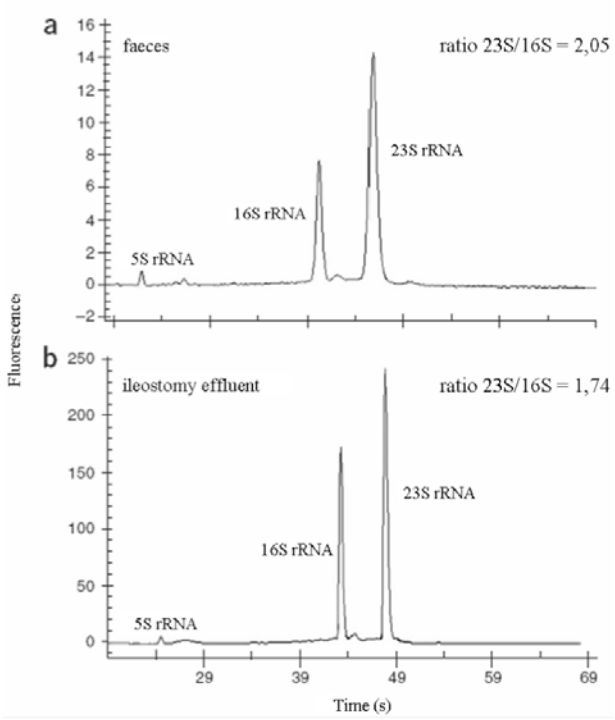


FIGURE 4-3 RNA plots showing examples of RNA that has been isolated from faecal samples quenched with RNA*later* (**a**) and ileostomy effluent quenched with methanol/HEPES buffer. (**b**) The peaks corresponding to 23S, 16S and 5S rRNA (and tRNA) are indicated. Note that the scales of fluorescence are different due to measurement on two different automated gel electrophoresis systems.

mRNA enrichment procedure based on RLM-RACE

The efficiency of the first step of the RLM-RACE enrichment, the phosphatase treatment, was tested with DNA as starting material. Reduction of the transformation efficiency of ligase-treated pUC18 plasmid after phosphatase treatment was used as a measure to determine the effectiveness of the phosphatase treatment. When a mock phosphatase treatment was performed, ~1,500 colonies were counted, including 50 white (background) colonies, indicative for a successfull ligation and transformation. Starting with the same quantity of linear pUC18 plasmid, phosphatase treatment with the recommended concentration of 0.05 U CIAP per 1 pmol DNA resulted in 5 white and no blue transformants per ml. In all cases, control experiments with pUC18 tranformation into *E.coli* showed similar transformation efficiencies. The reduction of blue colonies by 100% in case of phosphatase treatment shows that the removal of the 5'-phosphate group is very efficient.

The second enzymatic step of the RLM-RACE enrichment, the pyrophosphatase reaction, was studied with the use of a 5'-radioactively labelled DNA oligo. After pyrophosphatase treatment, no radioactivity could be detected on the oligos. This is indicative for a conversion of 5' triphosphates to mono- or diphosphates with a high efficiency.

Effects of enzymatic treatments on RNA quality

As the phosphate group(s) at the 5'-site of the mRNA and rRNA molecules might have a protective function on the RNA molecules, the effect of the removal of these groups on the RNA quality was investigated. Therefore, total RNA extracted from *Lactobacillus plantarum* WCFS1 was treated with pyrophosphatase (TAP) or pyrophosphatase and phosphatase (TAP + CIAP) prior to ligation of the RNA oligo. The quality of the treated RNA samples was compared by means of RNA plots (BioAnalyzer) to untreated total RNA (Fig. 4-4). Also total RNA after ligation of the RNA oligo without enzymatic treatment was compared. These results showed clearly that the RNA was partially degraded after TAP and TAP + CIAP treatment as can be seen by the decrease in peak height of the 23S rRNA peak with respect to the 16S rRNA peak (Fig. 4-4b). In addition, the increase of the baseline level between the 5S and 16S rRNA peaks is indicative for RNA degradation. Apparently, the phosphate group(s) at the 5'-site of the RNA molecules does(/do) have a protective effect, as removal has such a deteriorating effect on the RNA quality. In addition, mock experiments with either one of the three enzymatic steps, showed that the precipitation step in between the enzymatic treatments has a negative effect on the RNA quantity and quality (results not shown). Without successfull application of these enzymatic steps, enriched mRNA of sufficient quality cannot be obtained. Therefore, we decided to abandon this tedious enrichment procedure.

mRNA enrichment procedure based on MICROBExpress™ Kit

The second mRNA enrichment procedure, explored in this study, was the MICROB*Express* kit from Ambion. This relatively easy and straightforward procedure was tested on RNA from *Lactobacillus plantarum* WCFS1 as well as faecal RNA as starting material to verify its applicability on complex RNA samples. The original RNA samples as well as the enriched fractions were loaded on the BioAnalyzer. The effect of the enrichment was clearly visible on the gel views by a

large reduction of the strength of the 16S and 23S rRNA bands in the mRNA enriched samples compared to the original RNA samples (Fig. 4-5a). Similar results were obtained for mRNA enrichments from both *L. plantarum* RNA and faecal RNA, although the reduction of the 16S and 23S rRNA was a bit less explicit in case of the faecal RNA material, probably due to the heterogeneous mixture of rRNA molecules present in the faecal samples, compared to the homogeneous *L. plantarum* rRNAs.

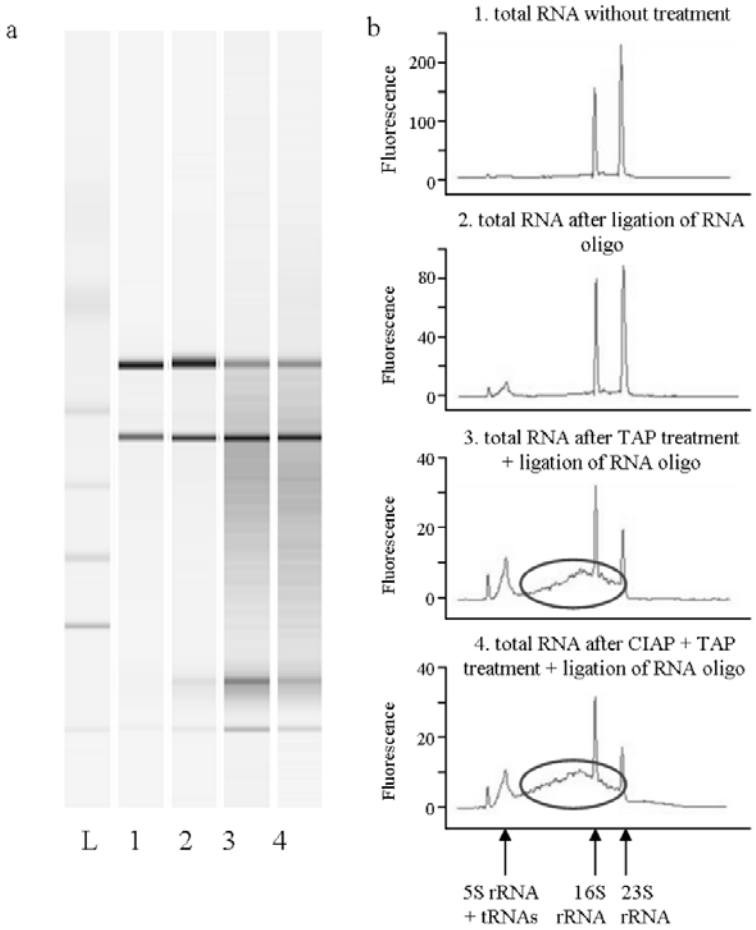


FIGURE 4-4 (**a**) Gel views of BioAnalyzer profiles of RNA isolated from *Lactobacillus plantarum* WCFS1. Different treatments were performed preceding the profiling as follows: 1 = total RNA without treatment; 2 = total RNA after ligation of the RNA oligo; 3 = total RNA after pyrophosphatase (TAP) treatment and ligation of the RNA oligo; 4 = total RNA after phosphatase (CIAP) and pyrophosphatase treatment followed by ligation of the RNA oligo. L indicates the RNA 6000 ladder. (**b**) RNA plots (BioAnalyzer) showing RNA of *L. plantarum* WCFS1 with and without treatments, corresponding to the gel views of (**a**). The amounts loaded were normalised to the original amount of RNA input. Note the decrease of the peaks corresponding to 16S and 23S rRNA as well as RNA degradation in case of TAP and CIAP treatment indicated with black ellipses (profiles 3 and 4). Furthermore, addition of an internal RNA marker is visible as an additional band at the bottom of the gel (Fig. 4-4a) and an additional peak left to the 5S rRNA peak (Fig. 4-4b).

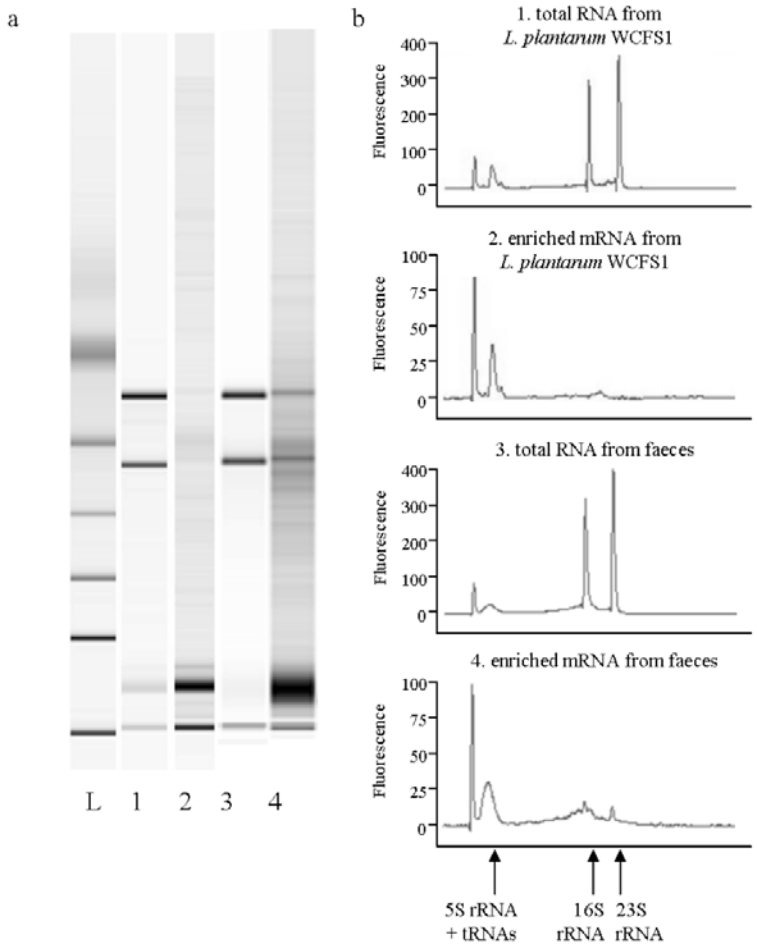


FIGURE 4-5 (**a**) Gel views of BioAnalyzer profiles of total RNA from *L. plantarum* WCFS1 (1); enriched mRNA from *L. plantarum* by MICROB*Express* (2); total RNA from a faecal sample (3); and enriched mRNA from the same faecal sample. L indicates the 6000 RNA ladder. (**b**) RNA plots (BioAnalyzer) of the total RNA and enriched mRNA by MICROB*Express* of *L. plantarum* WCFS1 and a faecal sample are shown. The peaks corresponding to the 5S + tRNAs, 16S and 23S rRNA are indicated by arrows. The peak preceding the 5S rRNA + tRNAs peak is corresponding to an internal RNA marker added to the samples prior to loading them on the BioAnalyzer, also visible at the bottom of the gel.

The removal of rRNA molecules was confirmed by the BioAnalyzer profiles, where the peaks corresponding to 16S and 23S rRNA were reduced to approximately 1/100th fraction of the signal intensities in the original *L. plantarum* RNA sample. In contrast, the peak intensity corresponding to the 5S rRNA + tRNAs was hardly reduced. This result confirms the selectivity of the procedure used, since the probe mixture of the Microb*Express* does not target 5S rRNA nor tRNA molecules (Fig. 4-5b). On average, from 10 μg total faecal RNA as starting material, 3.3 μg enriched mRNA could be obtained, within the range of 2.5-3.5 μg enriched mRNA mentioned by the supplier. This indicated that the initial contribution of mRNAs to the total RNA was increased from 5% to approximately 14-20% with this enrichment procedure, thus a three- to four-fold increase of the relative proportion of mRNA molecules.

CONCLUSION

The RNA isolation protocol described in this manuscript was successfully applied to ileostomy effluent samples, faeces and tissue samples. From the two mRNA enrichment procedures tested, the enrichment procedure based on 16S and 23S rRNA-probes turned out to be most effective and fast. This follow-up mRNA enrichment procedure was also successfully applied to ileostomy effluent samples and faeces. Nucleic acids retrieved in this way have served as targets for 16S rRNA- and mRNA-based approaches (see chapters five and six), which argues strongly for their use in the functional analysis of GI tract communities. We believe that the use of the RNA isolation procedure and the subsequent mRNA enrichment is not limited to samples from the human GI tract used in this study, but that both methods can also be applied to other GI tract samples.

ACKNOWLEDGEMENTS

The authors would like to thank Luisa Trindade, Asun Fernandez del Carmen and Christian Bachem from Wageningen University for the usefull discussions concerning the optimisation of the RLM-RACE in combination with cDNA-AFLP. Ans Geerling is also thanked for the usefull discussions about testing the efficiency of the phosphatase and pyrophosphatase treatments. Hans Heilig is thanked for all technical assistance.

CHAPTER 5

Gene expression profiling of the human GI tract microbiota

Carien C.G.M. Booijink, Jos Boekhorst, Erwin G. Zoetendal,
Hauke Smidt, Michiel Kleerebezem & Willem M. de Vos

Manuscript in preparation

ABSTRACT

The human gastrointestinal tract (GI tract) provides home to a complex microbial community, collectively termed microbiota. Although major efforts have been made to describe the diversity and stability of the microbiota, functional studies have been largely restricted to intestinal isolates and include very few community studies. The aim of this study was to explore the *in situ* gene expression of the GI tract microbiota, as well as an evaluation of the RNA fingerprinting method cDNA-AFLP (cDNA Amplified Fragment Length Polymorphism) for this purpose. To this end, cDNA-AFLP analysis of mRNA-enriched RNA extracts obtained from faecal samples of two healthy individuals collected at two different time points was performed in duplicate. Comparison of duplicate profiles based on Pearson correlation revealed that the technical reproducibility of the cDNA-AFLP profiling is very high, including profiling of duplicate mRNA enrichments (similarity indices of approximately 95%). Moreover, the microbiota of one of the two individuals was more stable in composition while the dynamics of their transcriptomes were similar. The subsequent excision and sequence determination of bands from the mRNA-enriched profiles resulted in 156 sequences of which 118 could be identified as Transcription Derived Fragment (TDF) or as rRNA gene sequence. The classification of retrieved TDFs into functional clusters based on COG (Cluster of Orthologous Genes) annotation showed that most assigned transcripts belonged to the metabolism cluster (25% and 27% of the sequences, respectively), underlining that even at the very end of the intestinal tract the microbiota is still very active. However, a large number of mRNA sequences (75% and 38% for subject 1 and 2, respectively), could not be assigned to any COG, indicating that the functions of their encoding proteins cannot be determined. This study furthermore demonstrates the suitability of cDNA-AFLP for metatranscriptome analysis of the faecal microbiota.

INTRODUCTION

The human gastrointestinal (GI) tract provides a niche for a complex community of microorganisms, commonly referred to as microbiota. Until recently, research focussing on the human GI tract microbiota was mostly aimed at describing its diversity and, to some extent, its functionality [235]. Studies that investigated the functionality of this microbiota focussed mainly on predominating species found in the large intestine and faeces, including members of the genera *Clostridium*, *Bacteroides* and *Bifidobacterium* [77, 114, 172, 210]. The few studies that aimed at investigating the functionality of a complete microbial ecosystem used metaproteomics [93, 211], or the construction and analysis of a cDNA library [149]. In the metaproteome study described by Klaassens and colleagues temporal changes in production of a protein with high similarity to bifidobacterial aldolase were found in the faeces of an unweaned infant [93]. A shotgun metaproteomics approach of extracted proteins from two faecal samples revealed that proteins with functions related to translation, carbohydrate metabolism and energy production were among the most abundantly present proteins [211]. The construction and analysis of a cDNA library is an approach at transcriptome level that can be used to investigate gene expression of mixtures of microorganisms, although so far no studies are reported that applied this approach to GI tract samples. Analysis of a cDNA library from bacterioplankton communities already showed that several functional genes of environmentally important processes could be detected [149]. Though this latter study provided usefull information about the activity of the ecosystem under study, application to far more complex bacterial ecosystems like in the GI tract of human adults with its large number of different bacterial genomes present (>1,000 species; [234]) will be far more complicated.

The last decade, the common strategy used for gene expression analysis, namely the microarray platform, has seen a rapid development towards applicability for environmental microorganisms [214]. So-called functional gene arrays (FGAs) have been developed to investigate genes encoding key enzymes representative for certain metabolic pathways of interest [106, 199, 222]. This approach has already delivered usefull information about temporal changes in gene expression of for instance biodegradation and metal resistance [162] and dechlorinating activity [32] in certain microbial communities. Nevertheless, since only differences in gene expression of already known sequences (at the time of platform construction) can be detected, application is still limited to communities for which the microbial diversity is largely known [214]. Furthermore, the application of such a mixed genomes platform to a complex ecosystem like the human GI tract would always have the drawback that sequence variation in closely related genes causes artifacts in the quantitative interpretation of the gene specific signals obtained by array analysis, like observed in the application to other microbial ecosystems [199, 222].

The RNA fingerprinting method cDNA Amplified Fragment Length Polymorphism (cDNA-AFLP) has been shown to be a very powerful tool to investigate (temporal) gene expression in several biological processes of a single specimen [7]. In terms of sensitivity, reproducibility and specificity, cDNA-AFLP results match very well with GeneChip results [159]. Relative to microarray-based approaches, cDNA-AFLP has the advantage that any unknown genome or set of genomes can be studied without prior sequence knowledge [7]. Therefore, this

technique should in principle be applicable to any microbial ecosystem. Moreover, the cDNA-AFLP approach allows the detection of lowly expressed genes and may allow the discrimination between homologous genes. So far, this method has been applied most extensively (about 100 studies) to analyse the transcription profiles in single species, mostly plants [12]. In addition, cDNA-AFLP was applied to a lesser extent to study the gene expression in mammalian cells [44, 117, 164], yeasts [159], nematodes [141], and some bacteria, mainly plant pathogens [177]. The aim of this study was to gain insight into the *in situ* gene expression of the GI tract microbiota and the evaluation of cDNA-AFLP as a potentially suitable technology for this purpose. To this end, total DNA and RNA extracts from faecal samples, as well as mRNA enriched fractions were used for (cDNA-)AFLP profiling. The technical reproducibility of the nucleotide isolation method (DNA or RNA), the mRNA enrichment and the AFLP profiling were quantified. In addition, profiles of two faecal samples from two individuals were compared to get insight into the temporal stability of the gene expression. Excision of bands from the mRNA-enriched profiles and subsequent sequence analysis was performed to investigate the functional distribution of genes expressed by the GI tract microbiota. To the best of our knowledge, this is the first study that explores the use of cDNA-AFLP to analyse the metatranscriptome of a complex microbial ecosystem such as the human faecal microbiota.

MATERIALS & METHODS

Sample collection

Fresh faecal samples were collected from two healthy volunteers (26-year old female; 33-year old male) with a time interval of three days. Samples were stored at -20 °C for DNA isolation or quenched with methanol-HEPES buffer of -80°C (±1:4 vol/vol) [148] and homogenised thoroughly for RNA isolation. Samples were processed within 24 hours after collection. These subjects had not been subjected to any dietary intervention, specific diet, or antibiotic treatment for at least one year prior to sampling.

DNA and RNA isolation, quantity and quality check

DNA was isolated from faecal samples as described by Zoetendal and colleagues [233] by using the Stool DNA Isolation Kit (Qiagen, Venlo, the Netherlands) and quantified spectrophotometrically (Nanodrop, Isogen Life Science B.V, IJsselstein, the Netherlands). Samples with an A260/280 ratio of ~2.0 were considered as pure and included for further analysis. Total RNA was isolated from fresh faecal samples quenched in methanol-HEPES using a Macaloid-based RNA isolation procedure [231]. Total RNA was treated with RNase-free DNase I (Roche, Almere, the Netherlands) (10 U DNase per 20 μg RNA) and incubated for 20 min at room temperature. The RNA was then heat-denatured at 65°C for 10 min, simultaneously inactivating DNase I. Absence of DNA was verified by using 100 ng of total RNA in a PCR reaction with primers complementary to conserved regions of the 16S ribosomal RNA (rRNA) gene. The RNA was quantified spectrophotometrically (Nanodrop) where an A260/280 ratio of ~1.8 was considered indicative for pure RNA. The quality of RNA was determined by electrophoresis (2100 Bioanalyzer, Agilent Technologies, Amstelveen, the

Netherlands). A 16S/23S rRNA ratio of 1.7 or higher was considered as indicative for sufficient quality for use in subsequent cDNA-AFLP profiling.

Adaptation of the cDNA-AFLP procedure for profiling of prokaryotic transcripts

The cDNA-AFLP method was originally developed to study gene expression in eukaryotic species. For such applications, the mRNA fraction of the total RNA can selectively be profiled by making use of the 3'-polyA tail of mRNA molecules for their reverse transcription into cDNA. Since this selective cDNA synthesis cannot be applied to prokaryotic messengers, several methods to enrich for prokaryotic mRNA were investigated. Of these, subtractive hybridisation by rRNA-probe based fishing (MICROB*Express* Bacterial mRNA enrichment kit, Ambion, Applied Biosystems, Niewerkerk a/d IJssel, the Netherlands) was found to be most efficient in enriching for mRNA from total RNA (as described in Chapter four) and was therefore, added to the original cDNA-AFLP procedure [7] prior to cDNA-synthesis (Fig. 5-1), as follows; an aliquot of 10 μg DNase-treated RNA was used for mRNA enrichment by MicrobExpress (Ambion) and enriched mRNA was quantified and its quality was checked. Since the quality of the mRNA enriched samples could not be assessed from the 16S/23S ratio, quality judgments for these samples were based on the presence of the tRNA and 5S rRNA peaks, since these molecules should remain unaffected during the mRNA enrichment procedure.

cDNA synthesis

Single stranded cDNA was synthesised using AMV reverse transcriptase (Invitrogen Life Technologies B.V., Breda, the Netherlands) and random hexamers (200 ng μl^{-1}). Double stranded cDNA was generated using protocols described by Sambrook *et al.* [171]. The double stranded cDNA was purified and concentrated in 10 μl using MinElute PCR Purification Kit (Qiagen).

cDNA-AFLP analysis

Double stranded cDNA (50 ng) was digested with TaqI and MseI or with TaqI and BfaI, respectively (New England Biolabs, Westburg, Leusden, The Netherlands) [8, 213]. The choice for these restriction enzyme combinations was based on a preliminary experiment with four different enzyme combinations in which these two resulted in profiles with ~50 clearly distinctive bands in the range of 50-500 bp. After digestion and pre-amplification, the pre-amplification mixtures were diluted to get equal concentrations for all samples, and 5 μl was used as template for the final amplifications using a ^{33}P-labelled TaqI primer (containing one or two selective nucleotides) and MseI or BfaI primer (containing two or three selective nucleotides). In total, 15 different sets of selective primer combinations were used for cDNA-AFLP analysis (supplementary table 5-S1).

Visualisation on gel, fragment isolation and sequence analysis

Polyacrylamide gels (4.5%; 7M Urea/1.0× TBE) were prepared following the manufacturers' instructions (SequaGel, AGCT Bioproducts, Hessle, UK). Amplification products (1.4 μl) were loaded to the prewarmed gel (45˚C) and Sequamark 10 bp ladder (Invitrogen Life Technologies) was loaded every 10th lane for normalisation purposes. The gel was run for ~2 h at 45˚C.

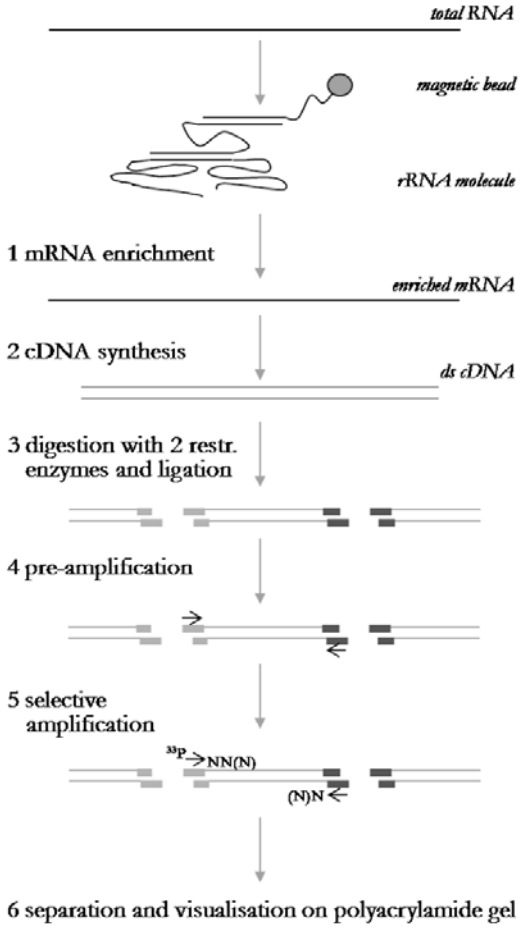


FIGURE 5-1 Schematic overview of the cDNA-AFLP procedure applied to RNA isolated from GI tract samples, adapted from Bachem *et al.* [7]. (1) Total RNA isolated from GI tract samples (faeces) is enriched for mRNA by rRNA-probe based fishing as indicated by the manufacturer (MicrobExpress, Ambion, Huntingdon, UK). (2) cDNA is synthesised by reverse transcription using a mixture of random decamers, followed by second strand synthesis. (3) The double-stranded cDNA is digested by two different four-bp cutting enzymes, followed by ligation of adaptors. The adaptors used are designed in such a way that once ligated to the sticky ends of the fragments, the sequence is changed and no longer recognised by the restriction enzymes. Therefore, restriction and ligation can be performed simultaneously. (4) Non-selective primers (indicated by arrows), which anneal to the ligated adaptors, are used to PCR pre-amplify the cDNA molecules. (5) In the selective amplification, a small aliquot of the pre-amplified fragments is used in a second PCR. These selective primers extend one, two or three bases inwards. This reduces the number of fragments by a factor 64, 256 or 1024, respectively. One of the primers contains a radioactive dye (^{33}P). (6) The fragments are size separated and visualised on a poly-acrylamide gel.

From these cDNA-AFLP gels, the lanes with the profiled enriched mRNA were selected and all well-separated bands within these lanes were excised and re-amplified as described before [159]. Re-amplification products were checked by loading 5 μl of the reaction mixture on a 1.2% agarose gel and the remaining products were purified prior to sequencing to remove excess primers and dNTPs. Sequencing reactions were performed using a Dyenamic ET Dye Terminator cycle sequencing kit and were analysed on a MegaBace automated sequencer (GE Healthcare, Diegum, Belgium). The sizes of the retrieved sequences were compared with the

sizes of the corresponding bands on the cDNA-AFLP gel, and only those that corresponded in size were taken into account for further analysis.

Bioinformatics and statistical analysis

Normalisation of the AFLP profiles was performed based on the 10 bp ladder profiles with the software package BioNumerics (Applied Maths N.V., Sint-Martens-Latem, Belgium). Subsequent statistical comparison of the technical duplicates and the temporal profiles was performed by UPGMA (Unweighted Pair Group Method, based on Arithmic mean) clustering based on pairwise Pearson correlation coefficients as incorporated in the BioNumerics software suite. The taxonomic assignment of the cDNA-AFLP sequences was performed by searching for homologous sequences in the GenBank database (including bacterial and human ST sequences, STS, GSS, environmental samples and phase 0, 1 or 2 HTGS sequences) using BLAST [3] with an e-value threshold of $1e^{-10}$ or >40% identity with protein sequences. Putative proteins encoded by the cDNA sequences were assigned to orthologous groups of the COG database [200] as described before [190]. The cDNA-AFLP sequences were also compared to the sequences obtained in the two metagenomic studies of human faecal microbiota [64, 95] with an e-value threshold of $1e^{-10}$.

Quantification of transcripts by real-time PCR

To verify whether the transcripts detected by cDNA-AFLP analysis could also be detected with another approach, and to determine to what relative level compared to total 16S rRNA copies these transcripts were expressed, real-time PCR was employed for six transcripts that were detected at both time points (t_0 and t_1). The fold-changes in the ratio of transcript versus 16S rRNAs between both time points were determined for the six chosen transcripts. These transcripts had been detected from profiles generated with the same primer selectivity (TaqI+A/MseI+GC) during cDNA-AFLP analysis, to exclude possible differences introduced by the profiling. Real-time PCRs were performed in a reaction volume of 25 µl containing 1× SYBR Green PCR Master Mix (Applied Biosystems, Foster City, USA), and 100 nM of each of the primers (MWG, Ebersberg, Germany). To make comparison of transcript levels between the temporal profiles possible, total 16S rRNA was determined with primers Bact-1369f with Prok-1492r [198]. In each reaction, 5 µl of 10× diluted cDNA synthesised with random hexamers from faecal RNA was used as template, as described before. Plasmids containing the six chosen transcript fragments were used, (in a range of 10^8 copies μl^{-1} to 10^0 copies μl^{-1} as standard curve) to determine the transcript copy number in triplicate in the faecal cDNA samples of subject 1 or 2. For total 16S rRNA determination, plasmid containing the complete 16S rRNA gene of *Lactobacillus gasseri* was used as standard in the same dilution range. Real-time PCR amplification and detection of cDNA were performed with the iQ5 (Bio-Rad Laboratories B.V., Veenendaal, the Netherlands). Data analysis was conducted with iQ5 Optical System Software Version 1.1.

RESULTS

mRNA enrichment from total RNA

The mRNA enrichment procedure based on selective removal of 16S and 23S rRNA molecules with magnetic beads (MicrobExpress) was used on the total RNA extracted from the faecal samples. The RNA samples were analysed both by conventional agarose gel- as well as by capillary electrophoresis (BioAnalyzer) before and after the enrichment procedure to determine the efficiency of the procedure and the quality of the RNA. The quality of the isolated RNA with a 16S/23S rRNA ratio of 1.7 or higher was judged as good (Fig. 5-S1). After enrichment the peaks corresponding to 16S and 23S rRNA disappeared almost completely, whereas the peak corresponding to tRNAs and 5S rRNA remained quantitatively equal as compared to its original level in the total RNA sample. This result indicates that the enrichment procedure selectively removed 16 and 23 S rRNA molecules, supporting that mRNA remained unaffected.

For identification of the genes expressed by the faecal microbiota, a subset of the dominant and well-separated bands (93 in total) from the total RNA and mRNA enriched profiles was excised and subjected to comparative sequence analysis using Blast searches against DNA, protein and EST databases. This revealed that 42% of the sequences obtained from the mRNA enriched profiles showed highest homology to sequences from mRNA origin (EST or encoded protein) versus 20% of the sequences obtained from the total RNA profiles, underlining the importance of the mRNA enrichment procedure. Subsequently, bands were excised only from the mRNA-enriched profiles to maximise the recovery of protein-encoding gene fragments of the faecal metatranscriptome.

Selection of restriction enzyme / primer combination

The optimal number of bands per lane within a cDNA-AFLP profile for subsequent profile comparison as well as excision of bands and sequence analysis, is around 50-70 [8]. To investigate, which combination of restriction enzymes and what number of selective nucleotides was needed to obtain this number of bands for faecal samples, several restriction enzymes and primer pairs were used for initial transcript profiling of faecal samples. Of the four tested enzyme combinations, MboI/TaqI, TaqI/Csp6I, TaqI/BfaI, and TaqI/MseI, the latter two showed approximately 50, more or less equally distributed and distinctive bands, and were selected as enzyme combinations of choice for further transcript profiling, whereas the other two primer combinations in general gave less than 50 bands that were unevenly distributed over the gel. The total selectivity of the primer pairs was varied between three and five additional nucleotides.

Technical reproducibility of cDNA-AFLP profiles

To analyse the fingerprint variation caused by the nucleic acid extraction procedure and subsequent (cDNA-)AFLP profiling, the Pearson-based correlation coefficients were calculated between both samples collected in time for all the primer pairs included. This revealed that for the DNA profiles, the similarity indices calculated between the DNA fingerprints of faecal samples collected at two time points ranged between 64 and 86% for all included primer

combinations (showed by gray shading in the distance tree of Fig. 5-2). This indicates that the variation introduced by the DNA extraction procedure and subsequent profiling is about 12%. Correlation analysis of the mRNA enriched fingerprints profiled with the same set of primer combinations indicated that the similarity indices ranged between 41 and 79% (showed by gray shading in the distance tree of Fig. 5-2), indicating that 38% of fingerprint variation is caused by the RNA isolation and subsequent mRNA enrichment procedure.

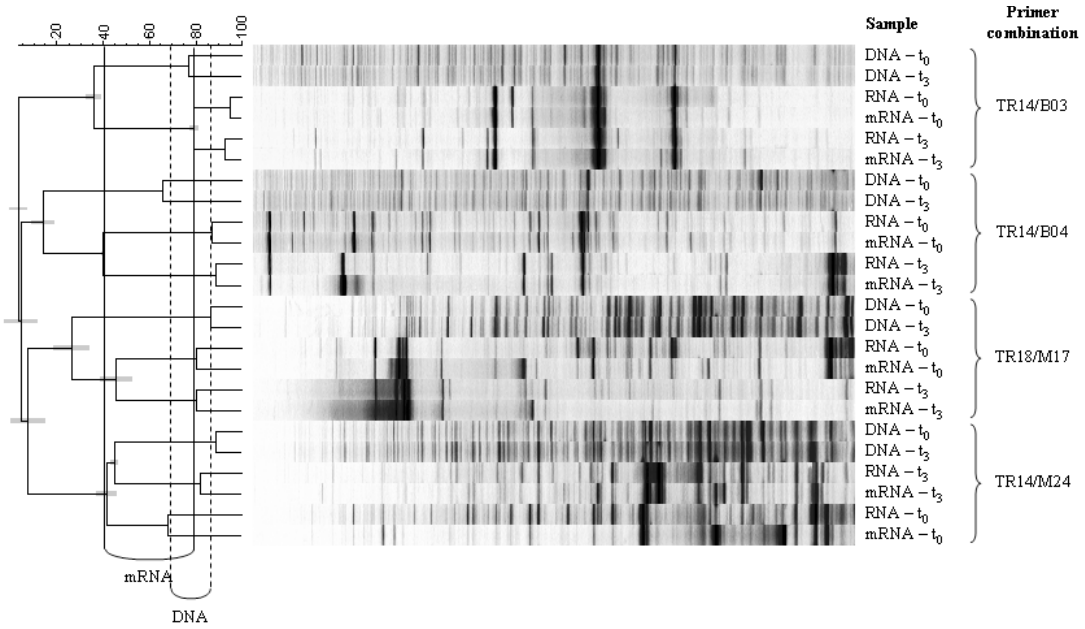


FIGURE 5-2 Cluster analysis based on UPGMA (Unweighted Pair Group Method, based on Arithmic mean) of temporal faecal (c)DNA-AFLP profiles of subject 1. DNA, total RNA as well as enriched mRNA were used as starting material. On the left, similarity indices based on Pearson correlation are indicated, whereas the gray bars indicate the error. In total six different primer combinations were used: TR14/B03 = Taq+AT/Bfa+G; TR14/B04 = Taq+AT/Bfa+T; TR14/M24 = Taq+AT/Mse+TC; TR14/M20 = Taq+AT/Mse+GC; TR18/M17 = Taq+CT/Mse+CG; TR14/M24 = Taq+AT/Mse+TC, respectively. The different time points are indicated by t_0 and t_3. The solid lines in the distance tree indicate the variation introduced by mRNA isolation (left), whereas the dashed lines indicate the variation due to DNA isolation (right).

Furthermore, the additional mRNA enrichment step in case of RNA profiles, showed an average reproducibility of 96.9 (± 1.9) for selectivity profiles generated with +2/+2 selective primers (Fig. 5-3). This indicates that less than 4% of the variation observed in mRNA enriched profiles can be attributed to the enrichment procedure, whereas the residual 34% is likely to be introduced by either the RNA extraction procedure or the steps involved in the cDNA-AFLP preparation, such as cDNA synthesis with random hexamers and (pre)-amplification. In contrast, when two profiles generated with the same primer specificity were compared at random, the average similarity was in the range of 6.3 to 19.5% for either DNA, or enriched mRNA profiles. This indicates that the reproducibility of the enrichment procedure is very high, which is essential for comparison of gene expression profiles over time.

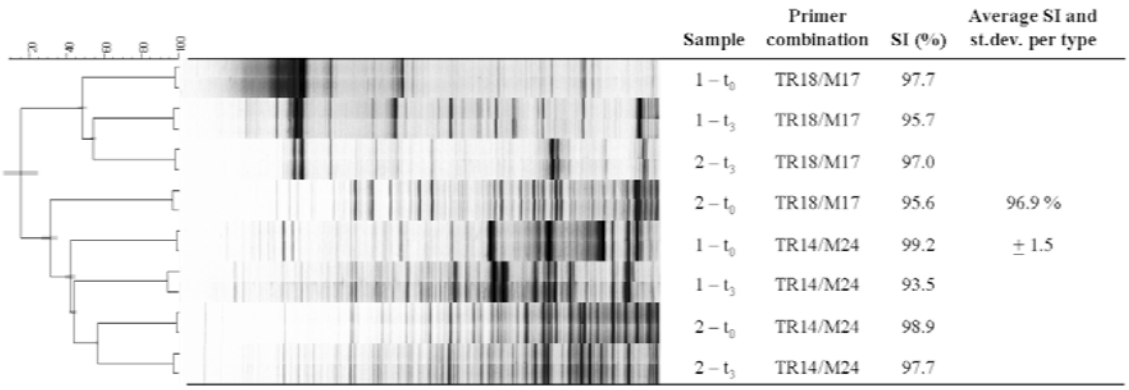


FIGURE 5-3 Cluster analysis based on UPGMA (Unweighted Pair Group Method, based on Arithmic mean) of technical profiling duplicates of faecal cDNA-AFLP profiles from subjects 1 and 2. Enriched mRNA of two time samples (t_0 and t_3) were used as starting material per subject. For cDNA-AFLP profiling of these four samples, two different primer pairs were used: TR18/M17 = Taq+CT/Mse+CG; and TR14/M24 = Taq+AT/Mse+TC. Similarity Index (SI) values based on Pearson correlation for the technical replicates are given with the overall average for +2/+2 primer selectivity. The tree on the left indicates the similarity, the error bars are indicated in gray.

Temporal stability of the metatranscriptome

To gain insight into the temporal stability of the faecal metatranscriptome two samples per subject taken with an interval of 3 days were compared. This revealed that for almost all restriction enzymes and primer combinations used, the DNA profiles clustered together, whereas the enriched mRNA profiles were more similar to the total RNA profiles of the same time point, as shown for subject 1 in Figure 5-2. Cluster analysis of the profiles of subject 2 revealed a similar sample type based clustering (enriched mRNA, total RNA and DNA, respectively) was observed, though with slightly lower similarity indices. The difference between subject 1 and 2 as reflected by the Pearson correlation coefficient, showed a higher mean similarity index for DNA profiles of subject 1 (avg. similarity index of 66.7% ± 19.3) compared to subject 2 (avg. similarity index of 30.1% ± 20.8), and this difference was statistically significant based on 1-tailed Student's t-test for both +1/+2 and +2/+2 primer selectivity ($p = 0.03$ and $p = 0.02$, respectively). As expected, profiles obtained for total RNA and mRNA enriched samples grouped more closely together than to the DNA sample. Subsequent sequence analysis of excised bands revealed that the six pairs of bands observed at the same height in the temporal profiles of subject 1 and 2 on the cDNA-AFLP gels resulted in sequences that were identical, with minor differences in sequence length.

High recovery of transcripts involved in metabolism

Distinct and well separated bands of a set of seven enriched mRNA profiles of both subjects profiled at both time points were excised from the gel and used for sequence analysis. Of the 273 excised bands, 56% resulted in high quality sequences with sizes that correlated with their position in the gel. Sequence analysis using Blast searches against DNA, protein and EST databases revealed that 9.1% and 47% of the sequences obtained from the mRNA enriched

profiles of subject 1 and 2, respectively, showed highest hits with prokaryotic protein-encoding gene sequences whereas 23% and 26% had no significant hit in the database (Fig. 5-4a). The residual 68% and 28% of the sequences obtained from the mRNA enriched profiles of subjects 1 and 2, respectively, displayed highest homology to rRNA sequences (mostly with 16S or 23S rRNA gene sequences of the genera *Bacteroides* and *Clostridium*). The functional distribution of the protein encoding TDFs into clusters of orthologous genes (COGs) is depicted in Figure 5-4b.

Based on COG distribution, most genes found to be expressed by the faecal microbiota are involved in metabolism, in particular carbohydrate and energy metabolism (Fig. 5-4b and Table 5-S2). Within this functional cluster, sequences encoding for glyceraldehyde-3-phosphate dehydrogenase and pyruvate flavodoxin/ferredoxin oxidoreductase were detected most frequently. Furthermore, response regulators, hydrogenases, SSS sodium solute transporters and hypothetical proteins were among the retrieved sequences from these faecal samples. Remarkably, many of the mRNA sequences from subject 1 encode enzymes involved in the glycolysis, whereas enzymes involved in amino acid conversion where observed in the microbiota of subject 2 (Table 5-S2).

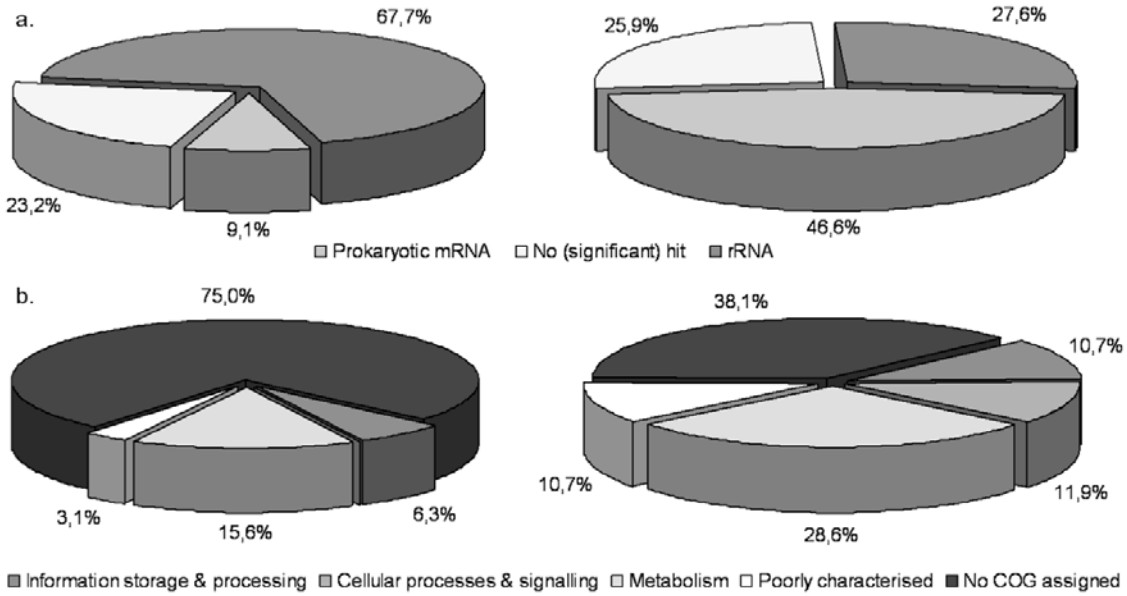


FIGURE 5-4 (**a**) Classification of the sequences retrieved from the mRNA enriched cDNA-AFLP profiles of faecal samples from person 1 and 2. Groups depicted are: prokaryotic mRNA; no or no significant similarity with sequences in GenBank; and rRNA (prokaryotic and eukaryotic). The prokaryotic mRNAs and no (significant) hit sequences are sub-classified into functional clusters in part (**b**). Blast similarity of retrieved sequences of subject 1 (32 seqs) and subject 2 (43 seqs), respectively. Classification of the sequences is based on the Cluster of Orthologous Genes (COG) as defined by Tatusov *et al.* [200].

Expression of six identified genes was verified and quantified by real-time PCR. Two investigated transcripts (2F02 with a significant hit to phosphonate ABC transporter, and 2G02 with a significant hit to an unknown protein within the phylum Proteobacteria) had transcription levels too low to be quantified by real-time PCR. Two other transcripts (1E07a with no significant hit, and 2G01 with a significant hit to asparte kinase) were expressed just above detection level. The

latter two transcripts (1E04a and 1E09a; both with no significant hits compared to GenBank) were found to be detected in numbers as high as 2.3×10^7 and 1.9×10^7 cDNA copies for transcripts 1E04a and 1E07a, respectively on t_0, whereas these levels were 1.9×10^7 and 2.0×10^7 on t_3, respectively.

DISCUSSION

In this study we evaluated cDNA-AFLP as a fingerprinting method for the characterisation of the GI tract microbiota metatranscriptome. Our study revealed individual-specific differences with respect to population composition, gene expression and their variation in time. Previous studies have shown that the cDNA-AFLP technique is very robust, sensitive and highly reproducible when applied to RNA extracted from a single species [159]. In this study it was shown that the average similarity of duplicate profiles was 95%, indicating a high level of reproducibility. This finding demonstrates that cDNA-AFLP profiling can be applied for metatranscriptome analysis of a complex ecosystem, such as the human GI tract.

Enriched mRNA, total RNA and DNA extracted from two faecal samples collected in time were profiled with (cDNA-)AFLP. From subsequent UPGMA cluster analysis of all profiles, it became clear that the DNA type samples are clustering together per primer combination as expected, whereas the mRNA enriched profiles clustered closer to the corresponding total RNA profile than to the corresponding temporal profiles for both subjects 1 and 2 (Fig. 5-4a and b). This is most likely due to the residual rRNA that is present in the enriched mRNA extracts. This is supported by the efficiency of rRNA removal, which was approximately 50% (the retrieval of mRNA sequences increased from 20% to 42% upon mRNA enrichment). Thus, the residual rRNA will still contribute significantly to the enriched mRNA profiles, therefore resulting in clustering more closely to the RNA profiles than the corresponding time points. Profiling of both DNA and RNA revealed individual specific differences, which is in line with previous observations based on 16S rRNA gene diversity [230]. Remarkably, the temporal variation with respect to DNA and mRNA based profiles was different between the two individuals. The DNA profiles were more stable in time than the mRNA profiles in subject 1 (avg. similarity between temporal profiles was 67% for DNA and 43% for mRNA enriched profiles) which indicates that the microbiota composition is more stable in time, compared to its fluctuating activity. In contrast the DNA and mRNA profiles of subject 2 were fluctuating similarly in time (similarities for DNA and enriched mRNA profiles were 43% and 40%, respectively), which might indicate that both the microbiota composition and activity are varying over time. Although the difference between subject 1 and 2 was found to be statistically significant, only a fraction of the total diversity in DNA or RNA sequences was profiled and compared due to the selective amplification step in the (cDNA-)AFLP profiling ($1/64^{th}$ of the total metatranscriptome is profiled per each of the six +1/+2 selective primer combinations used). To study the dynamics of the total metatranscriptome, it is therefore recommended to include the total set of 64 different +1/+2 selective primers, although smaller subsets can already indicate trends as was shown here.

The success rate of band excision and sequence determination of all selected profiles was on average 57% for the enriched mRNA profiles of faecal origin. For the +2/+3 primer combinations the success rate was on average higher than for the +2/+2 primer combinations, most likely due to interference of weak transcript signals in the background of the profiles. Blast analysis of the retrieved sequences allowed the functional annotation of 51% of the expressed genes, even though the majority of sequences did not show a significant hit with any of the available sequences in GenBank. This relatively high percentage of COG-assigned genes is comparable to functional assignments of sequences retrieved from faecal metagenomes (on average 48.1% assigned COGs to sequences obtained from 13 Japanese subjects [95], and an average of 54.5% assigned COGs to the sequences retrieved from two American subjects [64]. From the annotated sequences, the largest portion belonged to genes encoding enzymes involved in carbohydrate metabolism (15% and 11%, respectively for subject 1 and 2). Remarkably, genes encoding enzymes involved in the glycolysis were more prominently expressed in subject 1, while the metatranscriptome of subject 2 contained transcripts encoding proteins involved in amino acid conversions. This could indicate that the GI tract microbiota in subject 1 is involved in the fermentation of carbohydrates, while protein fermentation is more prominent in subject 2. However, this remains speculative, since only a fraction of the total faecal metatranscriptome of two individuals was studied.

Comparison of the cDNA-AFLP sequences with those obtained in the framework of two other recently published faecal metagenomes [64, 95] showed that on average about 51% of the bacterial mRNA sequences showed significant identity of >90% to sequences present in the Japanese metagenomic library (containing ~479 Mbp) whereas this was only 19% with sequences of the American library (containing ~145 Mbp; results not shown). This 2.5-fold difference is most likely explained by the size difference between the two metagenomic datasets and the far from saturated retrieval of novel sequences by metagenome analysis. Moreover, when the distribution of sequences of subjects 1 and 2 into COG clusters is compared with the faecal metagenome of two Americans [64] the most striking difference was the absence of transcripts involved in lipid, amino acid and vitamin metabolism in subject 1 and the relatively small group of transcripts in this cluster for subject 2 (11%) while this functional cluster has been reported to make-up about 25% of the American metagenomes [64]. Although the microbiotas of different subjects are compared in this case, this could give an indication that the clusters of genes involved in lipid, amino acid and vitamin metabolism are expressed in relatively low abundance in both subjects that were studied in this experiment.

In conclusion, this study illustrates the suitability of cDNA-AFLP as a technology to study the metatranscriptome of the GI tract microbiota in healthy subjects. It enables assessment of the metatranscriptome dynamics, and can provide novel insights in spatio-temporal variations in microbiota functionality. When combined with other functional metagenomic approaches, this technology can eventually be employed to achieve a better understanding of the ecology, and functionality of the human GI tract microbiota in health and disease.

ACKNOWLEDGEMENTS

Christian Bachem, Asun Fernandez del Carmen and Luisa Trindade are thanked for their usefull suggestions for modifications of the cDNA-AFLP procedure for whole community analysis. We are very grateful to Michiel Wels for performing part of the bioinformatics analysis. Finally, the authors would like to thank Hanneke Witsenboer and Jenny Peters from Keygene in Wageningen for cDNA-AFLP analysis.

CHAPTER 6

Metatranscriptomics approach to study the functionality of the human ileostomy effluent microbiota

Carien C.G.M. Booijink, Sahar El-Aidy, Jos Boekhorst, Hauke Smidt,
Willem M. de Vos, Erwin G. Zoetendal & Michiel Kleerebezem

Submitted for publication

ABSTRACT

The metatranscriptome of the largely unexplored human ileostomy effluent microbiota was investigated with the RNA fingerprinting method cDNA-AFLP. Ileostomy effluent of one subject was collected on two consecutive days. From the two ileal microbial metatranscriptome profiles obtained, 95 prokaryotic transcripts could be identified by sequence analysis and annotation, mostly within the metabolism COG-cluster (25%). Especially transcripts involved in carbohydrate uptake and degradation were abundant. Of all protein-encoding transcripts with significant sequence similarity to public database entries, 70% had the highest similarity with genes of the genus *Streptococcus*. The stability of the predominating species of the microbiota composition on the two consecutive days was limited (similarity index below 60%), while a total of five protein-encoding transcripts could be detected at both time points. The majority of retrieved transcript sequences seem to be commonly present in the microbiota of the small intestine of humans, since 15 of a total of 16 transcript sequences detected in one person, could also be detected in DNA samples of ileostomy effluent of at least one other ileostomy subject. Furthermore, in two cases multiple protein-encoding transcripts appeared to be functionally related as well as genetically linked in several streptococcal genome sequences in the public databases indicative for a coherent expression of functional pathways by the luminal streptococci. Overall, these data indicate that the genus *Streptococcus* is not only numerically abundant, but also predominates randomly generated metabolic activity profiles of the microbial ecosystem of the ileostomy effluent microbiota. Important functions exerted are related to metabolism, especially carbohydrate metabolism and transport. The fast transit of the ileal contents appears to generate an environment in which the capacity to rapidly metabolise available carbohydrates is a predominant selective advantage.

INTRODUCTION

The human gastrointestinal tract (GI tract) consists of several physiologically distinct parts. One of the regions of the GI tract is the small intestine, which includes duodenum, jejunum and ileum, and is located between the stomach and the large intestine. It is this location that seriously hampers *in vivo* sampling, and consequently studies aimed at describing the microbial diversity or functionality of the human ileum are very limited [18]. One way to circumvent sampling difficulties is to use subjects with an ileostomy as a model system. An ileostomy subject underwent surgical removal of the colon, and the terminal part of the ileum is connected to an ileostomy in the abdominal wall [89]. The contents of the ileum of an ileostomy subject flows into the appliance due to the peristaltic movement of the GI tract and can be sampled as ileostomy effluent in a non-invasive way. Sample collection of ileal luminal content is therefore relatively straightforward and can be performed at multiple sequential time points, thereby generating an attractive model system for the study of the human ileal microbiota.

The physiology and physico-chemical conditions of the lower small intestine can be anticipated to play an important role in shaping the endogenous microbiota. For example, local nutrient availability probably undergoes major fluctuations as a consequence of the variable dietary intake by the host, while the fluctuating pH will also have a continuous impact on the microorganisms residing in this niche [46, 176]. The changing availability of dietary components in the small intestinal lumen is likely to demand a more generalist genome capacity from the species present there [227]. At the same time, the presence of a mucus layer ensures a carbohydrate source for some specialist genera, such as *Bacteroides*, which have been detected in the ileal mucus layer [216, 217].

As described in Chapter three, species belonging to the genera *Streptococcus* and *Veillonella* were found to be abundantly present in ileostomy effluent samples of five healthy ileostomy subjects. Both groups have been detected higher up in the gastrointestinal tract before [188]. However, the exact role of these microorganisms in the ileum is largely unknown. The more than 40 cultured representatives belonging to the genus *Streptococcus* have an efficient usage of carbohydrates in common due to their wide range of enzymes involved in carbohydrate uptake and degradation [129]. This effective and broad spectrum capacity to harvest available nutrient sources might explain their relative dominance in the luminal part of the small intestine, where the flow and concentration of carbohydrates is high [18]. In contrast, the ten isolates within the genus *Veillonella* show hardly any enzyme activity related to sugar utilisation, nor extracellular glycosidic enzyme activity *in vitro* [5, 26, 86, 127]. However, *Veillonella* are able to ferment lactate to acetate and propionate [55], and this fermentative ability may explain their relative abundance in the ileostomy effluent, since lactate can be predicted to be abundantly present due to the carbohydrate fermentation by the streptococci.

In this study we investigated the genes expressed by the ileostomy effluent microbes and their functions. Two ileostomy effluent samples were collected from a single ileostomy subject at separate moments and microbiota transcripts were profiled by the RNA fingerprinting method cDNA-AFLP (cDNA Amplified Fragment Length Polymorphism). Since this method is independent of prior sequence information it is applicable to diverse microbial ecosystems, as

was shown with the profiling of faecal samples (see also Chapter five). Sequence analysis of a subset of mRNA derived cDNA-AFLP fragments was used to identify the expressed genes and to provide a starting point for the functional description of the ileal microbiota.

MATERIALS & METHODS

Sample collection

A healthy subject with an ileostomy (72-year old male; subject i4) participated in this study. This subject had not been subjected to any dietary intervention, specific diet, or antibiotic treatment for the last year. Apart from absence of the colon, he had no known abnormalities of the digestive system. The subject was informed about the aims and the protocol of the study orally and in written form and he signed a written informed consent before participation. The study was approved by the Universtiy Hospital Maastricht Ethical Committee. The subject collected the bulk of the ileostomy effluent in an unused appliance in the morning on two consecutive days. Directly following collection, the sample was split in two equal portions; one portion was quenched with methanol-HEPES buffer on dry ice (±1:4) [148] and homogenised thoroughly for RNA isolation and one portion was immediately frozen on dry ice for DNA isolation. Samples were processed within four days after collection. The subsequent analysis of these collected samples is outlined in Figure 6-S1.

DNA and RNA isolation, quantity and quality check

DNA was isolated from the ileostomy effluent samples of subject i4 as recently described by Zoetendal and colleagues [233] by using the Stool DNA Isolation Kit (Qiagen, Venlo, the Netherlands). Quantification was subsequently performed spectrophotometrically (Nanodrop, Isogen Life Science B.V, IJsselstein, the Netherlands). An A260/280 ratio of ~2.0 was considered as pure DNA, and included for further analysis. Total RNA was isolated from fresh ileostomy effluent samples quenched in methanol-HEPES using a Macaloid-based RNA isolation procedure [231]. Total RNA was treated with RNase-free DNase I (Roche, Almere, the Netherlands) (10 U DNase per 20 μg RNA) and incubated at room temperature for 20 min. The RNA was then heat-denatured at 65˚C for 10 min, simultaneously inactivating DNase I. To verify if all DNA was degraded, 100 ng of total RNA was used in a PCR reaction with primers complementary to conserved regions of the 16S ribosomal RNA (rRNA) gene. The RNA was quantified spectrophotometrically (Nanodrop) where an A260/280 ratio of ~1.8 was considered pure, and qualified by electrophoresis (2100 Bioanalyzer, Agilent Technologies, Amstelveen, the Netherlands).

Enrichment of mRNA from total RNA

The mRNA enrichment procedure based on selective retraction of 16S and 23S rRNA molecules with magnetic beads (MicrobExpress) was used on the total RNA extracted from the ileostomy effluent samples. The RNA samples were analysed both by electrophoresis and by BioAnalyzer before and after the enrichment procedure to determine the efficiency of the procedure and the quality of the RNA. Enriched samples that showed considerably reduced peaks corresponding to

16S and 23S rRNA a but equal peak size corresponding to tRNAs and 5S rRNA were used for subsequent analysis (details of the method and its results are described in Chapter four).

PCR-DGGE

To investigate the diversity and temporal stability of the dominating microorganisms in the two ileostomy effluent samples of one individual, PCR Denaturing Gradient Gel Electrophoresis (PCR-DGGE) was used. A ten-fold dilution of DNA isolated from the two ileostomy effluent samples was used as a template to amplify the V6 to V8 regions of 16S rRNA gene with primers U-968-GC-f and L-1401r [143]. The integrity of the nucleic acids was checked visually after electrophoresis of 5 μl of PCR products on a 1.2% agarose gel containing ethidium bromide.

PCR amplicons were subsequently separated by DGGE [73] using the Decode system (Bio-Rad Laboratories, Hercules, California, USA) with some minor modifications. In short, the polyacrylamide gels (8% [vol/vol] polyacrylamide (ratio of acrylamide-bisacrylamide, 37.5:1)) were made with denaturing gradients ranging from 30 to 50% to separate the generated amplicons of the total bacteria community. The gel was stained after completion of electrophoresis with $AgNO_3$ and developed [173]. DGGE gels were scanned at 400 dpi after overnight drying at 60˚C, and the statistical software package BioNumerics 3.5 (Applied-Maths, Sint-Martens-Latum, Belgium) was used to compare the DGGE profiles. Similarity indices (SI's) of the compared profiles were calculated based on the densitometric curves of the scanned DGGE profiles by using the Pearsons product-moment correlation coefficient [69].

cDNA synthesis and quality control

Single stranded cDNA was synthesised with AMV reverse transcriptase (Invitrogen Life Technologies B.V., Breda, the Netherlands) and random hexamers (200 ng μl^{-1}). Double stranded cDNA was generated with protocols described by Sambrook *et al.* [171]. The double stranded cDNA was purified and concentrated in 10 μl using MinElute PCR Purification Kit (Qiagen).

cDNA-AFLP analysis

Double stranded cDNA (50 ng) was digested with TaqI and MseI or with TaqI and BfaI, respectively (New England Biolabs, Westburg, Leusden, The Netherlands) [8, 213]. After digestion and pre-amplification, the pre-amplification mixtures were diluted and 5 μl was used as template for the final amplifications using a ^{33}P-labelled TaqI primer containing one or two selective nucleotides and MseI or BfaI primer containing either two or three selective nucleotides. In total, 25 different sets of selective primer combinations were used for cDNA-AFLP analysis (Table 6-S1).

Visualisation on gel, fragment isolation and sequence analysis

Polyacrylamide gels (4.5%; 7M Urea/1.0×TBE) were prepared following the manufacturers' instructions (SequaGel, AGCT Bioproducts, Hessle, UK). Amplification products (1.4 μl) of total RNA and enriched mRNA in duplicate were loaded to the prewarmed gel (45˚C). Sequamark 10 bp ladder (Invitrogen Life Technologies) was loaded every 10th lane for normalisation purposes. Duplicate cDNA-AFLP profiles of the enriched mRNA extracted from ileostomy

effluent samples collected in time were run on the same gel, to make temporal comparisons possible. The gel was run at 45˚C for approximately 2 h. From these cDNA-AFLP gels, the lanes with the profiled enriched mRNA were selected and all free laying bands within these lanes were excised and re-amplified as described before [159]. Re-amplification products were checked by loading 5 µl of the reaction mixture on a 1.2% agarose gel. The remaining products were purified prior to sequencing to remove excess primers and dNTPs. Sequencing reactions were performed using a Dyenamic ET Dye Terminator cycle sequencing kit and were analysed on a MegaBace automated sequencer (GE Healthcare, Diegum, Belgium). The sizes of the retrieved sequences were compared to those of the corresponding bands on the cDNA-AFLP gel, and only those that corresponded in length were taken into account for further analysis.

Bioinformatics and statistical analysis

Normalisation of the AFLP profiles was performed based on the 10 bp ladder profiles with the software package BioNumerics (Applied Maths N.V., Sint-Martens-Latem, Belgium) and adjusted manually if necessary. Subsequent statistical comparison of the technical duplicates and the temporal profiles based on UPGMA (Unweighted Pair Group Method, based on Arithmic mean) and Pearson correlation was performed to calculate similarity indices as implemented in BioNumerics. The taxonomic assignment of the cDNA-AFLP sequences was performed by searching for homologous sequences in the GenBank database (including bacterial and human ST sequences, STS, GSS, environmental samples and phase 0, 1 or 2 HTGS sequences) using BLAST [3] with an e-value threshold of $1e^{-10}$ or ≥40% similarity with protein sequences. Putative proteins encoded by the cDNA sequences were assigned to orthologous groups of the COG database [200] as described before [190].

Detection of identified transcripts in the microbiomes of other GI tract samples

To investigate whether the sequences obtained with transcript profiling are also present in the microbiome of other subjects with an ileostomy, PCR analysis was performed. A set of 14 primer-pairs corresponding to sequences longer than 100 bp and distributed over the different functional clusters, including four transcripts without assigned functional cluster, was used for PCR amplification (Table 6-S3). The annealing temperature during PCR amplification was in the range of 56-58°C for the chosen primers with product sizes in the range of 140-211 bp. DNA isolated from previously collected samples that were obtained from five healthy unmedicated ileostomy subjects (i1 to i5), two ileostomy subjects that were on medication (m1 and m2), a single ileorectal pouch subject (p1; see Chapter three) and faecal samples of two healthy adults (f1 and f2; see Chapter five) were used as template for PCR amplification.

RESULTS AND DISCUSSION

Temporal stability of the microbiota

To investigate the temporal stability of the dominating microbiota in the ileostomy effluent, PCR-DGGE of the 16S rRNA genes was used and compared to the AFLP profiles generated from the extracted DNA from the ileostomy effluent samples. The microbial profiles of the

ileostomy effluent samples were relatively simple (Fig. 6-1) as expected based on findings in Chapter three. The similarity between the two ileostomy effluent microbiota profiles as visualised by DGGE and AFLP was calculated based on the Pearson correlation. This coefficient was 51,4% for the ileostomy effluent DGGE profiles, whereas the average correlation coefficient of the AFLP profiles from the same samples was 55,6% ±10,4 irrespective of the number of selective nucleotides used (Fig. 6-1).

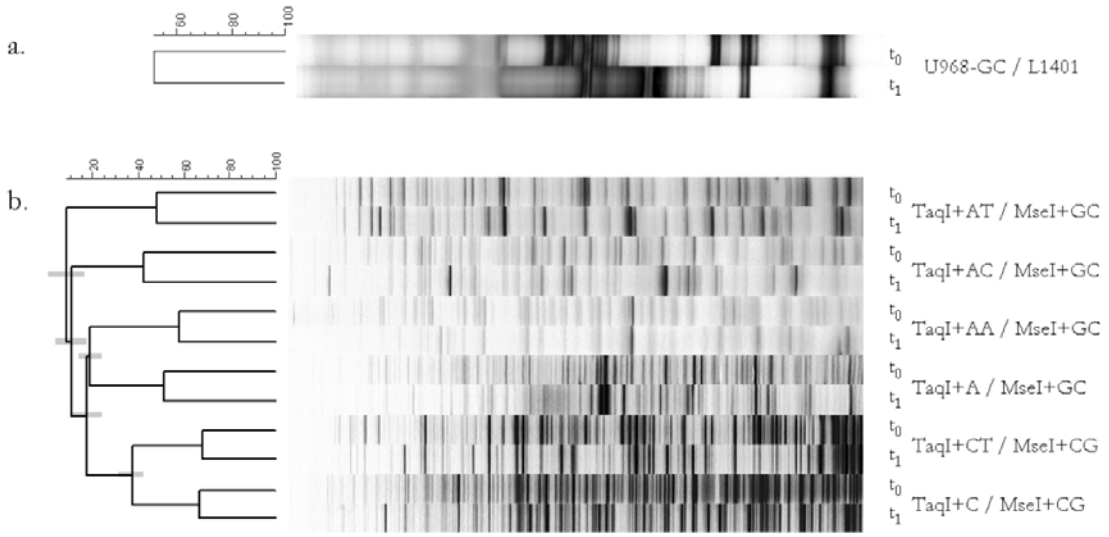


FIGURE 6-1 (**a**) DGGE profiles of the V6 to V8 region of 16S rRNA genes of two ileostomy effluent samples collected at day 1 and 2, reflecting the predominant bacterial community. The percentage similarity between the two profiles based on the Pearson correlation coefficient with UPGMA (Unweighted Pair Group Method, based on Arithmic mean) are indicated in the dendrogram. (**b**) AFLP profiles of DNA isolated from the same ileostomy effluent samples. After digestion with TaqI and MseI, six different selective primer combinations were used, as indicated. Cluster analysis is based on Pearson product-moment correlation with UPGMA whereas the gray bars indicate the error.

This indicates that about half of the predominating species can be detected on two consecutive days which is in good agreement with the observed dynamics of the ileostomy effluent microbiota composition described in Chapter three. Furthermore, it shows that AFLP profiling can be used as an alternative to DGGE to investigate the temporal stability of the microbiota.

cDNA-AFLP profile generation, comparison and transcript analysis

The metatranscriptome of the ileostomy effluent microbiota was investigated using cDNA-AFLP fingerprinting as described in Chapter five. In the metatranscriptomic profiles derived from enriched mRNA, the bands corresponding to partial transcripts were well separated making subsequent band excision and sequence analysis possible (Fig. 6-2). From the enriched mRNA profiles, a total of 230 bands were excised in duplicate, resulting in 155 sequences. Of these, 96 could be classified as bacterial transcripts based on sequence similarity of the encoded proteins (≥ 40% identity) with protein sequences in GenBank [3]. The database entries recovered as highest hit by comparison with the cDNA-AFLP sequences and their corresponding phylogeny

were analysed and are listed in Table 6-S2. The majority of the detected transcripts appeared to correspond with species within the Firmicutes phylum (81.1%), within the orders Lactobacillales (72.7%), Bacillales (4.7%) and Clostridiales (3.7%).

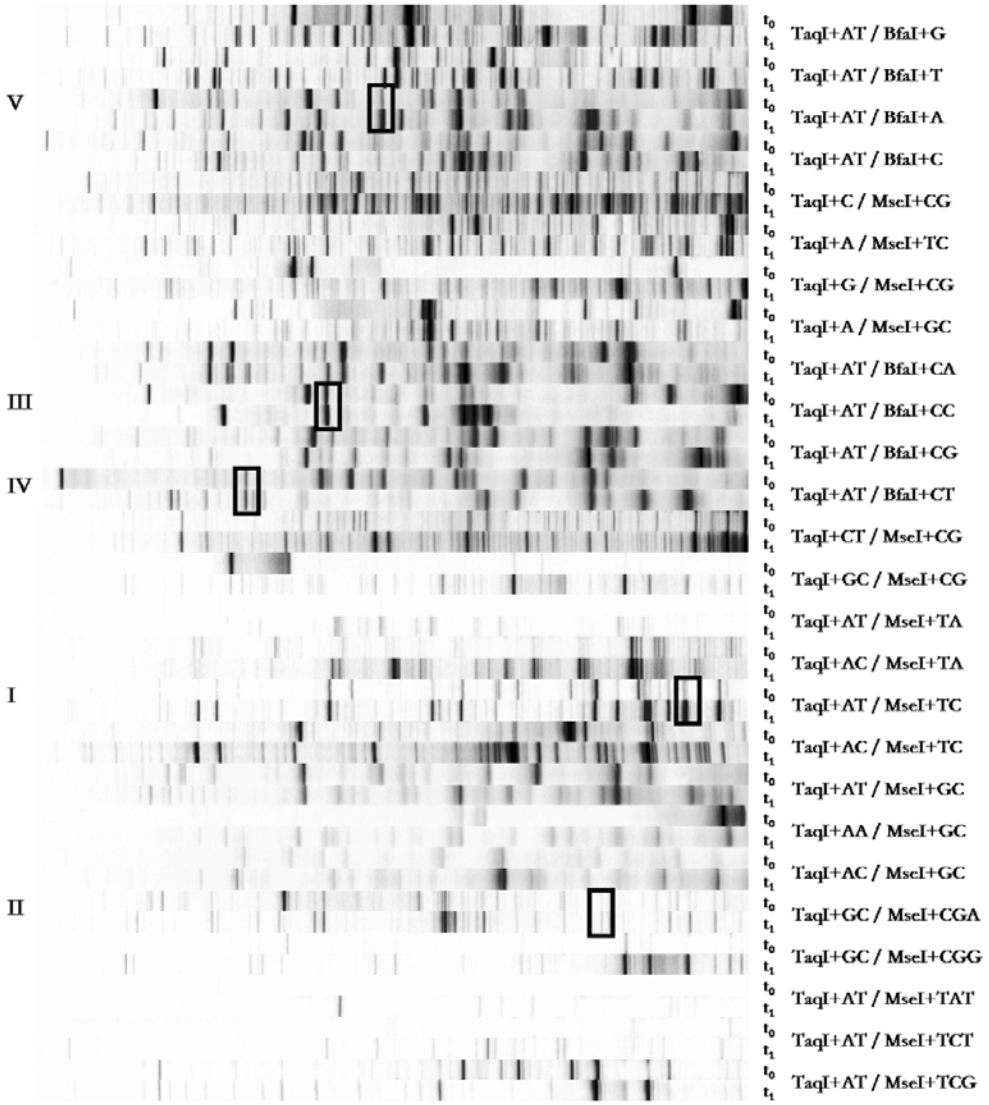


FIGURE 6-2 Overview of all temporal cDNA-AFLP profiles, using enriched mRNA from ileostomy effluent samples as starting material. After digestion with TaqI/BfaI or TaqI/MseI different selective primer combinations were used. Black boxes indicate transcript sequences (I to V) that could be retrieved at both time points. Roman numerals I to V represent transcripts with highest sequence similarity to the following encoded proteins: I = hypothetical protein; II = putative receptor; III = conserved hypothetical protein; IV = ribosomal subunit interface protein; and V = sugar hydrolase.

The residual 19% of the detected transcripts displayed the highest similarity with species within the phyla Proteobacteria (15%), Bacteroidetes (3%) and Verrucomicrobia (1%). Next to the identified bacterial gene transcripts, 21 sequences of the 156 generated did not show significant similarity with either DNA or protein sequences present in the public databases, while

39 sequences displayed high similarity with 16S or 23S rRNA genes. The majority of the latter sequences (29 sequences) displayed high level (≥ 97%) similarity with known rRNA genes within the phylum Firmicutes. Many of the rRNA derived sequences within the Firmicutes phylum had highest similarity scores with 16S rRNA genes of *Streptococcus* species (13 sequences), suggesting that this genus might be among the metabolically most active in the ileal lumen. However, since these analyses included removal of rRNA by MicrobExpress, this indicates that this rRNA probe removal was sub-optimal for species of the genus *Streptococcus*. Furthermore, also the phylogenetic distribution of protein-encoding transcripts supports a major role of the order Lactobacillales (see also below) in the meta-transcriptome of the ileostomy effluent, indicating that these groups of microorganisms play a predominant role in the microbiota activity in the ileal lumen.

Functional groups

The 96 bacterial, protein-encoding transcript sequences that were retrieved were investigated to determine their distribution among the functional clusters as defined by COG (Cluster of Orthologous Genes) classification. This analysis indicated that a large proportion (36.2%) of retrieved transcripts could not be classified to a COG, due to their short sequence length. Remarkably, the majority of sequences belonged to the functional cluster of metabolism (24.6% of all transcripts); especially gene sequences involved in carbohydrate metabolism and transport were retrieved frequently (38% of the metabolism cluster) (Fig. 6-3). Comparison of the transcripts retrieved at t_0 to t_1 revealed a similar distribution over the functional clusters. Within the carbohydrate metabolism and transport cluster, transcripts encoding enzymes belonging to the classes hydrolase, transferase and isomerase were detected most frequently, most of these are predicted to be involved in the pentose phosphate pathway, and metabolism of sugars such as galactose, fructose and starch.

The functional distribution of the retrieved transcripts was compared to the COG classification of almost 20,000 end read sequences obtained from a metagenomic fosmid library that was constructed using DNA extracted from effluent samples of the same subject with an ileostomy (i4). This analysis indicated that the genes involved in the expected functional categories, including transcription and translation, were expressed at comparable level to their abundance at metagenome level. Furthermore, the cellular processes and signalling cluster showed a comparable distribution in metatranscriptome and metagenome for all detected groups (the clusters of cell cycle control, cell motility and intracellular trafficking were not detected in the metatranscriptome in contrast to the metagenome). However, an over-representation of genes related to carbohydrate metabolism and transport in the metatranscriptome in comparison to the metagenome was found (38% of the metabolism cluster for the transcripts vs. 20% for the metagenomic sequences), suggesting that this COG is highly expressed (results not shown).

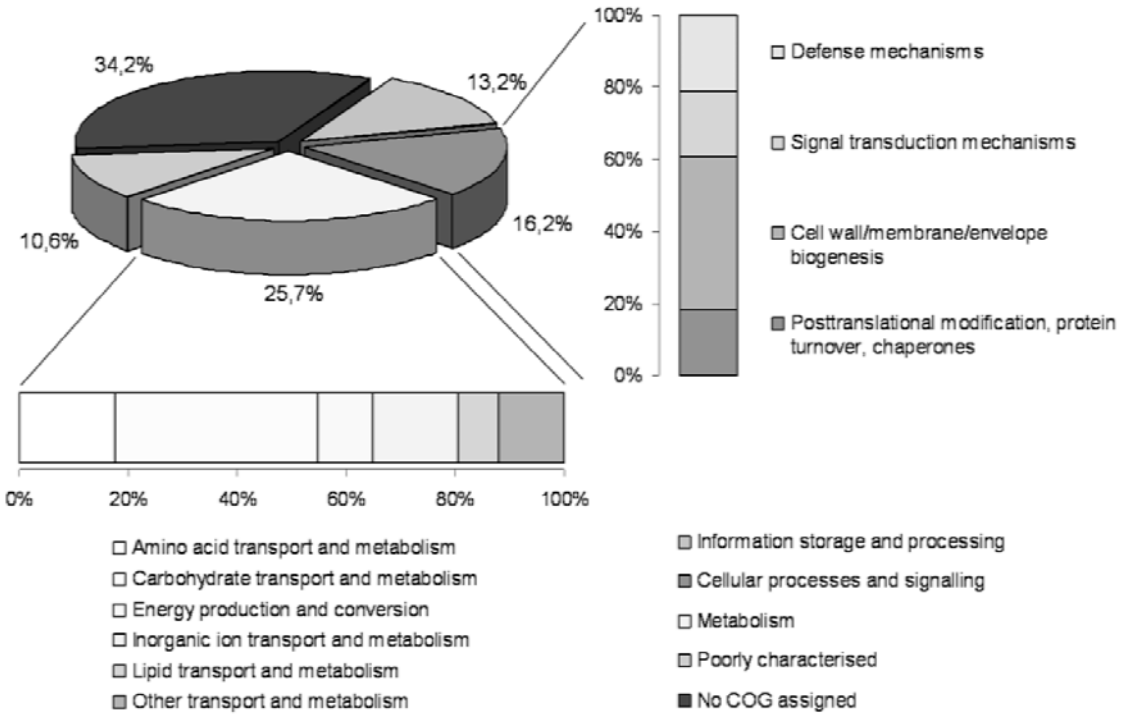


FIGURE 6-3 Classification of the sequences retrieved from the mRNA enriched cDNA-AFLP profiles of ileostomy samples from an ileostomy subject into Clusters of Orthologous Genes (COG) as defined by Tatusov *et al.* [200]. Different colours represent the following clusters: green = information storage and processing, orange = cellular processes and signalling, yellow = metabolism, blue = poorly characterised, and black = no functional COG assigned. A colour version of this picture can be found in de supplementary and colour figures section at the end of this thesis.

The functional annotation as well as the phylogenetic analysis of the transcripts based on sequence similarity is limited due to the relatively short length of the sequences. The metagenomic library offers extensive possibilities for further functional annotation of the transcripts, especially for those without functional assignment. Retrieval of these sequences on inserts of metagenome clones can give valuable additive information about the genetic context. Remarkably, the best-hits of three of the transcript sequences (7G12, 8B10b and 8B11b) appeared to be functionally related as well as genetically linked in several streptococcal genome sequences in the public databases (including *S. pyogenes* and *S. mutans*), which clearly supports the coherent expression of functional pathways by the luminal streptococci. Moreover, this notion is supported by preliminary analyses that trace these three transcripts back to a single metagenomic fosmid clone, which confirms the genetic linkage of these transcripts in the microbiota representative responsible for their expression.

Almost all transcripts that were classified within the carbohydrate metabolism cluster shared significant similarity with sequences from species within the order Lactobacillales (i.e. sugar hydrolases and epimerases, transketolase, phospho-sugar mutase, and aminopeptidase; see Table 6-S2). In contrast, faecal metatranscriptomes of two healthy idividuals indicated hardly any activity of species within the genus *Streptococcus* (Fig. 6-S2). Streptococci and other lactic acid

bacteria, belonging to the order Lactobacillales, are organisms that are well known for their rapid fermentative conversion of carbohydrates, which apparently provides a dominant selective advantage under the circumstances characteristic for the small intestine. These findings conceptually match well with the relatively high luminal flux rate and fluctuating conditions of the small intestinal lumen, which appears to play a key-role in the selective pressure on the microbiota and explains why bacteria within the Lactobacillales-order are not only dominantly present, but also appear to be very active carbohydrate metabolisers in the ileal lumen of this ileostomy subject.

Transcripts detected at both time points

From all metatranscriptome profiles, five excised bands and cognate sequences were detected at both time points. Two of these sequences (II and V) were assigned to the functional class of carbohydrate metabolism and transport based on conserved domains. Sequence II gave only one significant hit with a putative receptor from *Pseudomonas fluorescens*, whereas sequence V showed highest similarity with a sugar hydrolase from *Lactococcus lactis*, but the sequence was shared by other lactobacilli, including species within the genera *Streptococcus* and *Enterococcus*. The only hit with high similarity to sequence I (only 63 nucleotides) was a glutamate synthase (large subunit) from *S. mutans*, which tentatively places this transcript within the amino acid and energy metabolism clusters. A hypothetical protein shared by several streptococci showed high similarity with sequence III. Although this sequence was assigned to the cluster "function unknown" (S), hits with lower similarity included several sequences encoding transcription termination NusA-like proteins, suggestive for involvement in transcription. Finally, sequence IV showed highest sequence similarity with a ribosomal subunit interface protein (functional cluster of translation, ribosomal structure and biogenesis (J)) shared by streptococci and lactobacilli. Overall, these results indicate that only a few genes are found to be expressed on both days. In addition, it demonstrates that cDNA-AFLP allows monitoring of gene expression in time, despite the fact that only a small fraction of the complete metatranscriptome is analysed.

Retrieval of transcripts in other GI tract samples

To investigate the distribution of the retrieved sequences in other ileum and colon contents, 14 sequences were selected for which primers were designed that enabled the recovery of corresponding PCR-amplicons in the DNA prepared from ileostomy effluent from the same subject (i4). Subsequently, the same primer pairs were employed to investigate the presence of the identical or closely related gene fragments in DNA samples prepared from GI tract samples of nine other individuals. The results of this PCR survey as well as the sequence similarity of several recovered PCR amplicons in comparison with the original transcript sequences are depicted in Table 6-3. Remarkably, six sequences that were detected in all healthy subjects with an ileostomy showed a very high sequence similarity (>95% sequence similarity for most of them) to the original corresponding transcript sequence. Overall, more than 64% of all transcript sequences could be retrieved in the four healthy ileostomy subjects. Although intra-individual differences in microbiota composition as well as daily fluctuations were considerable, about 40% of the detected phylotypes in each healthy ileostomy subject were found in common with the other four

healthy subjects with an ileostomy, as described in Chapter three. Taken into account the fact that within genera or species considerable variation in gene distribution can exist, retrieval of 64% of the analysed transcript sequences in other ileostomy effluent samples indicates a very coherent ecological niche, on the level of functionality employed by the ileostomy effluent microbiota. In addition, 56% and 50% of the transcript sequences could be retrieved from the subject with the ileorectal pouch and the two medicated ileostomy subjects, respectively. The fact that these numbers are slightly lower might already have been expected based on basis of their more deviating microbiota compositions, especially in case of the medicated ileostomy subjects. In the faecal samples of two healthy volunteers only 19% and 56% of the sequences could be detected, respectively. This low number in at least one of the faecal samples might be explained by the fact that in the terminal part of the colon the amount of microorganisms outnumbers that of the ileum by a factor of 100 to a 1000, and the low abundance of ileum derived microbes may place them below the PCR-detection limit.

TABLE 6-1 Overview of the gene sequences obtained by cDNA-AFLP profiling that could be retrieved in the DNA of five healthy ileostomy subjects (i1 to i5), two medicated ileostomy subjects (m1 and m2), effluent from one subject with an ileorectal pouch (p1), and faecal samples from two healthy adults (f1 and f2). White boxes indicate that the gene sequence could be detected, the numbers indicate sequence similarity of the PCR product with the sequence obtained from the cDNA-AFLP profiles. Gray boxes indicate that no gene sequence was detected, whereas asterisks indicate PCR artefacts.

Transcript	Highest hit	Ileal effluent samples								Faecal samples	
		i4	i1	i2	i3	i5	p1	m1	m2	f1	f2
4-G01	Hypothetical protein (*Lactobacillales*)		87	97							96
4-G09	Sortase B family protein (*Lactobacillales*)										*
4-G08	Organic radical activating enzyme (*Enterobacteria*)	100	98	97	97	98	100				
4-H02	7,8-Dihydropteroate synthase (*Enterobacteria*)		95			95	91	94			
4-G07	Macrolide transporter (*Enterobacteria*)		98	98							
4-H01	Melibiose carrier protein (*Enterobacteria*)	99	97	98	96			96			
4-E02	Putative GTP pyrophosphokinase (*Lactobacillales*)	99	98	99			91				
4-H11	Putative transcriptional regulator (*Bacteroidales*)										
4-H03	Major facilitator superfamily (*Enterobacteria*)				*						
4-A05	Sensory box histidine kinase (*Lactobacillales*)										
4-E01	Putative bèta-galactosidase (*Lactobacillales*)		97								
4-D07	Polysaccharide biosynthesis family (*Lactobacillales*)			*					*		
4-C01	Hypothetical protein (*Lactobacillales*)		92								
4-C02	Homoserine O-succinyltransferase (*Lactobacillales*)	100	94								
4-D12	No significant hit										
4-A01	No significant hit										

CONCLUSION

Overall we can conclude that genes related to carbohydrate metabolism and transport were dominantly expressed by the microbiota in the ileostomy effluent of the subject under study. This functional category encompassed almost 40% of the overall expressed genes belonging to the metabolism cluster. The transcripts detected belonging to this functional cluster were mostly involved in diverse pathways, including the pentose phosphate pathway and metabolism of other

sugars (fructose, galactose and starch), in contrast to the metatranscriptome of the faecal microbiota where all expressed carbohydrate metabolism genes were related to glycolysis. The relatively high luminal flux rate in the ileum compared to the colon, might play a key-role in the selective pressure on the microbiota, favouring the rapid carbohydrate-metabolisers. Furthermore, the phylogenetic distribution of the protein-encoding transcripts indicates that species belonging to the order Lactobacillales play a predominant role in the microbiota activity of ileostomy effluent. Finally, retrieval of on average 64% of transcript sequences in four other healthy ileostomy subjects suggests that the microbiota in this ecosystems functions coherently in the ecosystem of the ileal lumen.

ACKNOWLEDGEMENTS

We are very grateful to the ileostomy subject for providing ileostomy effluent samples. Freddy Troost from Maastricht University (Nutrim) is thanked for his help with the MEC approval, and usefull discussions regarding the sampling procedure. Furthermore, we would further like to thank Milkha Leimena from Wageningen University for providing PCR data. Finally, we are grateful to Peter Schaap from Wageningen University for BLAST analysis of the retrieved transcripts against sequences from the ileostomy effluent metagenome. Finally, the authors would like to thank Hanneke Witsenboer and Jenny Peters from Keygene in Wageningen for cDNA-AFLP analysis.

CHAPTER 7

Novel mucin degradation genes identified in a human ileal metagenome library

Carien C.G.M. Booijink, Marion Leclerc, Fabienne Béguet, Erwin G. Zoetendal, Michiel Kleerebezem, Willem M. de Vos & Joël Doré

Manuscript in preparation

ABSTRACT

Mucins in the human small intestine are providing a continuous carbon and nitrogen source to the residing microbiota. To investigate the potential of intestinal microbes to degrade mucin a human ileal mucosa metagenomic library, comprising more than 20,000 fosmid clones with average insert size of 30 kb in *Escherichia coli*, was screened for growth on mucin or mucin-related compounds. On purified pig gastric mucin, a total of 212 colonies were able to grow, whereas 79 could grow on N-acetyl-D-glucosamine. Enzyme screening of these metagenomic clones led to the identification of two fosmids (53C5 and 53H12) that coded for high N-acetylgalactosaminidase activity outside and inside the cells, respectively, compared to an empty fosmid clone. Sequence analysis showed that fosmid insert 53C5 contained two putative genes with sequence similarity to a glycosyltransferase and an inner-membrane sugar transporter probably originating from *Bacteroides*. Furthermore, fosmid insert 53H12 was predicted to encode a protein with similarity to the soluble lytic murein transglycosylase from *Enterococcus faecalis* known to contain a conserved domain with an N-acetylglucosamine binding site. In addition, fosmid insert 53H12 was found to contain five genes derived from the Tn*1549*-like transposons originally identified in *E. faecalis*, including genes encoding a methyl transferase, relaxase and a helicase. These findings represent the first functional screening of an ileal metagenome and indicate that genes involved in mucin-degradation are ubiquitous in ileal microbes and can be derived from Gram-negative as well as Gram-positive bacteria.

INTRODUCTION

In the human gastrointestinal tract (GI tract) mucins have an important role as defensive barrier. Mucins cover the epithelial cells as a mucus layer, that consists of a highly hydrated gel (~95% water) containing salts, lipids, and proteins such as defensins, immunoglobulins, growth and trefoil factors, as well as released epithelial cells and fragments thereof [34]. On a molecular level, mucus consists of a protein backbone (apomucin) that is glycosylated with the oligosaccharides such as N-acetylgalactosamine (GalNac), N-acetylglucosamine (GlcNac), fucose, galactose and sialic acid [54]. Next to its function as a protective physical barrier between the mucosal surface and the luminal contents, mucus also acts as a lubricant to facilitate food passage [186].

Representatives of several phylogenetic groups of the microbiota inhabiting the human GI tract are able to use mucus as a carbon and/or nitrogen source. These bacteria are assured of a steady supply, as mucus is constantly produced and secreted by the human host. This constant secretion is necessary due to the short life time of mucus, varying from a few minutes to several hours [10]. For the complete degradation of mucins, a wide range of enzymes is needed due to the high complexity and diversity of mucins. Enzymes like glycosidases, proteases, sulphatases and sialidases are involved in the complete degradation pathway [29, 79]. It is estimated that approximately 1% of the total cultivable intestinal microbiota is able to produce the panel of enzymes required to degrade the complex structure of mucin into free sugars [79]. Although some single bacterial species involved in mucus-utilisation are able to produce a panel of mucolytic enzymes, mucin degradation is considered as a cooperative activity of multiple species. Isolates belonging to the mucin-degraders are members of the genera *Bifidobacterium*, *Bacteroides* and *Ruminococcus* and the recently described isolate *Akkermansia muninciphila* [35]. In addition, genome analysis showed that bacteria belonging to the genera *Clostridium*, *Prevotella*, *Vibrio* and *Streptomyces* also encode one or multiple enzymes involved in mucin-degradation [29] (see also Table 1-1). As these species are all isolated from colonic mucosa or faecal samples, little is known about their presence in the proximal GI tract. Mucin-degraders might be more abundant in the ileum, as the bacterial cell densities in mucosa samples of the terminal ileum were found to be higher compared to colonic mucosa samples [1]. It has been shown that species within the phyla Bacteroidetes and Verrucomicrobia, and *Clostridium* clusters XIVa and IV were most abundant in the ileal mucosa [216, 217]. However, there is presently no information on the mucin-degrading capacity of the ileal mucosa microbiota.

So far, the encoding genes and proteins involved in mucin degradation have been detected based on culture-dependent approaches or sequence analysis of bacterial genomes. Metagenomics provide researchers with another means to screen for enzymes in diverse microbial ecosystems like the GI tract [68]. Here, we used a metagenomic library constructed from DNA derived from bacterial cells of the mucosal layer of the human ileum to screen for genes encoding enzymes involved in the degradation of mucin and its components.

MATERIALS & METHODS

Specimen

Part of the distal ileum was obtained from a 51-year old male patient undergoing colonoscopy and surgery for lower colon cancer suspicion. The patient had not taken antibiotics three months prior to the sampling and showed an average body mass index. A segment of two cm^2 was obtained and immediately frozen and kept at -80°C until processing. The protocol was approved by the local ethics committee and informed consent was obtained from the subject before sampling.

Recovery of microbial cells

In order to separate the microbial cells from the mucus layer, the sample was first incubated for one hour at room temperature in a Balch tube sealed with rubber stopper containing anaerobic PBS buffer with 0.1 % Urea. The 2 cm^2-specimen was then cut out, linearised and gently scraped under a 0.3 kPa CO_2 flow. Cell separation was performed using a 16% Nycodenz gradient in Tricine NaOH buffer according to Roy *et al.* [167]. The microbial cells were rapidly retrieved from the gradient using a Pasteur pipette and 10 µl were kept for microscopic examination and Gram staining. The cells were kept at -80°C.

The bacterial pellets were resuspended in a 50 mM Tris (pH 8.0), 100 mM EDTA buffer, mixed with an equal volume of molten 1.6% Incert agarose (BMA), and then transferred into disposable plug molds (Bio-Rad). The lysis of the microorganisms was then performed as described by Nalin *et al.* [138]. Agarose plugs were first transferred in 45 ml of LA lysis buffer (50 mM Tris [pH 8.0], 100 mM EDTA, 5 mg of lysozyme ml^{-1}, 0.5 mg of achromopeptidase ml^{-1}) and incubated at 37°C for 6 h. The agarose plugs were then incubated in 45 ml of SP lysis buffer (50 mM Tris [pH 8.0], 100 mM EDTA, 1% lauryl sarcosyl, 2 mg of proteinase K ml^{-1}) at 55°C for 24 h. An additional incubation for 24 h was performed with fresh SP buffer. Agarose plugs were finally equilibrated in a 10 mM Tris (pH 8.0), 1 mM EDTA storage buffer.

Construction of the metagenomic library

High-molecular-weight bacterial DNA trapped in agarose plugs was immediately inserted into the wells of a 0.8% low-melting-temperature gel (Bio-Rad) and separated for 18 h by pulsed-field gel electrophoresis at 4.5 V cm^{-1} with 5- to 40-s pulse times with a CHEFDRIII apparatus (Bio-Rad). DNA fragments ranging between 35 and 48 kbp were isolated after HaeIII digestion and then recovered from the gel with GELase (Epicentre Technologies). Metagenomic DNA was then cloned into fosmids by using the EpiFos fosmid library production kit (Epicentre Technologies) as recommended by the manufacturer. Recombinant colonies were transferred to 96-well microtiter plates containing freezing medium (Luria-Bertani, 20% glycerol complemented with 12.5 µg of chloramphenicol ml^{-1}). After growing at 37°C for 22 h, the plates were stored at -20°C. Overall, 20,160 clones were obtained.

Mucin purification

Porcine gastric mucin (PGM, type III, Sigma) was purified prior to addition to culture media as described before [225]. In short, 10 g crude pig gastric mucin was suspended in 0.02 M phosphate buffer (pH 7,8) with 0.1 M NaCl and a few drops of toluene. The solution was subsequently centrifuged (10,000× g; 30 min at 4°C) and the supernatant was taken and cooled till 4°C. Pre-chilled ethanol (4°C) was added to a final concentration of 60% (v/v). After centrifugation (10,000× g; 30 min at 4°C), the resulting precipitate was dissolved in 0.1 M NaCl followed by precipitation of the mucin with ethanol as described above. These steps were repeated once more and the final purified mucin pellet was washed once with 100% ethanol prior to dialysis against distilled water (16 h at 4°C). The resulting solution of purified Pig Gastric Mucin (pPGM) was freeze-dried and used as additive to the culture media.

Culture conditions and screening strategy

The *E. coli* DH10B strain was routinely grown in Luria-Bertani (LB) broth at 37°C. The growth medium for transformed *E. coli* cells was supplemented with 12 μg ml^{-1} chloramphenicol (Cm) to maintain the EpiFOS-5 vector. To make pools per 96 fosmid clones, all fosmid clones were grown aerobically on 96-deep-well plates containing LB broth with Cm during 16 h, while shaking at 180 rpm. After vortexing the deep-well plates for 1 min, the fosmid clones were pooled per 96-wells plate and a glycerol stock for storage at -80ºC was prepared. All 210 pools of fosmid clones were distributed over three new 96-wells plates, to make screening of multiple pools of fosmid clones feasible (see also Fig. 7-1).

The fosmid clone pools were inoculated on 96-wells plates with 200 μl MOPS minimal medium [139] supplemented with Cm and one of the following carbon sources: 4 mM of N-acetyl-D-glucosamine (GlcNac; Sigma), 4 mM N-acetyl-D-galactosamine (GalNac; Sigma), 4 mM L-fucose (Sigma) or 3% (v/v) purified porcine mucin, on both liquid and solid medium. In all culture experiments *E. coli* DH10B with an insert-free EpiFOS-5 vector was taken along as control. All fosmid clone pools that showed growth to an OD >0.3 after 48 h were selected and the individual fosmid clones within that pool were screened for growth in the same manner. From all individual fosmid clones, the best 30 growers were selected per substrate and subjected to additional enzyme activity tests.

Enzyme activity assays

The fosmids clones that showed growth on mucin or mucin components (GlucNac, GlcNac or fucose) were selected and glycerol stocks of these fosmids were organised in a 96-wells sublibrary. These mucin-utilisers were pre-grown on LB with Cm for about 16 hours (OD_{600} of ~0.5). Nac-glycosidase activities were determined by the hydrolysis of *p*-nitrophenol (*p*-NP) linked substrates (Sigma-Aldrich, St. Louis, USA) at 37°C during 1 h. The enzyme assay mixture contained 1 mg ml^{-1} *p*-NP solution in 190 ul MOPS medium with 50 mM potassium phosphate buffer (pH 7) and 10 μl of pre-grown fosmids in LB. The colour formation was measured spectrophotometrically at 405 nm and the enzyme activity was calculated with the applicable molar extinction coefficient of the *p*-NP substrate under these conditions (13,700 M^{-1} cm^{-1}). One

unit of enzyme activity was defined as 1 μmol of *p*-NP liberated per minute per mg protein at 37°C. Furthermore, the glycosidase enzyme activities of individual fosmids from the ileal mucosa library were investigated during 1 h in the supernatants and cell extracts separately after 48 h growth on MOPS medium with one of the carbon sources under study.

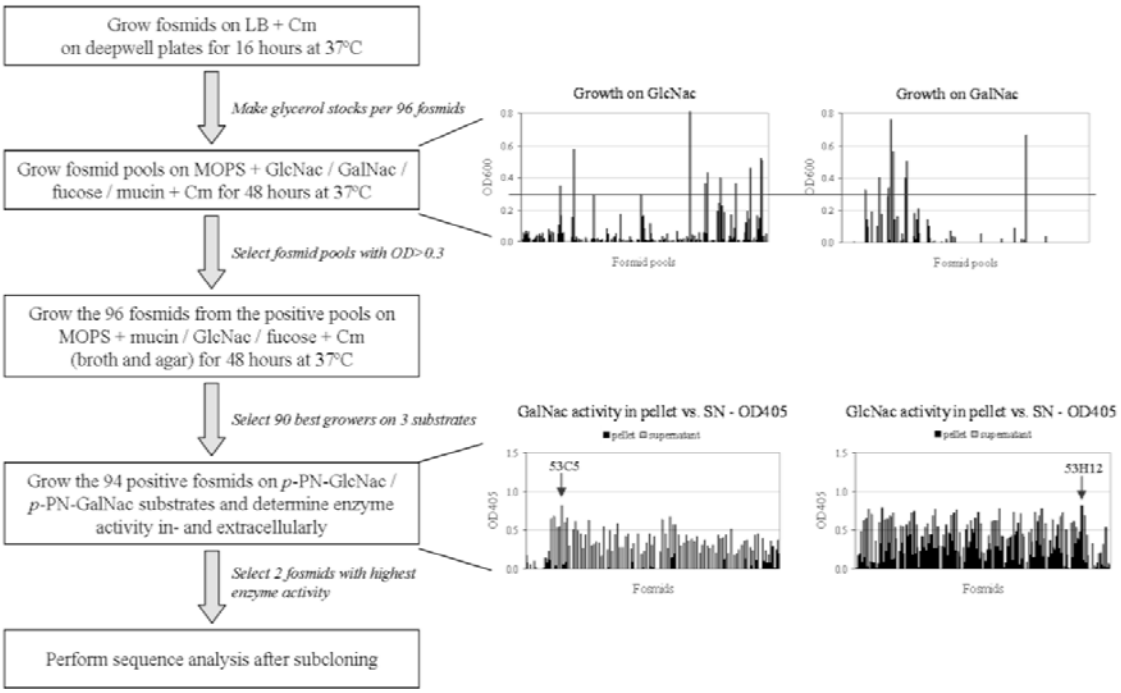


FIGURE 7-1 Layout of the experimental procedure for the functional screening of the ileal mucosa metagenome.

EpiFOS extraction

Prior to extraction of the EpiFOS vectors, the *E. coli* DH10B clones containing the vector inserts of interest were grown in 5 ml LB with Cm during 16 h, while shaking at 180 rpm. NucleoBond PC20 (Macherey-Nagel, Düren, Germany) was used for fosmid extraction. Yields were between 15 and 18 μg per fosmid as determined with NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Extracted fosmids were also inspected qualitatively by agarose gel electrophoresis on a 1% agarose gel.

Sequence analysis of clones growing on mucin

The inserts of the two vectors from fosmid clones growing on MOPS supplemented with GlcNac, GalNac or mucin were sequenced by Genoscope (France) after subcloning the fosmid insert. After assembly of the subclones into a contig, putative open reading frames (ORFs) were identified through the genomic database PEDANT (a Protein Extraction, Description and Analysis Tool) (http://pedant.gsf.de) [57, 163]. The predicted ORFs that were at least 90 bp in length were translated and their putative function was automatically annotated based on their similarities to sequences in the COG (Clusters of Orthologous Groups) and Pfam (Protein

Families) databases. The sequences without significant matches were further analysed with PSI-Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi). All sequences longer than 300 bp with an e-value of more than 10^{-3} in the BlastP searches were considered to be hypothetical. The cellular location of the putative gene products was predicted *in silico* with the protein subcelullar-location predictor LocateP [226].

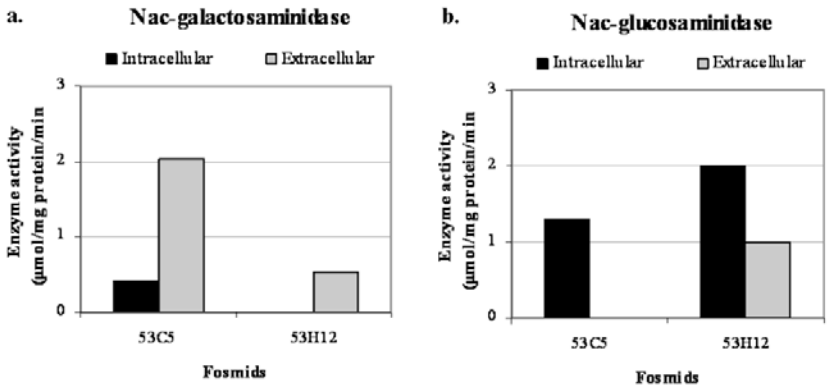
RESULTS AND DISCUSSION

The mucin degrading capacity of the microbiota residing in the mucus layer of the human ileum was investigated with a metagenomic approach. All 20,160 clones of an ileal mucosa metagenomic library were screened for the ability to use mucin or mucous components. From the 210 pools of 96 fosmid clones each, nine pools showed significant growth (OD_{600} >0.5 after 48 hours) on GlcNac, ten pools showed similar growth on GalNac, whereas six pools grew well on fucose. This growth could be attributed to the insert present in the selected fosmids, since *E. coli* DH10B harbouring an empty fosmid grew very poorly on any of the four tested substrates (OD_{600} <0.1).

Each of the 96 fosmid clones present in the positive pools was subsequently analysed for growth on mucin, fucose and GlcNac. Overall, 79 (0.39%) of the clones showed growth with an OD_{600} >0.3 on GlcNac after 48 h, on fucose a total of 25 (0.13%) fosmid clones could grow. On the more complex and diverse carbon source mucin, 212 (1.1%) fosmid clones showed growth. Subsequent enzyme activity tests of N-acetyl-β-D-glucosaminidase and N-acetyl-β-D-galactosaminidase were performed for 60 fosmid clones that were selected based on highest OD_{600} reached within 48 hours on MOPS supplemented with GlcNac or mucin. The majority of fosmid clones showed an average enzyme activity of N-acetyl-β-D-galactosaminidase on *p*-NP GalNac equal to *E. coli* DH10B control, both intra- and extracellularly. This was considered background enzyme activity by the *E. coli* host. However, fosmid 53C5 showed a two-fold higher enzyme activity compared to control, both in and outside the cell (Fig. 7-2a). In this context, it should be noted that the location of enzyme activity in the *E. coli* host does not necessarily correspond to the location of enzyme activity in the microbe from which the fosmid insert originates, this holds notably for Gram-positive bacteria. The other enzyme activity screen revealed that intracellular activity of N-acetyl-D-β-glucosaminidase was highest in fosmid 53H12 (2.0 μmol mg^{-1} protein min^{-1}), but also showed that the extracellular activity was higher compared to average and control (Fig. 7-2b). This fosmid was originally screened positively for growth on mucin. When the fosmids were grown on LB for 6 h (following lag-phase), it was found that fosmid 53H12 showed a more rapid increase in cell density compared to fosmid 53C5 or the average of all tested fosmids, reflected by the generation time (Fig. 7-2c). It should be noted that the fosmid clones initially grow slower compared to DH10B, possibly due to the presence of foreign DNA in the fosmid cells.

Sequence analysis of fosmids 53C5 and 53H12 was performed followed by automated annotation of the insert sequences. In fosmid 53C5 a total of 19 open reading frames (ORFs) were identified, most of which with considerable similarity to genes from *Bacteroides* (Fig. 7-3), a genus known for its wide range of sugar- and mucin-degrading enzyme-encoding genes

[110, 169, 223]. Many of these were annotated as encoding uncharacterised proteins (5 ORFs) or showed no significant similarity ($e<10^{-10}$) to known encoded proteins in the database (4 ORFs). Two putative genes were associated with the metabolism of sugars: ORF 10 showed similarity with a glycosyl transferase, with an intracellular location prediction; ORF 15 was identified as sugar transporter based on Pfam-assignment, with a location prediction on the inner membrane. Glycosyl transferases function in the catalysis of glycosidic bond formation and use donor sugar substrates that contain a phosphate leaving group [96]. A schematic overview of the transglycosylation reaction performed by glycosyl transferase is: R'-OH + R''-O-R → R-O-R' + R''-OH where R, R' and R'' represent sugar moieties [206]. It has been demonstrated previously that glycosyl hydrolases can also exert glycosyl transferase activity in for instance bifidobacteria [82, 208]. Therefore, it is conceivable that this presumed glycosyl transferase of *Bacteroides* may also hydrolyse glycosidic bonds, for instance in those derived from mucin.

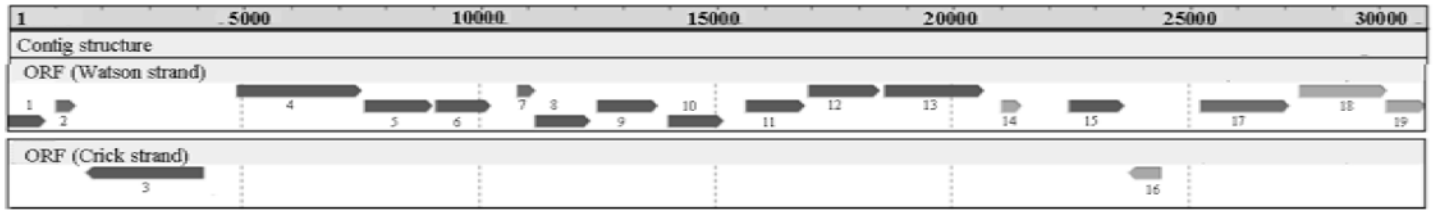


c.

	53C5	53H12	DH10B
GlcNac	142	188	364
GalNac	270	-	220
Mucin	126	134	112
LB	63	41	29

FIGURE 7-2 (**a**) Enzyme activity of N-acetyl-β-D-galactosaminidase (NacGal) for the two mucin-utilising fosmids 53C5 and 53H12, compared to the average enzyme activity of 90 selected fosmids and control (DH10B). The enzyme activity was determined on *p*-NP-linked NacGal, both intra- and extracellularly. (**b**) Enzyme activity of N-acetyl-β-D-glucosaminidase (NacGlc) of both fosmids compared to enzyme activity of the control. The enzyme activity was determined on *p*-NP-linked NacGlc, intra- as well as extracellularly. (**c**) Generation time (in min) of fosmid clones 53C5 and 53H12 on minimal medium (MOPS) supplemented with GlcNac, GalNac or mucin, compared to rich medium (LB). *E. coli* DH10B with an empty fosmid was taken along as control.

The majority of the 18 identified putative genes present on the insert of fosmid 53H12 showed highest similarity with genes of the species *Enterococcus faecalis* (Fig. 7-3). Interestingly, species belonging to the genus *Enterococcus* were so far not associated with mucin utilisation and this finding might direct towards a novel pathway for this genus. Again, a large proportion of putative genes (9 out of the 18) showed only similarity with uncharacterised proteins. Based on gene orthology, one putative gene was annotated as a soluble lytic murein transglycosylase (ORF 15). Bacterial lytic transglycosylases degrade murein via cleavage of the β-1,4-glycosidic bond between N-acetylmuramic acid and N-acetylglucosamine, while forming a 1,6-anhydrobond in the



ORF[a]	Length (aa)	Closest similar protein	Genbank acc nrs of identical protein	Organism	E-value	Putative function based on COG	Pfam	Protein location prediction[b]	Involved in pathway[b]
1	517	Methionyl-tRNA synthetase	ref|NP_811845	*Bacteroide thetaiotaomicron* VPI-5482	0	Methionyl-tRNA synthetase, COG0143	tRNA synthetases class I (M)	Intracellular	No pathway
2	138	Put. uncharacterised protein	ref|ZP_03209814	*Bacteroides plebeius* DSM 17135	4.E-46	Uncharacterized protein conserved in bacteria, COG2832	Protein of unknown function (DUF454)	Multi-transmembrane, innermembrane	Sec-(SPI)
3	831	TonB-dependent receptor domain prot.	ref|ZP_01883358	*Pedobacter* sp. BAL39	5.E-155	Outer membr. receptor proteins, mostly Fe transport, COG1629	TonB-dependent receptor plug domain	Multi-transmembrane (outer)	Sec-(SPI)
4	879	Put. TonB-linked outer membr. prot.	ref|YP_001298170	*Bacteroides vulgatus* ATCC 8482	0	Outer membr. cobalamin receptor protein, COG4206	TonB-dependent receptor plug domain	Intracellular	No pathway
5	477	Put. uncharacterised protein	ref|ZP_02072979	*Bacteroides uniformis* ATCC 8492	3.E-155			Multi-transmembrane (outer)	Sec-(SPII)
6	384	Put. iron uptake factor	gb|ABR38546	*Bacteroides vulgatus* ATCC 8482	5.E-126	Uncharacterized iron-regulated membr. protein, COG3182	PepSY-associated TM helix	Multi-transmembrane, innermembrane	Sec-(SPI)
7	125	Protein crcB homolog	ref|NP_662976	*Bacteroides thetaiotaomicron*	2.E-21	Integral membr. protein possibly involved in chromosome condensation, COG0239	CrcB-like protein	Multi-transmembrane, innermembrane	Sec-(SPII)
8	388	Put. uncharacterised protein	ref|ZP_01958711	*Bacteroides caccae* ATCC 43185	3.E-105		Thioredoxin-like superfamily	N-terminally anchored	Sec-(SPI)
		Thiol:disulfide interchange protein	gb|EAY32084	*Microscilla marina* ATCC 23134	8.E-17				
9	419	Phosphodiesterase I	ref|YP_099545	*Bacteroides fragilis* YCH46	2-111	Uncharacterized proteins of the AP superfamily, COG1524	Type I phosphodiesterase / nucleotide pyrophosphatase	N-terminally anchored	Sec-(SPI)
10	386	Put. glycosyl transferase	ref|YP_100690	*Bacteroides fragilis* YCH46	3.E-137	Glycosyltransferase, COG0438	Glycosyl transferases group 1	Intracellular	No pathway
11	410	Put. uncharacterised protein	ref|ZP_03300700	*Bacteroides dorei* DSM 17855	0	Uncharacterized membr. protein, required for spore maturation, COG2715	Nucleoside recognition	Multi-transmembrane, innermembrane	Sec-(SPI)
		Put. transmembr. spore maturation-like prot.	gb|YP_209865	*Bacteroides fragilis* NCTC 9343	4.E-170				
12	499	Polysaccharide biosynthesis protein	ref|YP001298602	*Bacteroides vulgatus* ATCC 8482	4.E-168	Membrane protein involved in the export of O-antigen and teichoic acid, COG2244	Polysaccharide biosynthesis protein	Multi-transmembrane, innermembrane	Sec-(SPI)
		Polysaccharide biosyntesis protein	gb|ABR38980	*Bacteroides vulgatus* ATCC 8482	3.E-167				
13	706	Put. uncharacterised protein	ref|YP_097449	*Bacteroides fragilis* YCH46	8.E-180		Peptidase	N-terminally anchored	Sec-(SPI)
		Put. exported protein	gb|ABR39912	*Bacteroides vulgatus* ATCC 8482	5.E-141				
14	138	Predicted phophohydrolase	ref|ZP_00379769	*Brevibacterium linens* BL2	5.E-01			Secreted to periplasm	Sec-(SPI)
15	391	Put. antiobitic resistance transport prot.	ref|YP_212544	*Bacteroides fragilis* NCTC 9343	1.E-75	Permeases of the major facilitator superfam., COG0477	Sugar (and other) transporter	Multi-transmembrane, innermembrane	Sec-(SPI)
16	227	Put. adhesin	gb|AAX16421	unc. murine large bowel bact. BAC 54B	2.E-01			Secreted to periplasm	Sec-(SPI)
17	619	Hemin-binding protein	gb|AAX16422	unc. murine large bowel bact. BAC 54B	2.E-19			Multi-transmembrane (outer)	Sec-(SPI)
18	614	YD repeat containing protein	ref|YP554026	*Burkholderia xenovorans* LB400	3.E-01			Multi-transmembrane (outer)	Sec-(SPI)
19	274	Unnamed protein product	emb|CAO14560	*Vitis vinifera*	9.E-01			Intracellular	No pathway

[a]ORFs possibly involved in mucin degradation are indicated with red boxes
[b]Location prediction and number of transmembrane helices as predicted by the program LocateP

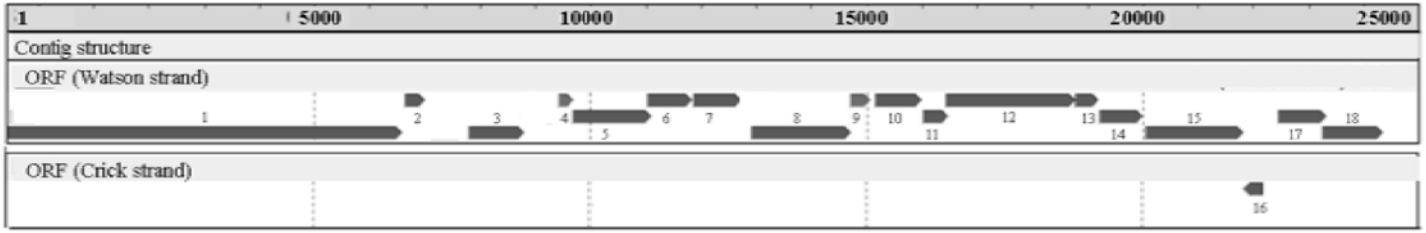
FIGURE 7-3 Overview of the 19 ORFs annotated in the DNA insert of mucin-utulising fosmid 53C5 comprising 30,742 bp. Grayscale used are corresponding to the alignment scores with the best BlastP hit: light gray = no BlastP hit; 80 < gray < 199; black >200. The description of the annotated ORFs in fosmid 53C5, including the observed sequence similarities and predicted protein locations, are depicted in the table. For all annotated genes with highest similarities with putative uncharacterized proteins, additional PSI-Blast results are shown in gray. Boxes indicate genes encoding proteins possibly involved in mucin metabolism.

muramic acid residue [197]. This function is related to cell wall metabolism, however, in ORF 15 of fosmid 53H12 no conserved domains corresponding to N-acetylmuramic acid binding sites could bedetected although these were detected for N-acetylglucosamine. Possibly, the encoded protein of ORF 15 can exert the cleavage of GlcNac-GlcNac bonds. Furthermore, mucoid types of *Enterococcus faeacalis* were detected before [19], suggesting that mucin or mucin-derived components can be transported over the membrane of *Enterococcus* species. An intracellular location was predicted for the encoded protein by ORF 15, as no signal peptide was detected (Fig. 7-4). This corresponded well with the higher enzyme activity that was detected inside the cells haboring fosmid 53H12 than outside these cells. Although no sugar transporter was annotated on the insert sequence of fosmid 53H12, transport of mucin-related compounds is most likely exerted by one of the nine uncharacterised ORF gene products or by the *E. coli* host.

The abundantly available sequence information from the so-called sequence-driven metagenome analysis, recently revealed that a set of 58 COGs were over-represented in the microbiome present in faeces of healthy infants and adults [95]. Remarkably, in fosmid 53H12 four of the in total six assigned COGs belonged to this enriched group (COGs 3451, 3505, 4509, 4646). Especially the assigned DNA methylase (COG4646) to ORF 1 and sortase B (COG4509) to ORF 14 were enriched about 30-fold compared to metagenomes of other environments. In addition, five of eighteen annotated genes in fosmid 53H12 (ORFs 1, 2, 3, 5 and 14) were found to be homologous to those on conjugative transposon *Tn*1549. Originally this transposon was identified in a strain of *E. faecalis*, but related transposons have also been detected in other species belonging to the genera *Clostridium* and *Streptococcus* [40, 146, 183]. Furthermore, the genes found in fosmid insert 53C5 related to transposon *Tn*1549 were retrieved numerously in the different faecal metagenomes [64, 95], indicating that the genes transferred by this transposon are advantageous to the recipient microbe including the vancomycin-resistance that this transposon can confer [153]. Thirty putative ORFs were assigned to *Tn*1549, including eight ORFs related to plasmid mobilisation and conjugation [62].

Next to the vancomycin-resistance genes, genes encoding two proteins involved in conjugative transfer (TrsE- and TrsK-like proteins) [136], an LtrC-like protein involved in the conjugation of a plasmid from *Staphylococcus aureus* [15] and a relaxase involved in plasmid mobilisation [151] were found on the left extremity of *Tn*1549. Homologues of these genes involved in conjugation and plasmid mobilisation were also found to be encoded by fosmid 53H12.

The function-based screening described here is especially suitable for the detection of novel functional genes or gene clusters. This brings about the problem that sequence analysis in case of previously uncharacterised genes will not provide predictive information concerning the function of the gene. More than 50% of the genes detected within the fosmid insert encoded conserved proteins of unknown function. Consequently, genes responsible for the phenotype that was screened for should be identified through additional experiments like transposon mutagenesis and sub-cloning, followed by assaying the derivative clones for growth on mucin or its derived compounds. The functional role of the ORFs predicted to be involved in the growth on mucin can be verified in this manner. Furthermore, this could also aid in the elucidation of the responsible gene(s) for the relatively fast growth of fosmid 53H12 compared to control.



ORF[a]	Length (aa)	Closest similar protein (UniProtKB/TrEMBL)	Other accession nrs of identical protein	Organism	E-value	Putative function based on COG	Pfam	Protein location prediction[b]	Involved in pathway[b]
1	2394	Superfam. II DNA + RNA helicase	gb|ABF36203; EF2307	*Streptococcus pyogenes* MGAS2096	0	DNA methylase, COG4646	N-6 DNA Methylase	Intracellular	No pathway
2	115	Putative uncharacterised protein	gb|AAQ16249; EF2306	*Enterococcus faecalis*	2.E-59			Intracellular	No pathway
		Unknown protein	gb|AAQ16249	*Enterococcus faecalis*	6.E-59				
3	329	Put. uncharacterised protein	ref|YP_600731; EF2305	*Streptococcus pyogenes* MGAS2096	5.E-137			Intracellular	No pathway
		LtrC	gb|ABF36187	*Streptococcus pyogenes* MGAS2096	3.E-137				
4	86	MocB-like protein	gb|AAQ16253	*Enterococcus faecalis*	7.0E-43			Intracellular	No pathway
5	470	Rlx-like protein	gb|AAQ16254; EF2303	*Enterococcus faecalis*	0		Relaxase/Mobilisation nuclease domain	Intracellular	No pathway
6	270	Put. uncharacterised protein	gb|AAQ16255	*Enterococcus faecalis*	2.E-137			Intracellular	No pathway
		Unknown protein	gb|AAQ16255	*Enterococcus faecalis*	2.E-137				
7	287	LtrC-like protein	gb|AAQ16256	*Enterococcus faecalis*	3.E-143			Intracellular	No pathway
8	596	TrsK-like protein	gb|AAQ16257	*Enterococcus faecalis*	0	Type IV secretory pathway, VirD4 components, COG3505	TraG/TraD family	Multi-transmembrane	Sec-(SPI)
9	118	VirB2-like protein	gb|AAQ16258	*Enterococcus faecalis*	4.E-46			Multi-transmembrane	Sec-(SPI)
10	278	Put. uncharacterised protein	ref|YP_60723	*Streptococcus pyogenes* MGAS2096	2.E-149			Multi-transmembrane	Sec-(SPI)
		Hypothetical protein MGAS2096	gb|ABF36179	*Streptococcus pyogenes* MGAS2096	3.E-149				
11	150	Put. uncharacterised protein	gb|AAQ16261	*Enterococcus faecalis*	1.E-59			Multi-transmembrane	Sec-(SPI)
		Unknown protein	gb| AAQ16261	*Enterococcus faecalis*	6.E-59				
12	777	TrsE-like protein	gb|AAQ16262	*Enterococcus faecalis*	0	Type IV secretory pathway, VirB4 components, COG3451		Intracellular	No pathway
13	134	Put. uncharacterised protein	gb|AAQ16262	*Enterococcus faecalis*	6.E-75			Intracellular	No pathway
		Unknown protein	gb|AAQ16263	*Enterococcus faecalis*	3.E-74				
14	257	Put. uncharacterised protein	gb|AAQ16264	*Enterococcus faecalis*	2.E-136	Unchar. prot. conserved in bacteria, COG4509	Sortase B	N-terminally anchored	Sec-(SPI)
		Sortase B family protein	gb|ABF36174	*Streptococcus pyogenes* MGAS2096	3.E-121				
15	591	Put. uncharacterised protein	gb|AAQ16265	*Enterococcus faecalis*	0	Soluble lytic murein transglycosylase & related regulatory proteins, COG0741		Intracellular	Possibly Sec-
		Antigen-like protein	gb|ABF36173	*Streptococcus pyogenes* MGAS2096	0				
16	119	Put. uncharacterised protein	gb|AAQ16275	*Enterococcus faecalis*	3.E-52		Helix-turn-helix	Intracellular	No pathway
		MerR family transcriptional regulator	gb|ABF36172	*Streptococcus pyogenes* MGAS2096	4.E-37				
17	288	Transcr. regulator, AraC family	gb|EAV71364	*Clostridium cellulolyticum* H10	8.E-101	AraC-type DNA-binding domain-containing proteins, COG2207	Bacterial regulatory helix-turn-helix proteins, AraC family	Intracellular	No pathway
18	361	Choloylglycine hydrolase family prot.	ref|YP_001787648	*Clostridium botulinum* Loch Maree Type A3	3.E-57	Predicted choloylglycine hydrolase, COG4927	Acyl-coenzyme A:6-aminopenicillanic acid acyl-transferase	Intracellular	No pathway

[a]ORFs possibly involved in mucin degradation are indicated with red boxes
[b]Location prediction and number of transmembrane helices as predicted by the program LocateP

FIGURE 7-4 Overview of the 18 ORFs annotated in the DNA insert of mucin-utulising fosmid 53H12 comprising 25,605 bp. Grayscale used is corresponding to the alignment scores with the best BlastP hit: 80 < gray < 199; and black >200. The description of the annotated ORFs in fosmid 53H12, including the observed sequence similarities and predicted protein locations, are depicted in the table. For all annotated genes with highest similarities with putative uncharacterized proteins, additional PSI-Blast results are shown in gray. The black box indicates a gene encoding a protein possibly involved in mucin metabolism. The gray boxes indicate genes that were homologous to genes on transposon *Tn*1549.

In conclusion, our results indicate that a functional screening strategy is suitable for high throughput analysis of a complete metagenomic library for mucin-utilising fosmids. The activity of N-acetyl-D-β-galactosaminidase could be detected extracellularly, whereas the highest activity of N-acetyl-D-β-glucosaminidase was detected intracellularly in the *E. coli* host. Three putative genes possibly involved in the utilisation of mucin could be detected by sequence analysis of the fosmid inserts: putative genes with sequence similarity to a glycosyltransferase and an inner-membrane sugar transporter (probably originating from *Bacteroides*) in fosmid 53C5 and a soluble lytic murein transglycosylase (identical to a gene from *Enterococcus faecalis*) in fosmid 53H12. In addition, the latter fosmid was found to contain five genes derived from the Tn*1549*-like transposons originally identified in *E. faecalis* but recently detected in high abundance in the faecal microbiota of a Japanese population [95] .

ACKNOWLEDGEMENTS

We are grateful to Maarten van de Guchte from INRA for usefull discussions concerning the screening strategy used. In addition, we would like to thank Roland Siezen and Michiel Wels from the Nijmegen Centre for Molecular Bioinformatics (CMBI) for assistance with the gene characterisation and annotation. Miaomiao Zhou from CMBI is thanked for his help with the cellular location prediction of the proteins. Also we would like to thank Muriel Derrien from the Laboratory of Microbiology in Wageningen for usefull remarks and discussions regarding bacterial mucin degradation. This work has been carried out with the financial support of a Marie Curie fellowship in the 6th Framework Programme within the LABHEALTH project (MEST-CT-2004-514428).

CHAPTER 8

Summary, concluding remarks & future perspectives

The research described in this thesis aimed at an improved understanding of the diversity and function of the microbiota residing in the human small intestine. In the following paragraphs, the major findings will be highlighted and discussed. Finally, possible directions are outlined for future research to relate the characteristics of the small intestinal microbiota and human health.

Evaluation of small intestinal model systems

Research focussing on the small intestinal microbiota is seriously hampered by the location of the small intestine in the normal human anatomy, as samples cannot be obtained without the use of invasive procedures via the oral or anal route. In addition to this invasive sample collection, pre-treatments with solutions that evacuate the bowel contents most likely have an effect on the ileal microbiota, especially the luminal microbiota will be almost completely removed by such treatments. In contrast, subjects with an ileostomy provide easy non-invasive access to the luminal contents of the ileum due to absence of the colon. The first and most extensive report that investigated the microbial diversity in the luminal ileum of ileostomy subjects dates back to 1967 [66]. In this study, the microbiota composition of ileostomy effluent samples, ileal perfusion samples obtained by peroral intubation and faecal samples was compared based on classical plating. This showed that lactobacilli, streptococci, staphylococci, fungi and total anaerobes could be cultured in higher numbers from ileostomy effluent compared to ileal perfusion samples, although the ratio between aerobes and anaerobes were comparable in both sample types [66]. As described in **Chapter 3**, the diversity and temporal stability of the predominant bacteria over a period of 28 days were studied in ileostomy effluent samples from five healthy and two medicated ileostomy subjects and one subject with an ileorectal pouch using culture independent small subunit ribosomal RNA (16S rRNA) gene based approaches. Overall, the total number of bacteria within these ileostomy effluents was in the range of 10^7-10^8 bacteria per gram and dominated by bacteria related to the genera *Streptococcus* and *Veillonella* and some Clostridial groups. The only other culture-independent study that aimed at a description of the microbiota of the human ileal lumen performed so far used samples obtained from sudden death victims [72]. The microbiota composition of the examined three diseased individuals with an average age of 81 years showed remarkably high numbers of Proteobacteria and relatively low abundance of *Clostridium*-related species in comparison to our findings in the ileostomy effluents of seven healthy subjects with an average age of 59 (see also **Chapter 3**). Besides the difference in age, differences in diet (ileostomy subjects consume more water and salt compared to subjects with a normal anatomy (dr. F. Troost, pers. comm.)) and health conditions of the subjects may also explain the distinct microbiota compositions. However, apart from other stochastic and methodological factors, this apparent discrepancy in composition could very well be explained by the difference in relapsed time between sample collection and processing in both studies, since the samples from the sudden death victims were collected several hours after death, whereas those from the ileostomy effluent were directly frozen on dry ice upon discharge from the body. The importance of direct and correct sample collection and processing is discussed in **Chapter 4** in which several protocols concerning sample storage, RNA isolation and mRNA enrichment were compared and contrasted. This study resulted in an optimised protocol for mRNA-based approaches for samples that can even be collected at the homes of the volunteers.

Ileal perfusion is another approach that can be used for obtaining samples from the human small intestine. To compare and contrast perfusion samples with those from ileostomy subjects, samples were collected from an individual that was intubated with a double-lumen tube via a catheter, from which one tube collects sample fluid while the other tube is used for perfusion with physiological salt solution proximal to the sample collection site (courtesy of prof. dr. R.J.

Brummer and dr. F. Troost). Though still quite invasive, collection of samples from the human body over a small time range is feasible without the need of prior bowel evacuation. The volunteer was a healthy 26-year-old male with normal GI tract anatomy and had not been subjected to antibiotic treatment 6 months prior to perfusion. The phylogenetic fingerprints of two ileal perfusion samples collected in time (65 minutes and 109 minutes after the start of the perfusion) from this volunteer were determined using HITChip analysis and compared to the phylogenetic profiles of two ileostomy effluent samples and two faecal samples of healthy adults. Cluster analysis (Fig. 8-1) indicated that the microbiota found in the ileal perfusion samples resembled the faecal microbiota more than the ileostomy effluent microbiota in terms of phylotype composition and diversity (Simpson's diversity indices of 137 ± 3.6, 136 ± 64.5 and 39 ± 7.5 for ileal perfusion, faecal and ileostomy effluent samples, respectively).

The same observation was made for all ileostomy effluent samples analysed previously by HITChip (**Chapter 3**). Nevertheless, several phylotypes belonging to the clusters Bacilli and *Clostridium* clusters I and XI were found in common between ileal perfusion and ileostomy effluent samples. However, typical colonic bacteria, such as Bacteroidetes and *Clostridium* clusters IV and XIVa were present in the ileal perfusion samples to a much higher extent as compared to the ileostomy effluent samples. This difference between the microbiotas of ileostomy effluent and ileal perfusion samples might be influenced by factors such as oxygen concentration. Strictly anearobic bacteria, like those belonging to the genera *Bacteroides* and *Clostridium*, are unlikely to be found under the potential (micro)aerobic conditions that are present at the terminal part of the ileum in case of an ileostomy.

Several other considerations related to the sampling should be taken into account when comparing the ileal perfusion samples to the ileal effluent samples. First of all, the tip of the sampler had passed through the ileocecal junction prior to the sampling as was observed by fluoroscopy. Although the tubes were air-flushed prior to sampling, it cannot be ruled out that colonic contents rather than ileal contents were sampled. In addition, retraction of the catheter tip from the region beyond the ileocecal junction may have led to relatively high levels of colonic reflux into the terminal ileum, thereby mixing the locally residing microbiotas. Furthermore, it is possible that the sampling tubes scratched along the mucosal layer of the ileum rather than sampling the luminal contents. However, comparison of the microbial diversity retrieved from ileal mucosa biopsies studied by clone library analyses [216, 217] with ileal perfusion samples on the other hand, revealed that especially the relative contribution of species belonging to the phylum Bacteroidetes was larger in the mucosa biopsy samples (50% and 37% of the retrieved clones) compared to ileal perfusion samples (23% on average). Moreover, both mucosa biopsy samples were collected after colon lavage with Picoprep™ or Phosphoral®. Effect of pre-treatment with these two solutions on the ileal microbiota is unknown; both were found to be as effective in rinsing the colon contents [182]. Further evaluation of ileal mucosa samples by in-depth phylotype characterisation, as well as inclusion of more ileal perfusion samples in the comparison, is necessary for evaluation of the ileal perfusion samples as an alternative to the invasive biopsy collection for obtaining ileal mucosa samples.

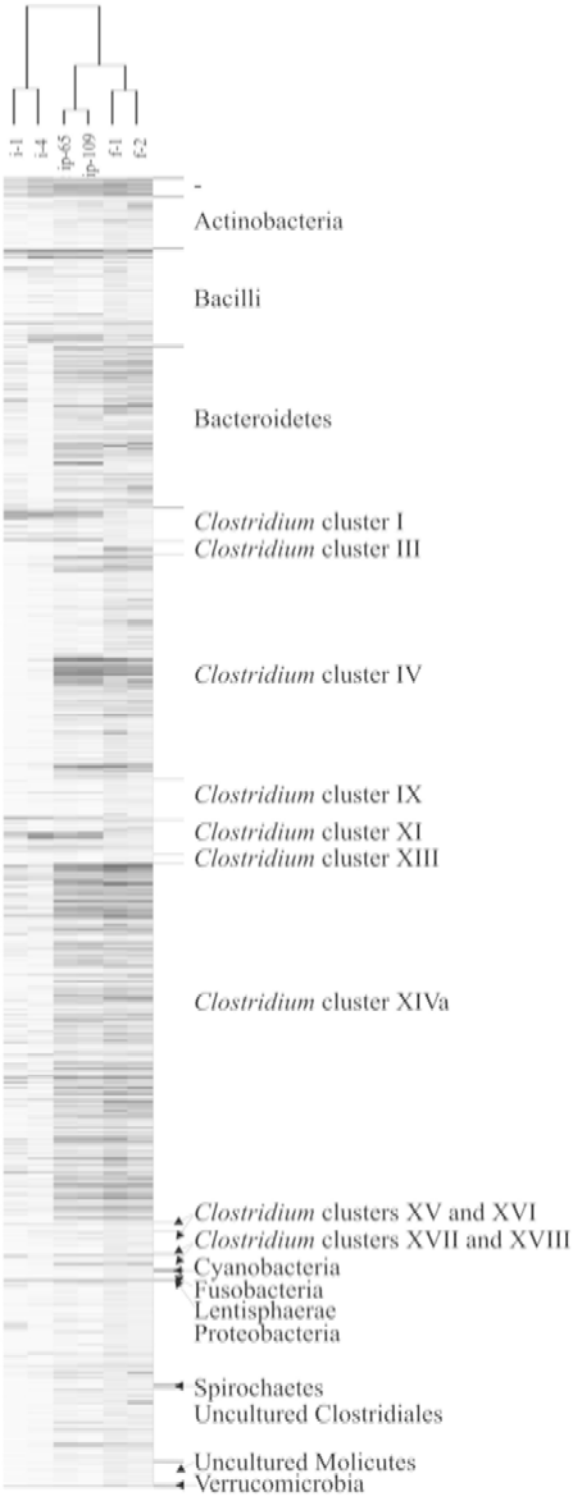


FIGURE 8-1 The two microbiota profiles of the ileal perfusion samples (ip-65 and ip-109 from a 26-year-old male) were analysed by HITChip and compared to the microbiota of effluent profiles of two healthy ileostomy subjects (i-1: a 55-year-old female; i-4: a 74-year-old male) and two faecal samples of healthy adults (f-1: a 54-year-old female; f-2: a 41-year-old female).

In conclusion, ileostomy subjects provide the only accessible source for collecting luminal samples in a non-invasive way over time at the moment. Therefore, ileostomy subjects provide an important and attractive model system that enables the analysis of temporal variation of the microbiota in the ileum (see also below). Finally, these base-line studies provide the starting point for follow-up studies where the effect of diet on the intestinal microbiota can be determined at the site where food has the primary large-scale interactions with the host and its microbes.

The diversity and stability of the small intestinal microbiota

In **Chapter 3** several culture-independent techniques were used to investigate the human luminal ileum microbiota diversity and the temporal stability. As a model for the human ileum, ileal effluent samples from subjects with an ileostomy were used for in-depth diversity analysis. Inter-individual differences in the diversity and temporal stability of the microbiota profiles were observed using HITChip analysis (**Chapter 3**). This indicated that each ileostomy subject had a host-specific ileal microbiota, as was demonstrated for the colonic microbiota previously [230]. However, the species diversity in ileostomy effluent was completely different from that of the colon with bacteria related to the genera *Streptococcus* and *Veillonella* and some Clostridial groups as being the dominant species. Typical faecal bacteria such as those belonging to the Bacteroidetes and *Clostridium* clusters IV and XIVa were hardly detected in ileostomy effluent (**Chapter 3**). Moreover, the microbiota profiles of all ileostomy subjects showed considerable fluctuations even within a single day, which is in clear contrast with the relatively high stability of the microbiota composition in the colon [209, 230]. These daily fluctuations of the microbiota profiles seemed to dangle around a group of stable colonisers, as was hypothesised for the colon before [41]. Moreover, inclusion of a sample collected one year prior to the other profiled samples collected over time for ileostomy subject i4, revealed that the bacterial fluctuations over a period of one year were similar to the daily fluctuations (Fig. 8-2). The average similarity coefficient between the ileal effluent microbiota profiles analysed over a period of eight days with the profile of the sample collected one year before was 47%, which is similar to the average similarity coefficient between the samples from the 8-day period (55%), indicating that about half of the microbiota remains stable in terms of diversity and relative abundance over short-term periods (one week) as well as long-term periods (one year). To investigate how these relative big fluctuations in microbiota composition related to the functions performed by the microbiota, subsequent metatranscriptome analysis of ileostomy effluent microbiota samples collected over time was performed as outlined below.

Metatranscriptome analysis starts with correct sample collection, and subsequent accurate extraction of a representative fraction of RNA is essential for studying the metatranscriptome of a microbial community, such as the GI tract microbiota. As described in **Chapter 4**, an RNA isolation procedure was optimised that was suitable for diverse GI tract samples, such as ileostomy effluent and faecal samples. Additional mRNA enrichment proved to be necessary from pilot cDNA-AFLP fingerprints and subsequent sequence analysis. The rRNA-probe based enrichment of mRNA was found most efficient in rRNA removal. A four-fold increase in the relative contribution of the mRNA molecules to the total pool of RNA was observed for RNA extracted from GI tract samples (**Chapters 5 and 6**). Enriched mRNA retrieved in this way from

ileostomy effluent and faecal samples have served as targets for mRNA-based approaches, showing their use in the functional analysis of GI tract communities. The enrichment used was based on probes targeting the conserved regions of 16S and 23S rRNA molecules of a set of microbes (listed on http://www.ambion.com/techlib/misc/microbe.html). Based on the rRNA sequences obtained from the ileostomy effluent cDNA-AFLP profiles, we anticipate that the mRNA enrichment can even be further improved for microbial ecosystems that are dominated by one or a few bacterial species as 33% of the retrieved rRNA sequences obtained by cDNA-AFLP analysis of enriched mRNA from ileostomy samples originated from *Streptococcus* species (**Chapter 6**).

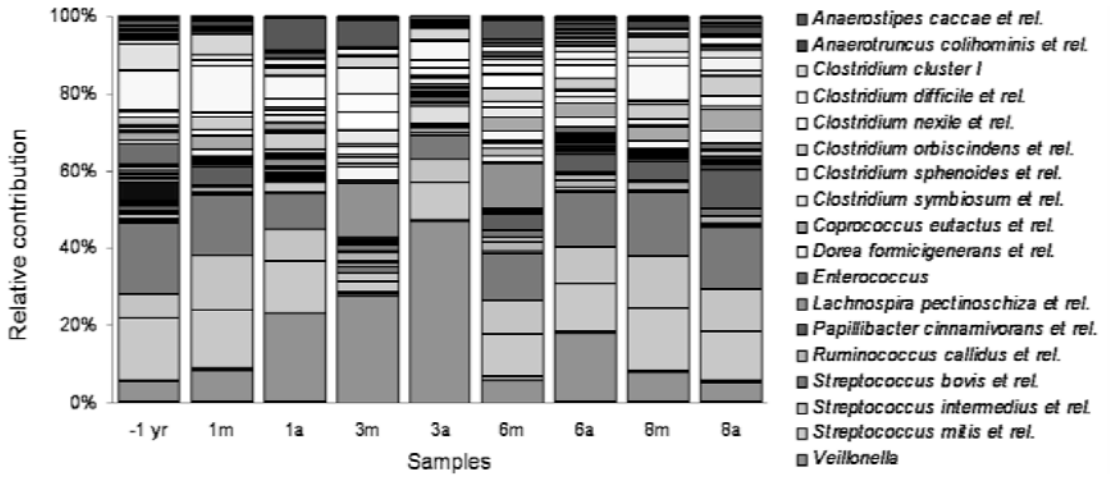


FIGURE 8-2 Relative contribution of the detected phylotypes with HITChip in nine ileostomy effluent samples collected over time of a healthy ileostomy subject (i4). The day of sampling is indicated under the bars, whereas m = morning and a = afternoon. In the legend, phylotypes that contribute at least 1% to the majority of the profiles are indicated. A colour version of this picture can be found in de supplementary and colour figures section at the end of this thesis.

cDNA-AFLP profiling was chosen for profiling the metatranscriptome of the ileostomy effluent and faecal microbiota. Since no prior sequence information is needed and temporal profiles can be compared relatively easy to reveal changes in gene expression over time, this approach was thought to be very suitable for metatranscriptome analysis of the human GI tract microbiota. Evaluating the cDNA-AFLP profiling in retrospect shows that the approach is indeed straightforward once mRNA of sufficient quality is obtained. Although only a fraction of the metatranscriptome was profiled due to the use of selective primers, this should be representative for the range of transcripts expressed by the microbiota, since the selective primer-subset was randomly chosen. The cDNA-AFLP approach for studying the *in situ* gene expression and its dynamics was validated for samples from the colon (**Chapter 5**). cDNA-AFLP analysis of mRNA-enriched RNA extracts obtained from faecal samples of two healthy individuals collected at two different time points was performed in duplicate. Comparison of duplicate profiles based on Pearson correlation revealed that the technical reproducibility of the mRNA enrichment and cDNA-AFLP profiling is very high (similarity indices of approximately 95%). The technical

reproducibility of duplicate DNA and RNA extraction proved to be lower (similarity index of 88% for DNA and 62% for total RNA) which could be due to the difficulty of homogenisation of samples in the storage buffer directly after collection. Moreover, this study clearly demonstrated that the microbiota of one of the two individuals was more stable in composition while the dynamics of their transcriptomes were similar. Last but not least, cDNA-AFLP allows post-analytic extraction and sequencing of the cDNA fragments from the gel with a success rate of approximately 60%. When combined with other functional metagenomic approaches, such as the one described in **Chapter 7** for the mucin degrading capacity of the metagenome, insight in the functional roles exerted by the human small instestinal microbiota can be increased, and possibly correlated with health and disease.

cDNA-AFLP analysis of faeces indicated a considerable fluctuation of the gene-expression over a period of three days (**Chapter 5**). Nevertheless and as expected, fluctuations in gene-expression patterns were more prominent in the highly dynamic microbiota of the ileum as was observed with samples taken over a two day period in a healthy ileostomy effluent subject (**Chapter 6**) with only six protein encoding transcripts being detected at both time points. This is in line with microbiota composition observed for ileostomy subjects which is already fluctuating within a day (**Chapter 3**). However, despite the small overlap on the transcript level, the distribution over functional gene classes that were expressed on two consecutive days by this dynamic microbiota was remarkably similar. In addition, phylogenetic profiling indicated that species of the genus *Streptococcus* contributed about 40% to the total microbiota (**Chapter 3**) while 70% of all protein-encoding transcripts with significant sequence similarity to public database entries had the highest similarity with genes of the genus *Streptococcus* (**Chapter 6**). This clearly demonstrates that the dominant populations of the ileostomy effluent appear also to be the most active ones.

Phylogenetic and Functional Cores in the GI tract

It has been hypothesised that each individual harbours a core of phylogenetic groups of bacteria that can be found in the colon at any time point during the lifespan of healthy adults [41, 234]. To analyse a potential individual core in the human ileum, samples collected from a single subject with an ileostomy over a one-year time span were compared based on HITChip profiling. This revealed a core group comprised of 24 different phylogenetically related groups (Fig. 8-3a). In addition, a hypothetical common core for all ileostomy subjects was investigated by comparing all the phylogenetic fingerprints of ileal effluent samples collected over time (23 ileal effluent samples from five ileostomy subjects). This analysis revealed that in total eight phylogenetic groups formed the common core, with species belonging to the genera *Streptococcus*, *Veillonella* and *Clostridium* as most abundant groups in all subjects investigated (contributing on average for 11%, 12% and 3%, respectively to the total probe signal intensity), whereas *Oxalobacter formigenes*-related, *Coprobacillus catenaformis*-related and *Enterococcus* species were detected with low relative abundance (less than 1% in all analysed profiles), indicative that the "common core" is not only determined by the abundant species (Fig. 8-3b).

It has also been suggested that a common core should not be formulated at the species level, but rather at the genetic level, by defining the function of microbial genes other than those

involved in housekeeping functions [204]. The existence of a stable core in individuals suggests a high degree of functional redundancy in the GI tract microbiota. In the ileostomy effluent metatranscriptome, we observed that 64% of a selection of retrieved transcripts, highly homologous gene fragments could be detected in the DNA extracted from the ileostomy effluent from four other subjects (**Chapter 6**), which is indicative for a stable functional core that might be in common to the ileum of these subjects. In the following section the contribution of the present research to an improved insight into the as yet largely unknown functions exerted by the human intestinal microbiota are outlined.

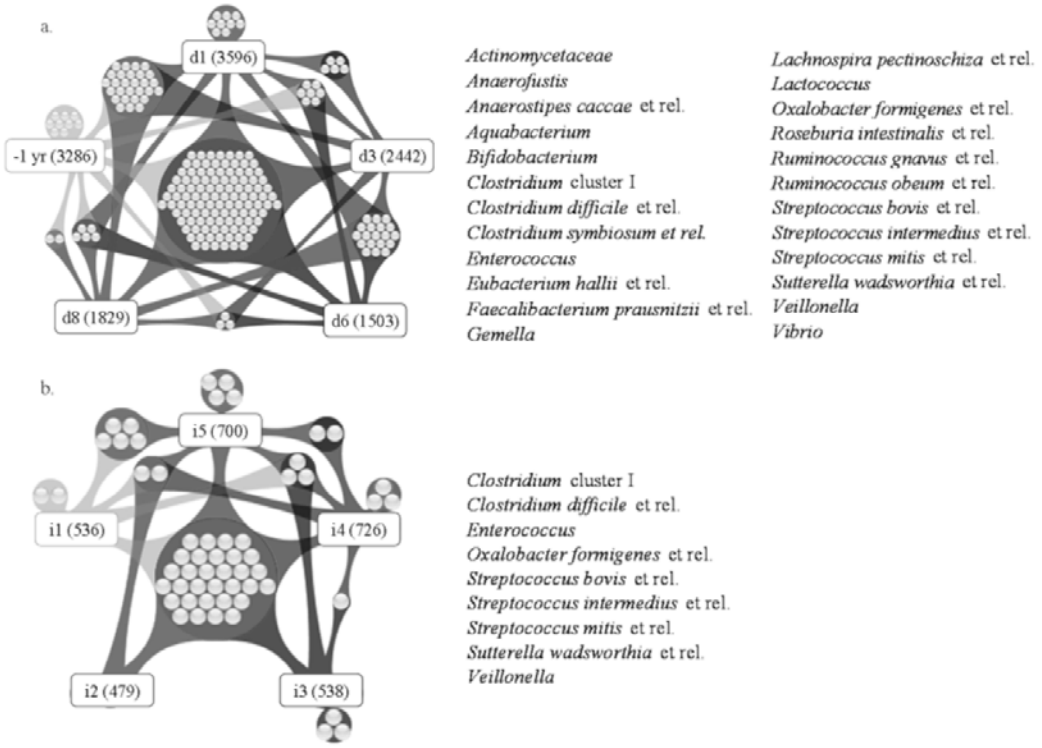


FIGURE 8-3 (**a**) Analysis of an hypothetical individual core among the microbiota of one subject with an ileostomy over a period of one year and a common ileal microbiota core shared by five ileostomy subjects over time (**b**) depicted in a Venn diagram. To determine the "common core" microbiota the probe signals of 23 available ileostomy effluent samples were included for comparison. The lists next to the Venn diagrams display the phylotypic groups that were found in common in the individual core (top) and the "common core" (bottom). The abbreviations i1 till i5 refer to the five subjects with an ileostomy included in the analysis of the common core. The numbers between parentheses indicate the total number of responding probes, whereas each sphere represents ten responding probes.

Genes involved in carbohydrate metabolism were abundantly found in the metatranscriptomes of one ileostomy subject (**Chapter 6**), but also in the faecal metatranscriptomes of two healthy adults (**Chapter 5**). From the ileostomy effluent samples, a broad range of transcripts encoding proteins involved in carbohydrate metabolism and energy production was found, including transporters, transferases, hydrolases and isomerases, involved in for example the pentose phosphate pathway (**Chapter 6**). Moreover, 65% of the transcripts involved in carbohydrate

metabolism and energy production were expressed by *Streptococcus*-like species, as was based on phylogenetic profiling and sequence similarity analyses of transcript sequences. This phylogenetic group is characterised by the capacity for very fast sugar fermentation and highly efficient carbohydrate transport systems, making this group well equipped competitors for substrates during the fast transit through the small intestine. Most likely, the environmental factors governing the luminal site of the ileum play a very important determining role in the microbiota composition as well as on the functions that are exerted.

Genes encoding proteins that are involved in the carbohydrate metabolism were also highly expressed in faeces, making up 26% of all sequenced transcripts (**Chapter 5**). This observation is in line with a recently reported metaproteomics study on the faecal microbiota [211] and suggests that carbohydrate metabolism is also an important function exerted by the microbiota in the colon. However, clear differences were observed beween the genes expressed by the ileostomy effluent microbiota and that of faeces (**Chapters 5 and 6**). Genes encoding proteins of for instance the pentose phosphate pathway of *Streptococcus* species were mainly found in ileostomy effuent while the carbohydrate metabolism-related transcripts from faeces originated almost all from species of the order *Clostridiales*, and encoded proteins involved in glycolysis and pyruvate synthesis (three times a glyceraldehyde-3-P dehydrogenase, and three times a pyruvate ferredoxin-oxidoreductase). In contrast to the ileum, the colon has a low transit rate and as a result the simple sugars are completely absorbed at the end of the colon, leaving the faecal microbes like for example *Clostridium* cluster XIVa with the more complex molecules for fermentation. Overall, these results indicate that the ileal and colonic microbiota are dominated by different active microbial groups, dictated by the differences in their functional niche.

Mucosa versus lumen

The aforementioned results were all obtained from the luminal side of the human small intestine or colon. Besides the lumen the mucosal layer is also a very important niche for the microbiota as it provides a steady carbon and nitrogen source to the microbes residing in there, mostly in the form of mucin. To investigate the mucin degrading capacity of this mucosal microbiota, a metagenomic library consisting of more than 20,000 fosmids was constructed from mucosal biopsy samples obtained from a healthy volunteer, and subsequently screened for its mucin degradative capacity by growth and enzyme analysis of all the fosmid clones (**Chapter 7**). This revealed that 1.1% of the fosmid clones were able to use mucin as a sole carbon source. Subsequent sequence analysis of the fosmid inserts of two clones that conferred high N-acetylgalactosaminidase and N-acetylglucosaminidase enzyme activity on the cloning host (*E. coli*) revealed three putative genes that are likely to be involved in mucin degradation.

The first fosmid encoded a glycosyltransferase and an inner-membrane sugar transporter, both originating from *Bacteroides* (**Chapter 7**). Interestingly, *Bacteroides* was not detected in the ileal effluent samples (**Chapter 3**), while this species was detected in the ileal mucosa before [216, 217]. This may suggest that *Bacteroides* species in the small intestine depend mainly on the mucin secreted by the host and previous gnotobiotic mouse studies revealed that specific host-microbe interactions are important for this interplay between mucin release and degradation [121].

The annotated ORFs in the other fosmid included a predicted soluble lytic murein transglycosylase gene from *Enterococcus faecalis* (**Chapter 7**). Remarkably, *Enterococcus* is a genus that has not been associated with mucin degradation before. This genus was detected in the hypothetical core of the five ileostomy subjects, though in low abundance (**Chapter 3**), indicating that it might be a generalist that can inhabit diverse niches in the human ileum.

Overall our functional screening of the ileal metagenome indicates that genes involved in mucin-degradation are ubiquitous in ileal microbes and can be derived from Gram-negative as well as Gram-positive bacteria.

Future perspectives

The work described in this thesis gives the first insight into the diversity, population dynamics and function of the small intestinal microbiota based on molecular analyses. This microbiota is less diverse and less stable in composition compared to the colon and is dominated by fastidious bacteria, such as streptococci that are likely to play an important role in the carbohydrate metabolism in the lumen of the ileum. In the near future, the research field focussing on the human GI tract microbiota is expected to develop towards meta-analyses of microbial ecosystem, by combining multiple approaches such as metagenomics in combination with metatranscriptomics [63] and metaproteomics. As metatranscriptomics or metaproteomics [94, 211] will reveal functional activities of the complete ecosystems, the detected genes or proteins found expressed *in situ* can be used as targets to trace back the corresponding microbe in the metagenomic library. A potentially suitable approach for this was recently investigated by targeting gene sequences that were found expressed *in situ* in the human ileostomy effluent samples using RING-FISH probes (S. El-Aidy, pers. comm.), as the suitability of RING-FISH to environmental samples other than the GI tract was demonstrated recently [150]. In such a combined approach the metagenome will merely function as a gene catalogue for the genetic potential of the active species present. Thereby sequence-based analysis will result in an overview of the genetic diversity of the microbiota and aid in the possible elucidation of a 'functional core'. In addition, comparative metagenomics combined with advanced bioinformatics software will make large scale comparisons of metagenomes possible. Subsequent post-metagenomic tools, will offer research possibilities for the analysis of spatial and temporal microbiota variation without the need of enormous sequencing efforts. This will result in the detection of genes common between the small intestine and large intestine, and derived from that, genes specific for species in either region of the GI tract. This knowledge could subsequently be used for the development of food ingredients and products or pharmaceutical components that can stimulate certain microorganisms or enhance specific microbial functionality at predestined sites in the GI tract. For example, diagnostic tools could be developed to determine the unique microbial profile of patients suffering from intestinal diseases like UC and Crohn's disease [118] and obesity [91, 105]. Derived from that, a justified choice for probiotic strains could be made that might improve the quality of life of these individuals [194].

The detection of many novel genes and gene clusters by function-driven analysis tools in the few studies performed will poise researchers to search for more high-throughput screening tools to analyse the huge functional potential of diverse microbial ecosystems, such as the human

GI tract. For example, a riboflavin-producing fosmid clone could be identified from an ileal mucosa library originating from *Faecalibacterium prausnitzii* recently (Booijink *et al.*, in prep.). For sure these type of functional approaches will provide new enzymes for new biotechnological applications, but also allow functional assignment of many proteins found in abundance in the databases currently designated as 'hypothetical' or 'conserved hypothetical' proteins. Also novel high-throughput sequencing techniques [6] will increase the number of analyses of bacterial genomes, metagenomic libraries as well as metatranscriptomes at low cost [56]. The next challenge will be the translation of findings from large studies (like the Human Microbiome Project and consortia like MetaHIT) into effective interventions that can improve health, possibly progressing towards personalised diet adaptations, tuned to an individuals residing microbiota and its functionality [116].

LITERATURE CITED

1. Ahmed, S., Macfarlane, G.T., Fite, A., McBain, A.J., Gilbert, P. & Macfarlane, S. 2007. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl. Environ. Microbiol.* **73**:7435-7442.
2. Alifano, P., Bruni, C.B. & Carlomagno, M.S. 1994. Control of mRNA processing and decay in prokaryotes. *Genetica.* **94**:157-172.
3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
4. Andersson, H. 1992. The ileostomy model for the study of carbohydrate digestion and carbohydrate effects on sterol excretion in man. *Eur. J. Clin. Nutr.* **46**:S69-76.
5. Arif, N., Do, T., Byun, R., Sheehy, E., Clark, D., Gilbert, S.C. & Beighton, D. 2008. *Veillonella rogosae* sp. nov., an anaerobic, Gram-negative coccus isolated from dental plaque. *Int. J. Syst. Evol. Microbiol.* **58**:581-584.
6. Asmann, Y.W., Wallace, M.B. & Thompson, E.A. 2008. Transcriptome profiling using next-generation sequencing. *Gastroenterology.* **135**:1466-1468.
7. Bachem, C.W., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. & Visser, R.G. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J.* **9**:745-753.
8. Bachem, C.W.B., Oomen, R.J.F.J. & Visser, R.G.F. 1998. Transcript Imaging with cDNA-AFLP: A Step-by-Step Protocol. *Plant Mol. Biol. Rep.* **16**:157-173.
9. Baker, G.C., Smith, J.J. & Cowan, D.A. 2003. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods.* **55**:541-555.
10. Bell, A.E., Sellers, L.A., Allen, A., Cunliffe, W.J., Morris, E.R. & Ross-Murphy, S.B. 1985. Properties of gastric and duodenal mucus: effect of proteolysis, disulfide reduction, bile, acid, ethanol, and hypertonicity on mucus gel structure. *Gastroenterology.* **88**:269-280.
11. Ben-Amor, K., Heilig, H., Smidt, H., Vaughan, E.E., Abee, T. & de Vos, W.M. 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.
12. Bensch, S. & Akesson, M. 2005. Ten years of AFLP in ecology and evolution: why so few animals? *Mol. Ecol.* **14**:2899-2914.
13. Berg, J.O., Lindqvist, L., Andersson, G. & Nord, C.E. 1983. Neuraminidase in *Bacteroides fragilis*. *Appl. Environ. Microbiol.* **46**:75-80.
14. Berg, R.D. 1996. The indigenous gastrointestinal microflora. *Trends Microbiol.* **4**:430-435.
15. Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A., Leelaporn, A. & Skurray, R.A. 1998. Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. *J. Bacteriol.* **180**:4350-4359.
16. Bik, E.M., Eckburg, P.B., Gill, S.R., Nelson, K.E., Purdom, E.A., Francois, F., Perez-Perez, G., Blaser, M.J. & Relman, D.A. 2006. Molecular analysis of the bacterial microbiota in the human stomach. *Proc. Natl. Acad. Sci. USA.* **103**:732-737.
17. Blaut, M. 2002. Relationship of prebiotics and food to intestinal microflora. *Eur. J. Nutr.* **41**:I11-16.
18. Booijink, C.C., Zoetendal, E.G., Kleerebezem, M. & de Vos, W.M. 2007. Microbial communities in the human small intestine: coupling diversity to metagenomics. *Future Microbiol.* **2**:285-295.
19. Bottone, E.J., Patel, L., Patel, P. & Robin, T. 1998. Mucoid encapsulated *Enterococcus faecalis*: an emerging morphotype isolated from patients with urinary tract infections. *Diagn. Microbiol. Infect. Dis.* **31**:429-430.
20. Boureau, H., Decre, D., Carlier, J.P., Guichet, C. & Bourlioux, P. 1993. Identification of a *Clostridium cocleatum* strain involved in an anti-*Clostridium difficile* barrier effect and determination of its mucin-degrading enzymes. *Res. Microbiol.* **144**:405-410.
21. Bratten, J.R. & Jones, M.P. 2007. Small intestinal motility. Curr. Opin. Gastroenterol. **23**:127-133.
22. Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P. & Rohwer, F. 2003. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**:6220-6223.

23. Bron, P.A., Grangette, C., Mercenier, A., de Vos, W.M. & Kleerebezem, M. 2004. Identification of *Lactobacillus plantarum* genes that are induced in the gastrointestinal tract of mice. *J. Bacteriol.* **186**:5721-5729.
24. Bron, P.A., Marco, M., Hoffer, S.M., Van Mullekom, E., de Vos, W.M. & Kleerebezem, M. 2004. Genetic characterization of the bile salt response in *Lactobacillus plantarum* and analysis of responsive promoters *in vitro* and *in situ* in the gastrointestinal tract. *J. Bacteriol.* **186**:7829-7835.
25. Bron, P.A., Meijer, M., Bongers, R.S., de Vos, W.M. & Kleerebezem, M. 2007. Dynamics of competitive population abundance of *Lactobacillus plantarum ivi* gene mutants in faecal samples after passage through the gastrointestinal tract of mice. *J. Appl. Microbiol.* **103**:1424-1434.
26. Byun, R., Carlier, J.P., Jacques, N.A., Marchandin, H. & Hunter, N. 2007. *Veillonella denticariosi* sp. nov., isolated from human carious dentine. *Int. J. Syst. Evol. Microbiol.* **57**:2844-2848.
27. Byun, R., Nadkarni, M.A., Chhour, K.L., Martin, F.E., Jacques, N.A. & Hunter, N. 2004. Quantitative analysis of diverse *Lactobacillus* species present in advanced dental caries. *J. Clin. Microbiol.* **42**:3128-3136.
28. Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., Garrity, G.M. & Tiedje, J.M. 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33**:D294-296.
29. Corfield, A.P., Wagner, S.A., Clamp, J.R., Kriaris, M.S. & Hoskins, L.C. 1992. Mucin degradation in the human colon: production of sialidase, sialate O-acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infect. Immun.* **60**:3971-3978.
30. Daniel, R. 2004. The soil metagenome-a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* **15**:199-204.
31. De Vries, M.C. 2006. Analyzing global gene expression of *Lactobacillus plantarum* in the human gasto-intestinal tract. PhD thesis Wageningen University, Wageningen, the Netherlands.
32. Dennis, P., Edwards, E.A., Liss, S.N. & Fulthorpe, R. 2003. Monitoring gene expression in mixed microbial communities by using DNA microarrays. *Appl. Environ. Microbiol.* **69**:769-778.
33. Derensy-Dron, D., Krzewinski, F., Brassart, C. & Bouquelet, S. 1999. Beta-1,3-galactosyl-N-acetylhexosamine phosphorylase from *Bifidobacterium bifidum* DSM 20082: characterization, partial purification and relation to mucin degradation. *Biotechnol. Appl. Biochem.* **29**:3-10.
34. Derrien, M. 2007. Mucin utilisation and host interactions of the novel intestinal microbe *Akkermansia muciniphila*. PhD thesis Wageningen University, Wageningen, the Netherlands.
35. Derrien, M., Vaughan, E.E., Plugge, C.M. & de Vos, W.M. 2004. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**:1469-1476.
36. Dethlefsen, L., Eckburg, P.B., Bik, E.M. & Relman, D.A. 2006. Assembly of the human intestinal microbiota. *Trends Ecol. Evol.* **21**:517-523.
37. Dethlefsen, L., McFall-Ngai, M. & Relman, D.A. 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature.* **449**:811-818.
38. Diaz-Torres, M.L., Villedieu, A., Hunt, N., McNab, R., Spratt, D.A., Allan, E., Mullany, P. & Wilson, M. 2006. Determining the antibiotic resistance potential of the indigenous oral microbiota of humans using a metagenomic approach. *FEMS Microbiol. Lett.* **258**:257-262.
39. DiGiulio, D.B., Romero, R., Amogan, H.P., Kusanovic, J.P., Bik, E.M., Gotsch, F., Kim, C.J., Erez, O., Edwin, S. & Relman, D.A. 2008. Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation. *PLoS ONE.* **3**:e3056.
40. Domingo, M.C., Huletsky, A., Bernal, A., Giroux, R., Boudreau, D.K., Picard, F.J. & Bergeron, M.G. 2005. Characterization of a Tn5382-like transposon containing the vanB2 gene cluster in a *Clostridium* strain isolated from human faeces. *J. Antimicrob. Chemother.* **55**:466-474.
41. Doré, J. 2006. Presented at the 11th International Symposium on Microbial Ecology - ISME 11; The hidden powers - microbial communities in action, Vienna, Austria, August 20-25.
42. Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. & Relman, D.A. 2005. Diversity of the human intestinal microbial flora. *Science.* **308**:1635-1638.
43. Egert, M., de Graaf, A.A., Smidt, H., de Vos, W.M. & Venema, K. 2006. Beyond diversity: functional microbiomics of the human colon. *Trends Microbiol.* **14**:86-91.

44. Ekkapongpisit, M., Wannatung, T., Susantad, T., Triwitayakorn, K. & Smith, D.R. 2007. cDNA-AFLP analysis of differential gene expression in human hepatoma cells (HepG2) upon dengue virus infection. *J. Med. Virol.* **79**:552-561.
45. Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W.R. 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl. Environ. Microbiol.* **67**:89-99.
46. Evans, D.F., Pye, G., Bramley, R., Clark, A.G., Dyson, T.J. & Hardcastle, J.D. 1988. Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut.* **29**:1035-1041.
47. Falk, P.G., Hooper, L.V., Midtvedt, T. & Gordon, J.I. 1998. Creating and maintaining the gastrointestinal ecosystem: what we know and need to know from gnotobiology. *Microbiol. Mol. Biol. Rev.* **62**:1157-1170.
48. Ferraris, R.P. 2001. Dietary and developmental regulation of intestinal sugar transport. *Biochem. J.* **360**:265-276.
49. Ferrer, M., Beloqui, A., Timmis, K.N. & Golyshin, P.N. 2009. Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* **16**:109-123.
50. Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N. & Golyshin, P.N. 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ. Microbiol.* **7**:1996-2010.
51. Finegold, S.M., Sutter, V.L. & Mathisen, G.E. 1983. Normal indigenous intestinal flora. , p. 3-31. *In* D. J. Hentgens (ed.), Human Intestinal Microflora in Health and Disease. Academic Press, New York, USA.
52. Fitzsimons, N.A., Akkermans, A.D., de Vos, W.M. & Vaughan, E.E. 2003. Bacterial gene expression detected in human faeces by reverse transcription-PCR. *J. Microbiol. Methods.* **55**:133-140.
53. Foerstner, K.U., von Mering, C. & Bork, P. 2006. Comparative analysis of environmental sequences: potential and challenges. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**:519-523.
54. Forster, J. & Forster, G. 1994. Gastrointestinal mucus, p. 1255-1282. *In* L. R. Johnson (ed.), Physiology of the gastrointestinal tract, 3rd ed, vol. 2. Raven Press, New York.
55. Foubert, E.L. & Douglas, H.C. 1948. Studies on the Anaerobic Micrococci: II. The Fermentation of Lactate by *Micrococcus lactilyticus*. *J. Bacteriol.* **56**:35-36.
56. Frank, D.N. & Pace, N.R. 2008. Gastrointestinal microbiology enters the metagenomics era. *Curr. Opin. Gastroenterol.* **24**:4-10.
57. Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A. & Mewes, H.W. 2001. Functional and structural genomics using PEDANT. *Bioinformatics.* **17**:44-57.
58. Fromin, N., Hamelin, J., Tarnawski, S., Roesti, D., Jourdain-Miserez, K., Forestier, N., Teyssier-Cuvelle, S., Gillet, F., Aragno, M. & Rossi, P. 2002. Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environ. Microbiol.* **4**:634-643.
59. Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H. & Yamamoto, K. 2005. Identification and molecular cloning of a novel glycoside hydrolase family of core 1 type O-glycan-specific endo-alpha-N-acetylgalactosaminidase from *Bifidobacterium longum*. *J. Biol. Chem.* **280**:37415-37422.
60. Furrie, E., Macfarlane, S., Thomson, G. & Macfarlane, G.T. 2005. Toll-like receptors-2, -3 and -4 expression patterns on human colon and their regulation by mucosal-associated bacteria. *Immunology.* **115**:565-574.
61. Gabor, E.M., de Vries, E.J. & Janssen, D.B. 2004. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ. Microbiol.* **6**:948-958.
62. Garnier, F., Taourit, S., Glaser, P., Courvalin, P. & Galimand, M. 2000. Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus* spp. *Microbiol.* **146**:1481-1489.
63. Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P. & Joint, I. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE.* **3**:e3042.

64. Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M. & Nelson, K.E. 2006. Metagenomic analysis of the human distal gut microbiome. *Science.* **312**:1355-1359.
65. Go, P.M., van Dieijen-Visser, M.P., Davies, B.I., Lens, J. & Brombacher, P.J. 1988. Microbial flora and bile acid metabolism in patients with an ileal reservoir. *Scand. J. Gastroenterol.* **23**:229-236.
66. Gorbach, S.L., Nahas, L., Weinstein, L., Levitan, R. & Patterson, J.F. 1967. Studies of intestinal microflora. IV. The microflora of ileostomy effluent: a unique microbial ecology. *Gastroenterology.* **53**:874-880.
67. Goso, Y., Ishihara, K., Sugawara, S. & Hotta, K. 2001. Purification and characterization of alpha-L-fucosidases from *Streptomyces* sp. OH11242. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **130**:375-383.
68. Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**:669-685.
69. Hane, B.G., Jager, K. & Drexler, H.G. 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. *Electrophoresis.* **14**:967-972.
70. Hao, W.L. & Lee, Y.K. 2004. Microflora of the gastrointestinal tract: a review. *Methods Mol. Biol.* **268**:491-502.
71. Hayashi, H., Sakamoto, M., Kitahara, M. & Benno, Y. 2003. Molecular analysis of fecal microbiota in elderly individuals using 16S rDNA library and T-RFLP. *Microbiol. Immunol.* **47**:557-570.
72. Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M. & Benno, Y. 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.
73. Heilig, H.G.H.J., Zoetendal, E.G., Vaughan, E.E., Marteau, P., Akkermans, A.D.L. & de Vos, W.M. 2002. Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl. Environ. Microbiol.* **68**:114-123.
74. Higham, S.E. & Read, N.W. 1990. Effect of ingestion of fat on ileostomy effluent. *Gut.* **31**:435-438.
75. Hold, G.L., Pryde, S.E., Russell, V.J., Furrie, E. & Flint, H.J. 2002. Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. *FEMS Microbiol. Ecol.* **39**:33-39.
76. Hooper, L.V., Midtvedt, T. & Gordon, J.I. 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **22**:283-307.
77. Hooper, L.V., Wong, M.H., Thelin, A., Hansson, L., Falk, P.G. & Gordon, J.I. 2001. Molecular analysis of commensal host-microbial relationships in the intestine. *Science.* **291**:881-884.
78. Hoskins, L.C., Agustines, M., McKee, W.B., Boulding, E.T., Kriaris, M. & Niedermeyer, G. 1985. Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J. Clin. Invest.* **75**:944-953.
79. Hoskins, L.C. & Boulding, E.T. 1981. Mucin degradation in human colon ecosystems. Evidence for the existence and role of bacterial subpopulations producing glycosidases as extracellular enzymes. *J. Clin. Invest.* **67**:163-172.
80. Hoskins, L.C., Boulding, E.T. & Larson, G. 1997. Purification and characterization of blood group A-degrading isoforms of alpha-N-acetylgalactosaminidase from *Ruminococcus torques* strain IX-70. *J. Biol. Chem.* **272**:7932-7939.
81. Hung, G.U., Tsai, C.C. & Lin, W.Y. 2006. Development of a new method for small bowel transit study. *Ann. Nucl. Med.* **20**:387-392.
82. Hung, M.N. & Lee, B.H. 2002. Purification and characterization of a recombinant beta-galactosidase with transgalactosylation activity from *Bifidobacterium infantis* HL96. *Appl. Microbiol. Biotechnol.* **58**:439-445.
83. Hunsaker, W.R., Badri, H., Lombardo, M. & Collins, M.L. 1989. Nucleic acid hybridization assays employing dA-tailed capture probes. II. Advanced multiple capture methods. *Anal. Biochem.* **181**:360-370.
84. Iwase, H., Ishii, I., Ishihara, K., Tanaka, Y., Omura, S. & Hotta, K. 1988. Release of oligosaccharides possessing reducing-end N-acetylgalactosamine from mucus glycoprotein in *Streptomyces* sp. OH-11242 culture medium through action of endo-type glycosidase. *Biochem. Biophys. Res. Commun.* **151**:422-428.
85. Jones, B.V. & Marchesi, J.R. 2007. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat. Methods.* **4**:55-61.

86. Jumas-Bilak, E., Carlier, J.P., Jean-Pierre, H., Teyssier, C., Gay, B., Campos, J. & Marchandin, H. 2004. *Veillonella montpellierensis* sp. nov., a novel, anaerobic, Gram-negative coccus isolated from human clinical samples. *Int. J. Syst. Evol. Microbiol.* **54**:1311-1316.
87. Kane, M.D., Poulsen, L.K. & Stahl, D.A. 1993. Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Appl. Environ. Microbiol.* **59**:682-686.
88. Katayama, T., Fujita, K. & Yamamoto, K. 2005. Novel bifidobacterial glycosidases acting on sugar chains of mucin glycoproteins. *J. Biosci. Bioeng.* **99**:457-465.
89. Kelly, S., Yeo, J., Robertson, G.M., Chapman, B., Wells, J.E. & Frizelle, F.A. 2004. Functional assessment of bacterial colonization in patients with ileal pouch-anal anastomosis and Brooke ileostomy. *Dis. Colon Rectum.* **47**:1386-1389.
90. Kibe, R., Sakamoto, M., Yokota, H., Ishikawa, H., Aiba, Y., Koga, Y. & Benno, Y. 2005. Movement and fixation of intestinal microbiota after administration of human feces to germfree mice. *Appl. Environ. Microbiol.* **71**:3171-3178.
91. Kinross, J.M., von Roon, A.C., Holmes, E., Darzi, A. & Nicholson, J.K. 2008. The human gut microbiome: implications for future health care. *Curr. Gastroenterol. Rep.* **10**:396-403.
92. Kitahara, M., Sakata, S., Sakamoto, M. & Benno, Y. 2004. Comparison among fecal secondary bile acid levels, fecal microbiota and *Clostridium scindens* cell numbers in Japanese. *Microbiol. Immunol.* **48**:367-375.
93. Klaassens, E.S., de Vos, W.M. & Vaughan, E.E. 2007. Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl. Environ. Microbiol.* **73**:1388-1392.
94. Klaassens, E.S. 2007. Exploring the functionality of intestinal Bifidobacteria: a post-genomic approach. PhD thesis Wageningen University, Wageningen, the Netherlands.
95. Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V.K., Srivastava, T.P., Taylor, T.D., Noguchi, H., Mori, H., Ogura, Y., Ehrlich, D.S., Itoh, K., Takagi, T., Sakaki, Y., Hayashi, T. & Hattori, M. 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**:169-181.
96. Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**:521-555.
97. Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H. & Rupp, S. 2007. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J. Biotechnol.* **127**:575-592.
98. Lane, D.J. 1991. 16S/23S rRNA sequencing, p. 115-175. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons, Chichester.
99. Langkilde, A.M., Andersson, H., Schweizer, T.F. & Wursch, P. 1994. Digestion and absorption of sorbitol, maltitol and isomalt from the small bowel. A study in ileostomy subjects. *Eur. J. Clin. Nutr.* **48**:768-775.
100. Lee, A. 1984. Neglected niches: the microbial ecology of the gastrointestinal tract, p. 115-162. *In* K. C. Marshall (ed.), Adv. Microb. Ecol., vol. 8. Plenum Press, New York.
101. Lepage, P., Seksik, P., Sutren, M., de la Cochetiere, M.F., Jian, R., Marteau, P. & Doré, J. 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.
102. Leps, J. & Smilauer, P. 2003. Multivariate analysis of ecological data using CANOCO. Cambridge University Press, Cambridge.
103. Leser, T.D., Amenuvor, J.Z., Jensen, T.K., Lindecrona, R.H., Boye, M. & Moller, K. 2002. Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. *Appl. Environ. Microbiol.* **68**:673-690.
104. Levy, G.N. & Aminoff, D. 1980. Purification and properties of alpha-N-acetylgalactosaminidase from *Clostridium perfringens*. *J. Biol. Chem.* **255**:11737-11742.
105. Ley, R.E., Turnbaugh, P.J., Klein, S. & Gordon, J.I. 2006. Microbial ecology: human gut microbes associated with obesity. *Nature.* **444**:1022-1023.

106. Liebich, J., Schadt, C.W., Chong, S.C., He, Z., Rhee, S.K. & Zhou, J. 2006. Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. *Appl. Environ. Microbiol.* **72**:1688-1691.
107. Liu, X. & Gorovsky, M.A. 1993. Mapping the 5' and 3' ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res.* **21**:4954-4960.
108. Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. & Schleifer, K.H. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363-1371.
109. Macdonald, T.T. & Monteleone, G. 2005. Immunity, inflammation, and allergy in the gut. *Science.* **307**:1920-1925.
110. Macfarlane, G.T. & Gibson, G.R. 1991. Formation of glycoprotein degrading enzymes by *Bacteroides fragilis*. *FEMS Microbiol. Lett.* **61**:289-293.
111. Macfarlane, G.T. & Macfarlane, S. 2007. Models for intestinal fermentation: association between food components, delivery systems, bioavailability and functional interactions in the gut. *Curr. Opin. Biotechnol.* **18**:156-162.
112. Macfarlane, G.T., Macfarlane, S. & Gibson, G.R. 1992. Synthesis and release of proteases by *Bacteroides fragilis*. *Curr. Microbiol.* **24**:55-59.
113. Macfarlane, S., Hopkins, M.J. & Macfarlane, G.T. 2001. Toxin synthesis and mucin breakdown are related to swarming phenomenon in *Clostridium septicum*. *Infect. Immun.* **69**:1120-1126.
114. Macfarlane, S., Woodmansey, E.J. & Macfarlane, G.T. 2005. Colonization of mucin by human intestinal bacteria and establishment of biofilm communities in a two-stage continuous culture system. *Appl. Environ. Microbiol.* **71**:7483-7492.
115. Mackie, R.I., Sghir, A. & Gaskins, H.R. 1999. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* **69**:S1035-1045.
116. Mai, V. & Draganov, P.V. 2009. Recent advances and remaining gaps in our knowledge of associations between gut microbiota and human health. *World J. Gastroenterol.* **15**:81-85.
117. Majima, S., Kajino, K., Fukuda, T., Otsuka, F. & Hino, O. 2000. A novel gene "Niban" upregulated in renal carcinogenesis: cloning by the cDNA-amplified fragment length polymorphism approach. *Jpn. J. Cancer Res.* **91**:869-874.
118. Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J. & Doré, J. 2006. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut.* **55**:205-211.
119. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* **437**:376-380.
120. Marteau, P., Pochart, P., Doré, J., Bera-Maillet, C., Bernalier, A. & Corthier, G. 2001. Comparative study of bacterial groups within the human cecal and fecal microbiota. *Appl. Environ. Microbiol.* **67**:4939-4942.
121. Martens, E.C., Chiang, H.C. & Gordon, J.I. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe.* **4**:447-457.
122. Maruyama, K. & Sugano, S. 1994. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene.* **138**:171-174.
123. Mattes, R.D. 2005. Fat taste and lipid metabolism in humans. *Physiol. Behav.* **86**:691-697.
124. Mattes, R.D. 2000. Nutritional implications of the cephalic-phase salivary response. *Appetite.* **34**:177-183.
125. Mattes, R.D. 2006. Orosensory considerations. Obesity (Silver Spring). 14:164S-167S.

126. Mattila-Sandholm, T., Blaut, M., Daly, C., de Vuyst, L., Doré, J., Gibson, G., Goossens, H., Knorr, D., Lucas, J., Lähteenmaki, L., Mercenier, A., Saarela, M., Shanahan, F. & de Vos, W.M. 2002. Food, GI-tract functionality and human health cluster: PROEUHEALTH. *Microb. Ecol. Health Dis.* **14**:65-74.
127. Mays, T.D., Holdeman, L.V., Moore, W.E.C., Rogosa, M. & Johnson, J.L. 1982. Taxonomy of the Genus *Veillonella* Prevot. *Int. J. Syst. Bacteriol.* **32**:28-36.
128. Mengheri, E. 2008. Health, probiotics, and inflammation. *J. Clin. Gastroenterol.* **42**:S177-178.
129. Milinovich, G.J., Burrell, P.C., Pollitt, C.C., Bouvet, A. & Trott, D.J. 2008. *Streptococcus henryi* sp. nov. and *Streptococcus caballi* sp. nov., isolated from the hindgut of horses with oligofructose-induced laminitis. *Int. J. Syst. Evol. Microbiol.* **58**:262-266.
130. Minekus, M., Marteau, P., Havenaar, R. & Huis in 't Veld, J.H.J. 1995. A multicompartmental dynamic computer-controlled model simulating the stomach and small intestine. *ATLA-Alternatives To Laboratory Animals.* **23**:197-209.
131. Minekus, M., Smeets-Peeters, M., Bernalier, A., Marol-Bonnin, S., Havenaar, R., Marteau, P., Alric, M., Fonty, G. & Huis in't Veld, J.H. 1999. A computer-controlled system to simulate conditions of the large intestine with peristaltic mixing, water absorption and absorption of fermentation products. *Appl. Microbiol. Biotechnol.* **53**:108-114.
132. Misra, S.P., Dwivedi, M. & Misra, V. 2006. Ileoscopy in 39 hematochezia patients with normal colonoscopy. *World J. Gastroenterol.* **12**:3101-3104.
133. Moore, W.E. & Holdeman, L.V. 1974. Human fecal flora: the normal flora of 20 Japanese-Hawaiians. *Appl. Microbiol.* **27**:961-979.
134. Morrissey, D.V. & Collins, M.L. 1989. Nucleic acid hybridization assays employing dA-tailed capture probes. Single capture methods. *Mol. Cell. Probes.* **3**:189-207.
135. Morrissey, D.V., Lombardo, M., Eldredge, J.K., Kearney, K.R., Groody, E.P. & Collins, M.L. 1989. Nucleic acid hybridization assays employing dA-tailed capture probes. I. Multiple capture methods. *Anal. Biochem.* **181**:345-359.
136. Morton, T.M., Eaton, D.M., Johnston, J.L. & Archer, G.L. 1993. DNA sequence and units of transcription of the conjugative transfer gene complex (trs) of Staphylococcus aureus plasmid pGO1. *J. Bacteriol.* **175**:4436-4447.
137. Muyzer, G., de Waal, E.C. & Uitterlinden, A.G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**:695-700.
138. Nalin, R., Robe, P. & Tran Van, V. 11 January 2001. Method for indirectly extracting non-cultivable DNA organisms and DNA by said method France patent EP1278831.
139. Neidhardt, F.C., Bloch, P.L. & Smith, D.F. 1974. Culture medium for enterobacteria. *J. Bacteriol.* **119**:736-747.
140. Neut, C., Bulois, P., Desreumaux, P., Membre, J.M., Lederman, E., Gambiez, L., Cortot, A., Quandalle, P., van Kruiningen, H. & Colombel, J.F. 2002. Changes in the bacterial flora of the neoterminal ileum after ileocolonic resection for Crohn's disease. *Am. J. Gastroenterol.* **97**:939-946.
141. Neveu, C., Charvet, C., Fauvin, A., Cortet, J., Castagnone-Sereno, P. & Cabaret, J. 2007. Identification of levamisole resistance markers in the parasitic nematode Haemonchus contortus using a cDNA-AFLP approach. *Parasitology.* **134**:1105-1110.
142. Novis, B.H., Bank, S. & Marks, I.N. 1971. The cephalic phase of pancreatic secretion in man. *Scand. J. Gastroenterol.* **6**:417-422.
143. Nubel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R.I., Ludwig, W. & Backhaus, H. 1996. Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *J. Bacteriol.* **178**:5636-5643.
144. Nugent, S.G., Kumar, D., Rampton, D.S. & Evans, D.F. 2001. Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs. *Gut.* **48**:571-577.
145. O'May, G.A., Reynolds, N. & Macfarlane, G.T. 2005. Effect of pH on an *in vitro* model of gastric microbiota in enteral nutrition patients. *Appl. Environ. Microbiol.* **71**:4777-4783.

146. Paulsen, I.T., Banerjei, L., Myers, G.S., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A. & Fraser, C.M. 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science*. **299**:2071-2074.
147. Pettker, C.M., Buhimschi, I.A., Magloire, L.K., Sfakianaki, A.K., Hamar, B.D. & Buhimschi, C.S. 2007. Value of placental microbial evaluation in diagnosing intra-amniotic infection. *Obstet. Gynecol.* **109**:739-749.
148. Pieterse, B., Jellema, R.H. & van der Werf, M.J. 2006. Quenching of microbial samples for increased reliability of microarray data. *J. Microbiol. Methods.* **64**:207-216.
149. Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A. & Hollibaugh, J.T. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* **71**:4121-4126.
150. Pratscher, J., Stichternoth, C., Fichtl, K., Schleifer, K.H. & Braker, G. 2008. Application of recognition of individual genes-fluorescence *in situ* hybridization (RING-FISH) to detect nitrite reductase genes (nirK) of denitrifier in pure cultures and environmental samples. *Appl. Environ. Microbiol.* **75**:802-810.
151. Projan, S.J., Kornblum, J., Moghazeh, S.L., Edelman, I., Gennaro, M.L. & Novick, R.P. 1985. Comparative sequence and functional analysis of pT181 and pC221, cognate plasmid replicons from *Staphylococcus aureus*. *Mol. Gen. Genet.* **199**:452-464.
152. Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. & Glockner, F.O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**:7188-7196.
153. Quintiliani, R., Jr. & Courvalin, P. 1994. Conjugal transfer of the vancomycin resistance determinant vanB between enterococci involves the movement of large genetic elements from chromosome to chromosome. *FEMS Microbiol. Lett.* **119**:359-363.
154. Raes, J., Korbel, J.O., Lercher, M.J., von Mering, C. & Bork, P. 2007. Prediction of effective genome size in metagenomic samples. *Genome Biol.* **8**:R10.
155. Rajilić-Stojanović, M. 2007. Diversity of the human gastrointesinal microbiota - novel perspectives from high throughput analyses. PhD thesis Wageningen University, Wageningen, the Netherlands.
156. Rakoff-Nahoum, S. & Medzhitov, R. 2006. Role of the innate immune system and host-commensal mutualism. *Curr. Top. Microbiol. Immunol.* **308**:1-18.
157. Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., 2nd, Shah, M., Hettich, R.L. & Banfield, J.F. 2005. Community proteomics of a natural microbial biofilm. *Science.* **308**:1915-1920.
158. Rediers, H., Rainey, P.B., Vanderleyden, J. & de Mot, R. 2005. Unraveling the secret lives of bacteria: use of *in vivo* expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. *Microbiol. Mol. Biol. Rev.* **69**:217-261.
159. Reijans, M., Lascaris, R., Groeneger, A.O., Wittenberg, A., Wesselink, E., van Oeveren, J., de Wit, E., Boorsma, A., Voetdijk, B., van der Spek, H., Grivell, L.A. & Simons, G. 2003. Quantitative comparison of cDNA-AFLP, microarrays, and GeneChip expression data in *Saccharomyces cerevisiae*. *Genomics.* **82**:606-618.
160. Relman, D.A. 2008. 'Til death do us part': coming to terms with symbiotic relationships. Forward. *Nat. Rev. Microbiol.* **6**:721-724.
161. Reysenbach, A.-L. & Pace, N.R. 1995. Protocol 16. Reliable Amplification of Hyperthermophilic Archaeal 16S rRNA Genes by the Polymerase Chain Reaction p. 101-107. *In* F. T. Robb and A. R. Place (ed.), Archaea: A Laboratory Manual - Thermophiles, vol. 3. Cold Spring Harbour Laboratory Press, New York.
162. Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X. & Zhou, J. 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl. Environ. Microbiol.* **70**:4303-4317.

163. Riley, M.L., Schmidt, T., Artamonova, II, Wagner, C., Volz, A., Heumann, K., Mewes, H.W. & Frishman, D. 2007. PEDANT genome database: 10 years online. *Nucleic Acids Res.* **35**:D354-357.
164. Rizos, D., Lonergan, P., Boland, M.P., Arroyo-Garcia, R., Pintado, B., de la Fuente, J. & Gutierrez-Adan, A. 2002. Analysis of differential messenger RNA expression between bovine blastocysts produced in different culture systems: implications for blastocyst quality. *Biol. Reprod.* **66**:589-595.
165. Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J. & Goodman, R.M. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**:2541-2547.
166. Rosenow, C., Saxena, R.M., Durst, M. & Gingeras, T.R. 2001. Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. *Nucleic Acids Res.* **29**:e112.
167. Roy, K., Anba, J., Corthier, G., Rigottier-Gois, L., Monnet, V. & Mistou, M.Y. 2008. Metabolic adaptation of *Lactococcus lactis* in the digestive tract: the example of response to lactose. *J. Mol. Microbiol. Biotechnol.* **14**:137-144.
168. Ruas-Madiedo, P., Gueimonde, M., Fernandez-Garcia, M., de los Reyes-Gavilan, C.G. & Margolles, A. 2008. Mucin degradation by *Bifidobacterium* strains isolated from the human intestinal microbiota. *Appl. Environ. Microbiol.* **74**:1936-1940.
169. Ruseler-van Embden, J.G., van der Helm, R. & van Lieshout, L.M. 1989. Degradation of intestinal glycoproteins by *Bacteroides vulgatus*. *FEMS Microbiol. Lett.* **49**:37-41.
170. Saif, L.J., Ward, L.A., Yuan, L., Rosen, B.I. & To, T.L. 1996. The gnotobiotic piglet as a model for studies of disease pathogenesis and immunity to human rotaviruses. *Arch. Virol. Suppl.* **12**:153-161.
171. Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual, 2 ed. Cold Spring Harbor, New York.
172. Samuel, B.S. & Gordon, J.I. 2006. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc. Natl. Acad. Sci. USA.* **103**:10011-10016.
173. Sanguinetti, C.J., Dias Neto, E. & Simpson, A.J. 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques.* **17**:914-921.
174. Sano, M., Hayakawa, K. & Kato, I. 1992. Purification and characterization of alpha-L-fucosidase from *Streptomyces* species. *J. Biol. Chem.* **267**:1522-1527.
175. Santavirta, J., Mattila, J., Kokki, M. & Matikainen, M. 1991. Mucosal morphology and faecal bacteriology after ileoanal anastomosis. *Int. J. Colorectal Dis.* **6**:38-41.
176. Sarles, H., Dani, R., Prezelin, G., Souville, C. & Figarella, C. 1968. Cephalic phase of pancreatic secretion in man. *Gut.* **9**:214-221.
177. Sarosh, B.R. & Meijer, J. 2007. Transcriptional profiling by cDNA-AFLP reveals novel insights during methyl jasmonate, wounding and insect attack in Brassica napus. *Plant. Mol. Biol.* **64**:425-438.
178. Satokari, R., Gronroos, T., Laitinen, K., Salminen, S. & Isolauri, E. 2009. *Bifidobacterium* and *Lactobacillus* DNA in the human placenta. *Lett. Appl. Microbiol.* **48**:8-12.
179. Savage, D.C. 1977. Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* **31**:107-133.
180. Schaefer, B.C. 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* **227**:255-273.
181. Schloss, P.D. & Handelsman, J. 2003. Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* **14**:303-310.
182. Schmidt, L.M., Williams, P., King, D. & Perera, D. 2004. Picoprep-3 is a superior colonoscopy preparation to Fleet: a randomized, controlled trial comparing the two bowel preparations. *Dis. Colon Rectum.* **47**:238-242.
183. Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeno-Tarraga, A.M., Wang, H., Holden, M.T., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B. & Parkhill, J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**:779-786.

184. Seksik, P., Lepage, P., de la Cochetiere, M.F., Bourreille, A., Sutren, M., Galmiche, J.P., Doré, J. & Marteau, P. 2005. Search for localized dysbiosis in Crohn's disease ulcerations by temporal temperature gradient gel electrophoresis of 16S rRNA. *J. Clin. Microbiol.* **43**:4654-4658.
185. Servin, A.L. & Coconnier, M.H. 2003. Adhesion of probiotic strains to the intestinal mucosa and interaction with pathogens. *Best Pract. Res. Clin. Gastroenterol.* **17**:741-754.
186. Shirazi, T., Longman, R.J., Corfield, A.P. & Probert, C.S. 2000. Mucins and inflammatory bowel disease. *Postgrad. Med. J.* **76**:473-478.
187. Silvester, K.R., Englyst, H.N. & Cummings, J.H. 1995. Ileal recovery of starch from whole diets containing resistant starch measured *in vitro* and fermentation of ileal effluent. *Am. J. Clin. Nutr.* **62**:403-411.
188. Simon, G.L. & Gorbach, S.L. 1984. Intestinal flora in health and disease. *Gastroenterology.* **86**:174-193.
189. Simpson, E.H. 1949. Measurement of Diversity. *Nature.* **163**:688-688.
190. Snel, B., Bork, P. & Huynen, M.A. 2002. The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. USA.* **99**:5890-5895.
191. Snyder, L.A., Loman, N., Pallen, M.J. & Penn, C.W. 2009. Next-Generation Sequencing-the Promise and Perils of Charting the Great Microbial Unknown. *Microb. Ecol.* **57**:1-3.
192. Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.H., Westover, B.P., Weatherford, J., Buhler, J.D. & Gordon, J.I. 2005. Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science.* **307**:1955-1959.
193. Stagg, A.J., Hart, A.L., Knight, S.C. & Kamm, M.A. 2004. Microbial-gut interactions in health and disease. Interactions between dendritic cells and bacteria in the regulation of intestinal immunity. *Best Pract. Res. Clin. Gastroenterol.* **18**:255-270.
194. Stappenbeck, T.S., Hooper, L.V., Manchester, J.K., Wong, M.H. & Gordon, J.I. 2002. Laser capture microdissection of mouse intestine: characterizing mRNA and protein expression, and profiling intermediary metabolism in specified cell populations. *Methods Enzymol.* **356**:167-196.
195. Steele, H.L., Jaeger, K.E., Daniel, R. & Streit, W.R. 2009. Advances in recovery of novel biocatalysts from metagenomes. *J. Mol. Microbiol. Biotechnol.* **16**:25-37.
196. Stewart-Tull, D.E., Ollar, R.A. & Scobie, T.S. 1986. Studies on the *Vibrio cholerae* mucinase complex. I. Enzymic activities associated with the complex. *J. Med. Microbiol.* **22**:325-333.
197. Suvorov, M., Lee, M., Hesek, D., Boggess, B. & Mobashery, S. 2008. Lytic transglycosylase MltB of Escherichia coli and its role in recycling of peptidoglycan strands of bacterial cell wall. *J. Am. Chem. Soc.* **130**:11878-11879.
198. Suzuki, M.T., Taylor, L.T. & DeLong, E.F. 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. *Appl. Environ. Microbiol.* **66**:4605-4614.
199. Taroncher-Oldenburg, G., Griner, E.M., Francis, C.A. & Ward, B.B. 2003. Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl. Environ. Microbiol.* **69**:1159-1171.
200. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J. & Natale, D.A. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* **4**:41.
201. Thomson, A.B., Drozdowski, L., Iordache, C., Thomson, B.K., Vermeire, S., Clandinin, M.T. & Wild, G. 2003. Small bowel review: Normal physiology, part 1. *Dig. Dis. Sci.* **48**:1546-1564.
202. Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P. & Rubin, E.M. 2005. Comparative metagenomics of microbial communities. *Science.* **308**:554-557.
203. Tsai, H.H., Sunderland, D., Gibson, G.R., Hart, C.A. & Rhodes, J.M. 1992. A novel mucin sulphatase from human faeces: its identification, purification and characterization. *Clin. Sci. (Lond.).* **82**:447-454.
204. Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R. & Gordon, J.I. 2009. A core gut microbiome in obese and lean twins. *Nature.* **457**:480-484.

205. Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* **23**:88-93.
206. Van den Broek, L.A.M. 2005. Glycosyl hydrolases from *Bifidobacterium adolescentis* DSM20088 - Their role in the metabolism and synthesis of oligosaccharides. PhD thesis Wageningen University, Wageningen, the Netherlands.
207. Van der Waaij, L.A., Harmsen, H.J., Madjipour, M., Kroese, F.G., Zwiers, M., van Dullemen, H.M., de Boer, N.K., Welling, G.W. & Jansen, P.L. 2005. Bacterial population analysis of human colon and terminal ileum biopsies with 16S rRNA-based fluorescent probes: commensal bacteria live in suspension and have no direct contact with epithelial cells. *Inflamm. Bowel Dis.* **11**:865-871.
208. Van Laere, K.M., Hartemink, R., Beldman, G., Pitson, S., Dijkema, C., Schols, H.A. & Voragen, A.G. 1999. Transglycosidase activity of *Bifidobacterium adolescentis* DSM 20083 alpha-galactosidase. *Appl. Microbiol. Biotechnol.* **52**:681-688.
209. Vanhoutte, T., Huys, F., De Brandt, E. & Swings, J. 2004. Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. *FEMS Microbiol. Ecol.* **48**:437-446.
210. Vaughan, E.E., Heilig, H.G., Ben-Amor, K. & de Vos, W.M. 2005. Diversity, vitality and activities of intestinal lactic acid bacteria and bifidobacteria assessed by molecular approaches. *FEMS Microbiol. Rev.* **29**:477-490.
211. Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M.G., Apajalahti, J., Tysk, C., Hettich, R.L. & Jansson, J.K. 2009. Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* **3**:179-189.
212. Vieites, J.M., Guazzaroni, M.E., Beloqui, A., Golyshin, P.N. & Ferrer, M. 2009. Metagenomics approaches in systems microbiology. *FEMS Microbiol. Rev.* **33**:236-255.
213. Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.
214. Wagner, M., Smidt, H., Loy, A. & Zhou, J. 2007. Unravelling microbial communities with DNA-microarrays: challenges and future directions. *Microb. Ecol.* **53**:498-506.
215. Walter, J., Mangold, M. & Tannock, G.W. 2005. Construction, analysis, and beta-glucanase screening of a bacterial artificial chromosome library from the large-bowel microbiota of mice. *Appl. Environ. Microbiol.* **71**:2347-2354.
216. Wang, M., Ahrne, S., Jeppsson, B. & Molin, G. 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.
217. Wang, X., Heazlewood, S.P., Krause, D.O. & Florin, T.H. 2003. Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J. Appl. Microbiol.* **95**:508-520.
218. Weber, A. & Jung, K. 2002. Profiling early osmostress-dependent gene expression in Escherichia coli using DNA macroarrays. *J. Bacteriol.* **184**:5502-5507.
219. Whitcomb, D.C. & Lowe, M.E. 2007. Human pancreatic digestive enzymes. *Dig. Dis. Sci.* **52**:1-17.
220. Woods, S.C., Lutz, T.A., Geary, N. & Langhans, W. 2006. Pancreatic signals controlling food intake; insulin, glucagon and amylin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**:1219-1235.
221. Wright, D.P., Rosendale, D.I. & Robertson, A.M. 2000. *Prevotella* enzymes involved in mucin oligosaccharide degradation and evidence for a small operon of genes expressed during growth on mucin. *FEMS Microbiol. Lett.* **190**:73-79.
222. Wu, L., Thompson, D.K., Li, G., Hurt, R.A., Tiedje, J.M. & Zhou, J. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl. Environ. Microbiol.* **67**:5780-5790.
223. Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V. & Gordon, J.I. 2003. A genomic view of the Human-*Bacteroides thetaiotaomicron* symbiosis. *Science.* **299**:2074-2076.
224. Yun, J. & Ryu, S. 2005. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microb. Cell Fact.* **4**:8.
225. Zhou, J.S., Gopal, P.K. & Gill, H.S. 2001. Potential probiotic lactic acid bacteria *Lactobacillus rhamnosus* (HN001), *Lactobacillus acidophilus* (HN017) and *Bifidobacterium lactis* (HN019) do not degrade gastric mucin *in vitro*. *Int. J. Food Microbiol.* **63**:81-90.

226. Zhou, M., Boekhorst, J., Francke, C. & Siezen, R.J. 2008. LocateP: genome-scale subcellular-location predictor for bacterial proteins. *BMC Bioinformatics*. **9**:173.
227. Zimmer, M. & Brune, A. 2005. Physiological properties of the gut lumen of terrestrial isopods (Isopoda: Oniscidea): adaptive to digesting lignocellulose? *J. Comp. Physiol. [B]*. **175**:275-283.
228. Zocco, M.A., Ainora, M.E., Gasbarrini, G. & Gasbarrini, A. 2007. *Bacteroides thetaiotaomicron* in the gut: molecular aspects of their interaction. *Dig. Liver Dis*. **39**:707-712.
229. Zoetendal, E.G., Akkermans, A.D.L., Akkermans-van Vliet, W.M., de Visser, A.G.M. & de Vos, W.M. 2001. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microb. Ecol. Health Dis*. **13**:129-134.
230. Zoetendal, E.G., Akkermans, A.D.L. & de Vos, W.M. 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.
231. Zoetendal, E.G., Booijink, C.C.G.M., Klaassens, E.S., Heilig, H.G.H.J., Kleerebezem, M., Smidt, H. & de Vos, W.M. 2006. Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat. Protocols*. **1**:954-959.
232. Zoetendal, E.G., Collier, C.T., Koike, S., Mackie, R.I. & Gaskins, H.R. 2004. Molecular ecological analysis of the gastrointestinal microbiota: a review. *J Nutr*. **134**:465-472.
233. Zoetendal, E.G., Heilig, H.G.H.J., Klaassens, E.S., Booijink, C.C.G.M., Kleerebezem, M., Smidt, H. & de Vos, W.M. 2006. Isolation of DNA from bacterial samples of the human gastrointestinal tract. *Nat. Protocols*. **1**:870-873.
234. Zoetendal, E.G., Rajilić-Stojanović, M. & de Vos, W.M. 2008. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut*. **57**:1605-1615.
235. Zoetendal, E.G., Vaughan, E.E. & de Vos, W.M. 2006. A microbial world within us. *Mol. Microbiol*. **59**:1639-1650.
236. Zoetendal, E.G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A.D.L. & de Vos, W.M. 2002. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl. Environ. Microbiol*. **68**:3401-3407.

CHAPTER 3

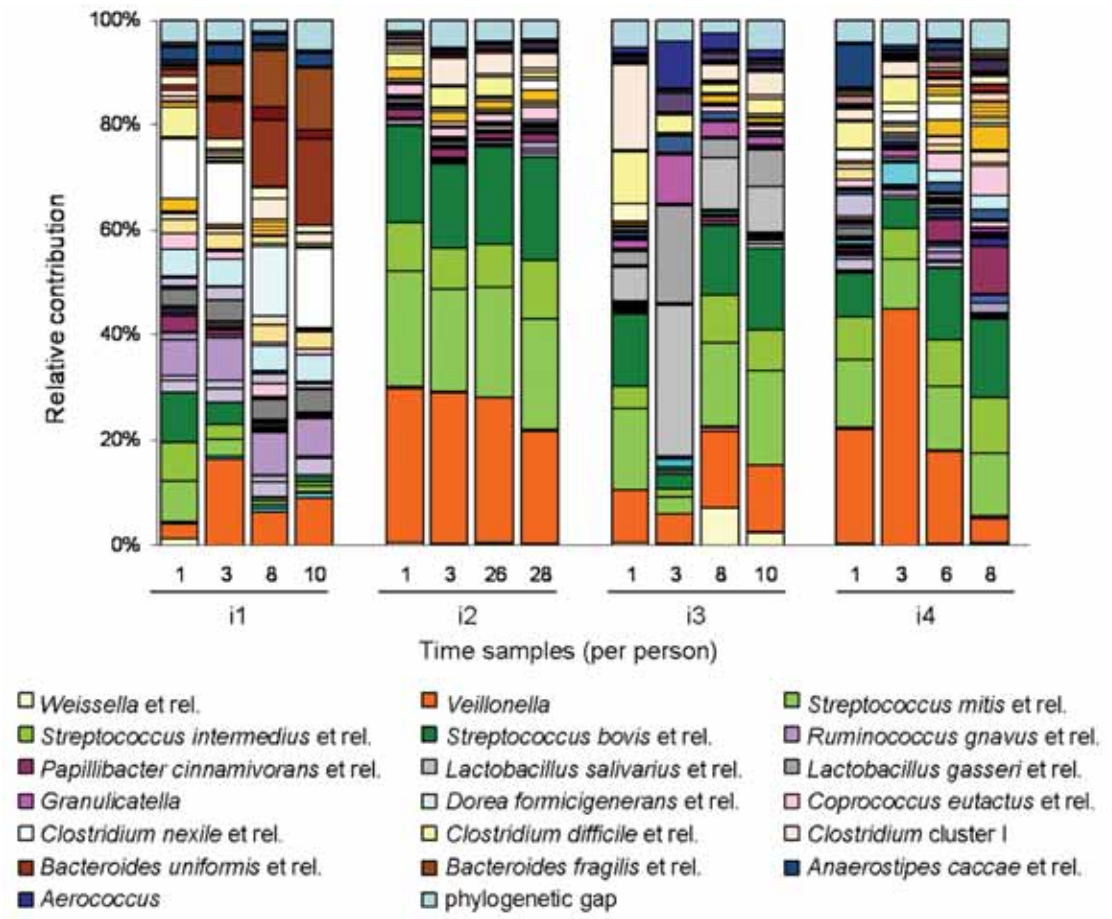
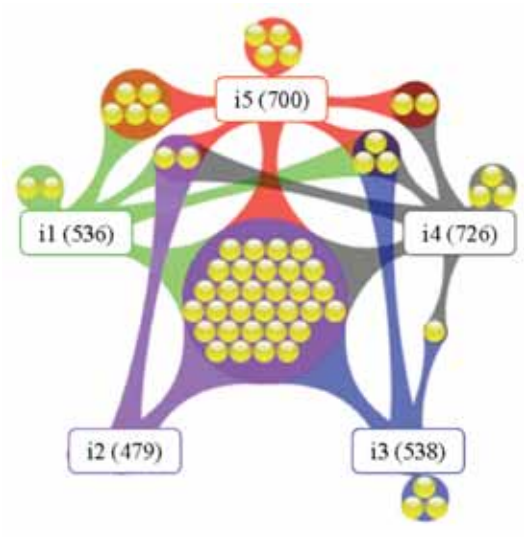


FIGURE 3-3 (**c**) Relative contribution of the detected phylotypes with HITChip in four ileostomy effluent samples collected over time of four healthy ileostomy subjects. The day of sampling per person is indicated under the bars. In the legend, phylotypes that contribute at least 5% to one of the profiles are indicated. The added spike is indicated as phylogenetic gap on top of the columns.



"Common core" microbiota detected in ileal effluent of five subjects with an ileostoma
Clostridium cluster I
*Clostridium dif f icil*et rel.
Enterococcus
*Oxalobacter f ormigenes*et rel.
Streptococcus bovis et rel.
Streptococcus intermedius et rel.
Streptococcus mitis et rel.
Veillonella

FIGURE 3-S1 Venn diagram showing the distribution of the taxonomic groups that had significant hybridisation signal in ileal effluent samples from five healthy subjects with an ileostomy over time. Each yellow sphere represents ten responding probes. The subjects are indicated with i1 to i5, whereas numbers in parentheses indicate the total number of responding probes within that sample. The eight detected taxonomic groups common for all five subjects are listed in the table. The image was created using AutoFocus version 5 software (Aduna B.V., the Netherlands).

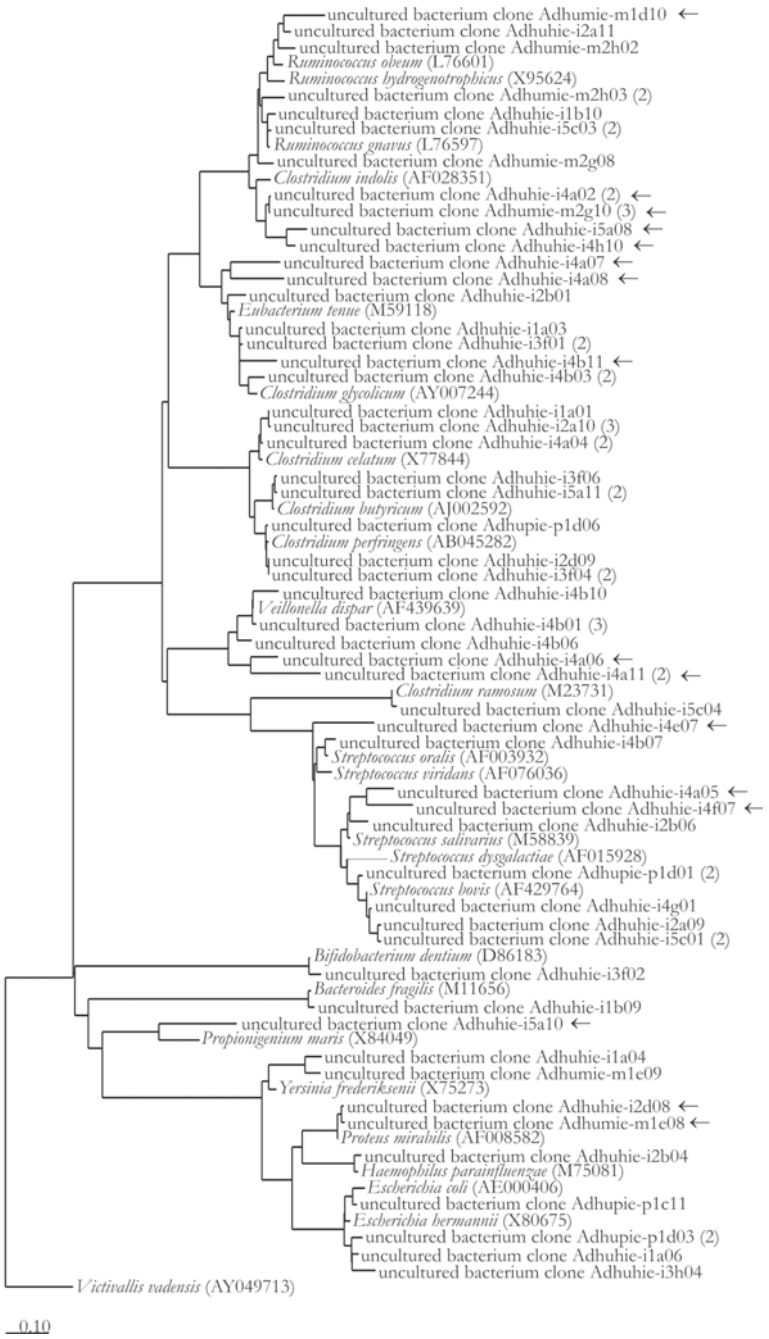


FIGURE 3-S2 Neighbour-joining tree of the phylotypes identified in a clone library constructed from universal 16S rRNA amplicons (± 430 bp) and closest related cultured representatives in GenBank (accession numbers in parentheses) identified in ileal effluent samples collected in time. A total number of 72 sequences were obtained from this ileal effluent clone library. The eight individuals are identified as five healthy ileostomy subjects (i1 to i5), two medicated ileostomy subjects (m1 and m2) and one person with an ileorectal pouch (p1). Numbers between parentheses behind the clone reference indicate the frequency of the sequence within the clone library. Arrows indicate sequences with <97% similarity to its closest relative. Sequences were aligned with Pearson correlation within ARB, and the distance matrix was calculated with the neighbour joining algorithm. The phylogenetic tree was rooted with the appropriate amplicon region of the 16S rRNA gene sequence of *Victivallis vadensis*. The scale bar represents the genetic distance.

a.

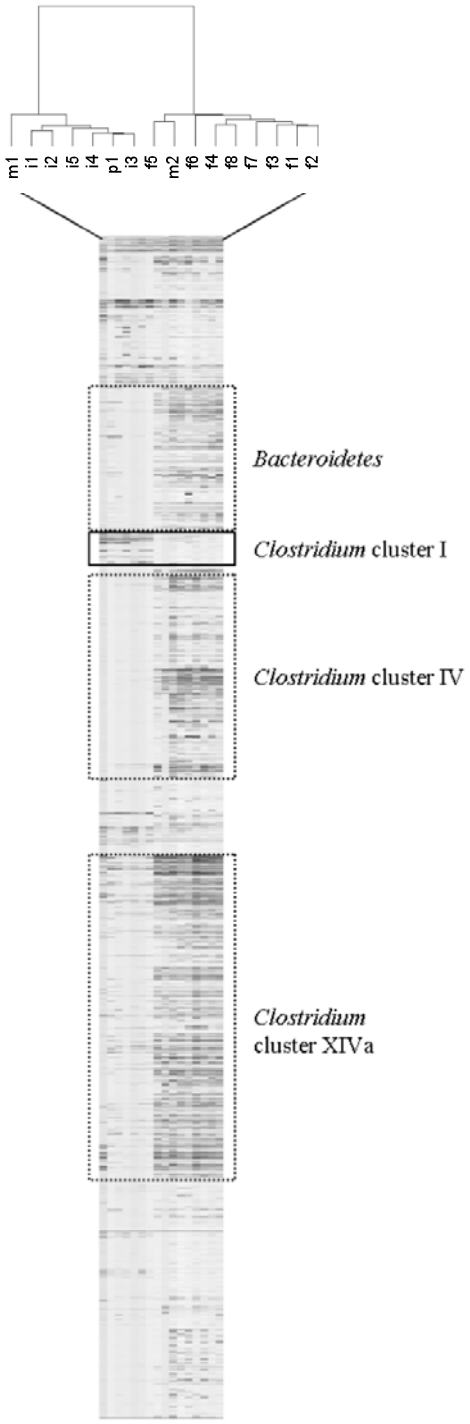


FIGURE 3-S3 (**a**) Taxonomic profile of the eight ileal effluent profiles compared with eight randomly chosen faecal profiles, analysed by HITChip. Taxonomic groups more dominant in ileal profiles compared to faecal are indicated with black rectangles, whereas groups less dominant in ileal profiles are indicated with dashed rectangles. For each individual, the last ileal effluent sample collected in the morning was used for HITChip analysis. Abbreviations used are: i1 to i5 for individuals with an ileostomy; m1 and m2 for ileostomy subjects under medication; p1 for an individual with an ileorectal pouch and f1 to f8 for faecal samples of eight healthy individuals.

b.

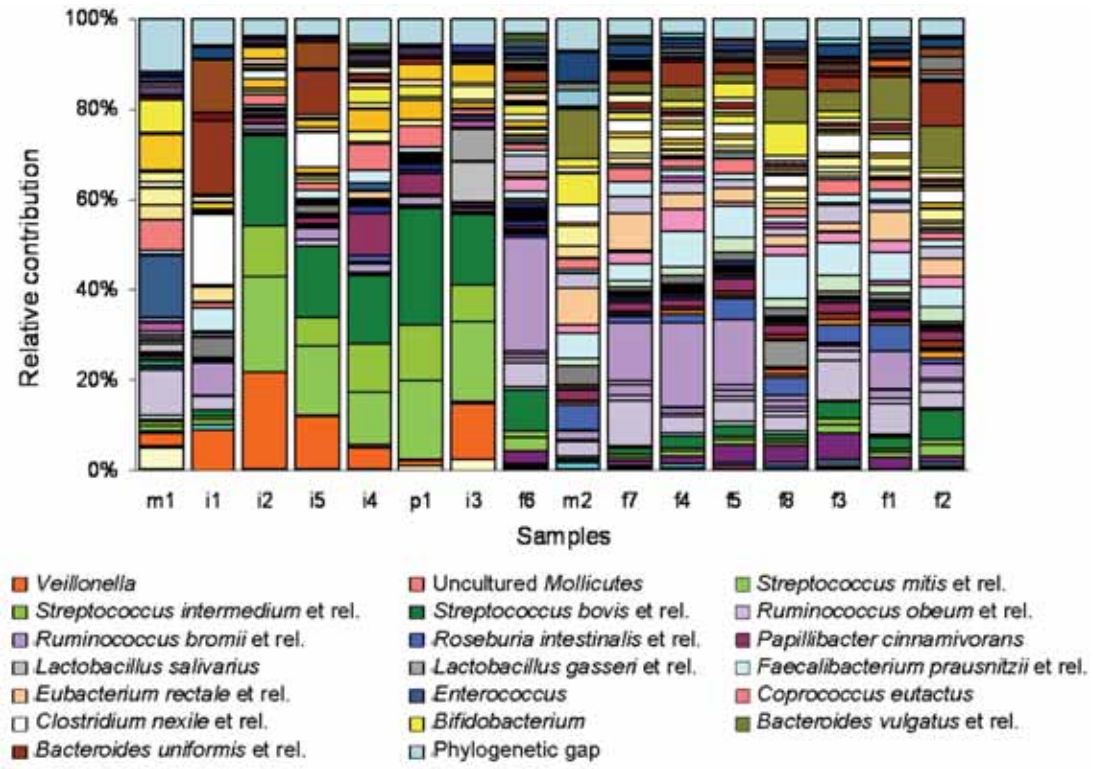


FIGURE 3-S3 (**b**) The relative contribution of each taxonomic unit to the total microbiota is depicted for each subject. In the legend all the phylotypes contributing >5% to two or more of the profiles are indicated, where the added spike is indicated as phylogenetic gap. The order of the samples in the bar plots is equal to the order on the taxonomic profile in (**a**). Abbreviations used are: i1 till i5 for the healthy ileostomy subjects, p1 for the person with an ileorectal pouch; m1 and m2 for ileostomy subjects taking medication and f1 till f8 for faecal profiles of eight random chosen healthy adults.

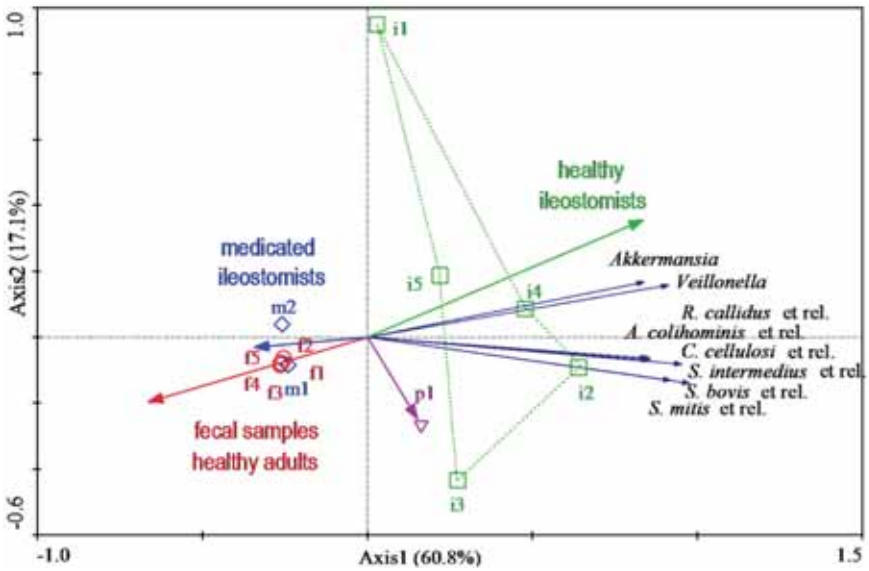
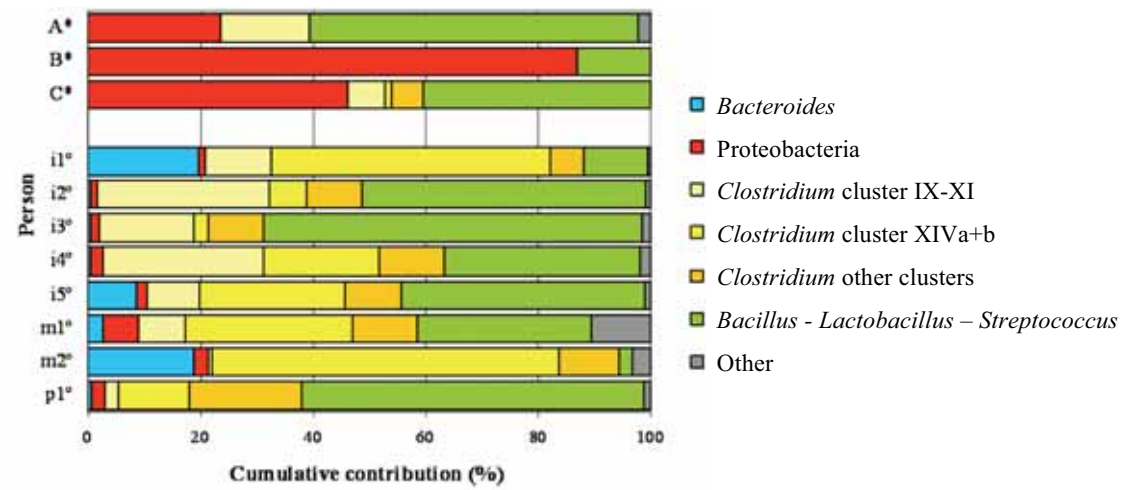


FIGURE 3-S4 PCA triplot depicting the microbiota of five healthy ileostomy subjects (i1 till 15, indicated by green squares), two medicated ileostomy subjects (m1 and m2, indicated by blue diamonds), one person with an ileorectal pouch (p1, indicated by a purple triangle), five faecal samples of healthy adults (f1 till f5, indicated by red circles). Percentage values at the axes indicate contribution of the principle components to the explanation of total variance in the dataset. The phylogenetic groups, named after its cultivated representative that contributed at least 70% to the explanatory axis used in the plot are presented as vectors.



* Based on random clone libraries of DNA (272 clones with an average insert size of 500 bp) from ileum samples of three subjects taken at autopsy (Hayashi *et al.* [72].)
Subject A: a 74-year-old male; subject B: a85-year-old female; subject C: a 87-year-old female.

° Based on HITChip profiles of ileal effluent samples of seven ileostomy subjects and one subject with an ileorectal pouch. Abbreviations used to indicate the subjects are: i1 = a 55-year-old female with an ileostomy; i2 = a 60-year-ol female with an ileostomy; i3 = a 57-year-old male with an ileostomy; i4 = a 74-year-old male with an ileostomy; i5 = a 48-year-old male with an ileostomy; m1 = a 41-year-old male with an ileostomy on medication (prednisone); m2 = a 52-year-old female with an ileostomy on medication (mesalamine); p1 = a 48-year-old female with an ileorectal pouch.

FIGURE 3-S5 The ileal microbiota composition in three individuals at autopsy as determined by sequence analysis of random clone libraries compared to the ileal effluent microbiota of seven ileostomy subjects and one subject with an ileorectal pouch as determined by HITChip analysis.

CHAPTER 5

TABLE 5-S1 An overview of the restriction enzymes and selective nucleotides used for profiling the nucleotides (DNA, RNA and enriched mRNA) extracted from the faecal microbiota of two healthy subjects. Primer combinations that were used for sequence analysis of the profiled RNA molecules, are indicated with an asterisk (*).

Primer codes		Primer1	Primer2
TR14/B01		TaqI+AT	BfaI+A
TR14/B03	*	TaqI+AT	BfaI+G
TR14/B04	*	TaqI+AT	BfaI+T
TR01/M20	*	TaqI+A	MseI+GC
TR01/M23	*	TaqI+A	MseI+TA
TR01/M24	*	TaqI+A	MseI+TC
TR03/M17		TaqI+G	MseI+CG
TR12/M24		TaqI+AC	MseI+TC
TR18/M17		TaqI+CT	MseI+CG
TR14/M20		TaqI+AT	MseI+GC
TR14/M23		TaqI+AT	MseI+TA
TR14/M24		TaqI+AT	MseI+TC
TR14/M82	*	TaqI+AT	MseI+TAT
TR14/M85	*	TaqI+AT	MseI+TCG
TR20/M55		TaqI+GC	MseI+CGA

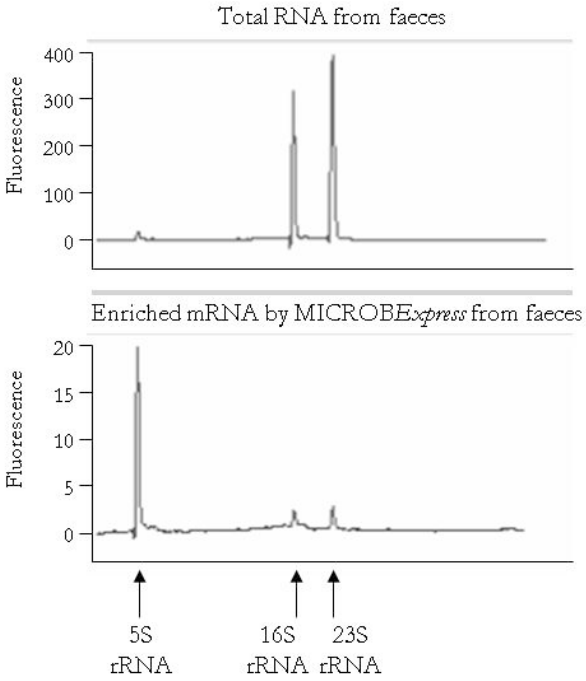


FIGURE 5-S1 MICROB*Express* treatment of total RNA from faeces analysed with the BioAnalyzer. RNA plots of the total faecal RNA and enriched mRNA are shown. The peaks corresponding to the 5S, 16S and 23S rRNA molecules are indicated by arrows.

TABLE 5-S2 Overview of the TDFs obtained from the cDNA-AFLP gels of subjects 1 and 2 that gave highest hits to gene sequences or encoded proteins available in GenBank. Per TDF, information is listed in columns, including sequence length, highest hit with corresponding Blast results and functional category based on COG. Abbreviations used in column Db (database) are: N = nr NCBI_nucleotides (BLASTN), U = Uniprot and E = Expressed Sequence Tags (TBLASTX).

	TDFs	Length (bp)	Db	Highest hit sequence	Phylum [order]	ID (%)	e-value	GenBank reference	COG
Peron 1	1D11a	215	U	Hypothetical prot.	Firmicutes [Clostridiales]	61.5	9.0E-01	ABZ84886	E
	1G09a	185	U	Glyceraldehyde-3-P dehydrogenase	Firmicutes [Clostridiales]	56.0	2.0E-04	CAC80449	G
	1G10a	191	U	Pyruvate flavodoxin/ferredoxin oxidoreductase	Firmicutes [Clostridiales]	81.5	4.0E-19	EAV696767	C
	1H06a	303	U	Hypothetical prot.	Firmicutes [Clostridiales]	98.0	3.0E-53	EDR46217	X
	1H07a	155	U	Glyceraldehyde-3-P dehydrogenase	Firmicutes [Clostridiales]	91.7	1.0E-08	CAC80449	G
	1H08a	190	U	Pyruvate-ferrodoxin oxidoreductase	Firmicutes [Clostridiales]	79.6	1.0E-14	EAV696767	C
	1A02b	140	U	Hypothetical prot.	Bacteroidetes [Bacteroidales]	74.5	4.0E-13	EEC98392	X
	1A09b	69	U	Glucose/galactose transporter	Bacteroidetes [Bacteroidales]	97.2	4.0E-24	CP000139	X
	1C11b	356	U	ATP-dependent DNA helicase RecG	Thermotogae [Thermotogales]	30.9	5.0E-07	EEB81428	LK
	1E12b	435	U	Hypothetical prot.	Firmicutes [Clostridiales]	47.1	8.0E-34	EDP22300	S
	1F05b	250	U	Glucose-6-phosphate isomerase	Bacteroidetes [Bacteroidales]	95.3	2.0E-26	ABR41711	J
Person 2	2A02	293	N	Massetolide A biosynthesis gene cluster	Proteobacteria [Pseudomonadales]	86.6	1.0E-30	EU199081	C
	2A04	220	U	Alcohol dehydrogenase zinc-containing	Proteobacteria [Alteromonadales]	67.3	2.0E-09	EAZ99052	C R
	2A06	203	N	5': Put. activator/repressor of mer operon/ 3': hyp. prot.	Proteobacteria [Pseudomonadales]	95.9	3.0E-58	AM235768	X
	2A09	173	U	Lipoyl synthase	Proteobacteria [Pseudomonadales]	58.7	4.0E-07	CAK17475	H
	2B04	121	N	Alanine dehydrogenase	Bacteroidetes [Bacteroidales]	96.5	2.0E-43	CP000139	E
	2C03	401	U	SSS sodium solute transporter superfamily	Proteobacteria [Desulfuromonadales]	53.8	2.0E-09	ACH38535	R
	2C05	268	N	Two-component response regulator, PprB	Proteobacteria [Pseudomonadales]	89.5	4.0E-42	AE004091	T K
	2C07	203	N	5': transposition helper prot. / 3': hypothetical prot.	Proteobacteria [Pseudomonadales]	93.0	6.0E-74	AE016853	C
	2C09	199	N	Cons. hyp. prot. with FAD/NAD(H) binding domain	Proteobacteria [Enterobacteriales]	83.2	3.0E-23	CP000948	S
	2D11	313	U	Lactose operon repressor	Proteobacteria [Enterobacteriales]	87.0	3.0E-39	ABR80251	K
	2D12	277	N	Transcriptional regulator, NifA subfamily, Fis Family	Proteobacteria [Enterobacteriales]	100.0	1.0E-136	CP000946	K T
	2E01	199	N	tRNA 2-selenouridine synthase, selenophosphatase-dep.	Proteobacteria [Enterobacteriales]	98.5	2.0E-94	CP000948	R
	2E02	201	U	Hypothetical prot.	Firmicutes [Clostridiales]	97.5	1.0E-14	EDM62187	G
	2E04	134	N	SSU ribosomal prot. S18	Firmicutes [Lactobacillales]	99.2	1.0E-56	CP000419	J
	2E06	111	N	Hypothetical prot.	Proteobacteria [Enterobacteriales]	94.6	2.0E-39	CP000783	E T
	2E08	124	U	GAPDH, type I	Firmicutes [Clostridiales]	76.9	2.0E-07	ACD23900	G
	2E09	184	N	Putative pyruvate ferredoxin oxidoreductase	Firmicutes [Clostridiales]	74.7	3.0E-20	CPA17727	C
	2F01	127	N	SSU ribosomal prot. S18	Firmicutes [Lactobacillales]	99.2	4.0E-57	CP000419	J
	2F02	398	N	Phosphonate ABC transporter, inner membrane subunit	Proteobacteria [Pseudomonadales]	81.6	2.0E-59	CP000680	P
	2F05	263	N	Probable acyl-CoA thiolase	Proteobacteria [Pseudomonadales]	81.0	2.0E-45	CP000744	I
	2F08	182	U	N-acetymuramoyl-1-alanine amidase	Proteobacteria [Enterobacteriales]	100.0	3.0E-11	CAQ88278	M
	2F09	214	U	4-Aminobutyrate aminotransferase, putative	Proteobacteria [Pseudomonadales]	89.2	1.0E-10	AAN69692	E
	2F10	164	N	5': Cons. domain prot. / 3': electron transfer flavoprot.	Proteobacteria [Pseudomonadales]	84.8	2.0E-22	CP000076	T K
	2G01	144	U	Aspartate kinase	Firmicutes [Clostridiales]	86.4	1.0E-05	EAX47162	E
	2G02	119	N	Prot. of unknown function	Proteobacteria [Enterobacteriales]	97.4	5.0E-46	CP000946	S
	2G09	402	U	Laccase	Proteobacteria [Pseudomonadales]	97.0	1.0E-30	ABL75157	V
	2G12	251	E	Eukaryotic mRNA	Streptophyta [Sapindales]	93.8	1.0E-81	EY714820	X
	2H01	159	U	Type II secretion system prot.	Proteobacteria [Pseudomonadales]	58.3	3.0E-03	ABA76558	NU
	2H02	162	U	SSS sodium solute transporter superfamily	Proteobacteria [Burkholderiales]	77.3	3.2E-01	ACD19513	R

CELLULAR PROCESSES AND SIGNALLING
[D] Cell cycle control, cell division, chromosome partitioning
[Y] Nuclear structure
[V] Defense mechanisms
[T] Signal transduction mechanisms
[M] Cell wall/membrane/envelope biogenesis
[N] Cell motility
[Z] Cytoskeleton
[W] Extracellular structures
[U] Intracellular trafficking, secretion, and vesicular transport
[O] Posttranslational modification, protein turnover, chaperones

INFORMATION STORAGE AND PROCESSING
[J] Translation, ribosomal structure and biogenesis
[A] RNA processing and modification
[K] Transcription
[L] Replication, recombination and repair
[B] Chromatin structure and dynamics

METABOLISM
[C] Energy production and conversion
[G] Carbohydrate transport and metabolism
[E] Amino acid transport and metabolism
[F] Nucleotide transport and metabolism
[H] Coenzyme transport and metabolism
[I] Lipid transport and metabolism
[P] Inorganic ion transport and metabolism
[Q] Secondary metabolites biosynthesis, transport & catabolism

POORLY CHARACTERISED
[R] General function prediction only
[S] Function unknown

NO FUNCTIONAL CLUSTER ASSIGNED
[X] No assignment of functional group possible

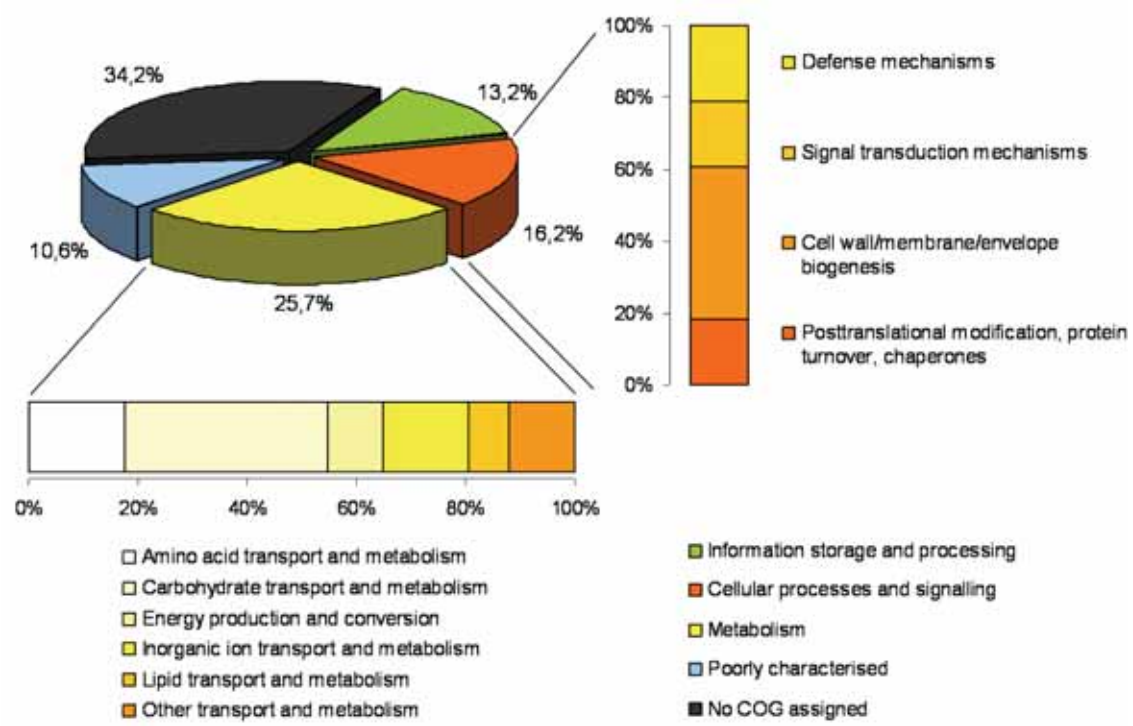


FIGURE 6-3 Classification of the sequences retrieved from the mRNA enriched cDNA-AFLP profiles of ileostomy samples from an ileostomy subject into Clusters of Orthologous Genes (COG) as defined by Tatusov *et al.* [200]. Different colours represent the following clusters: green = information storage and processing, orange = cellular processes and signalling, yellow = metabolism, blue = poorly characterised, and black = no functional COG assigned.

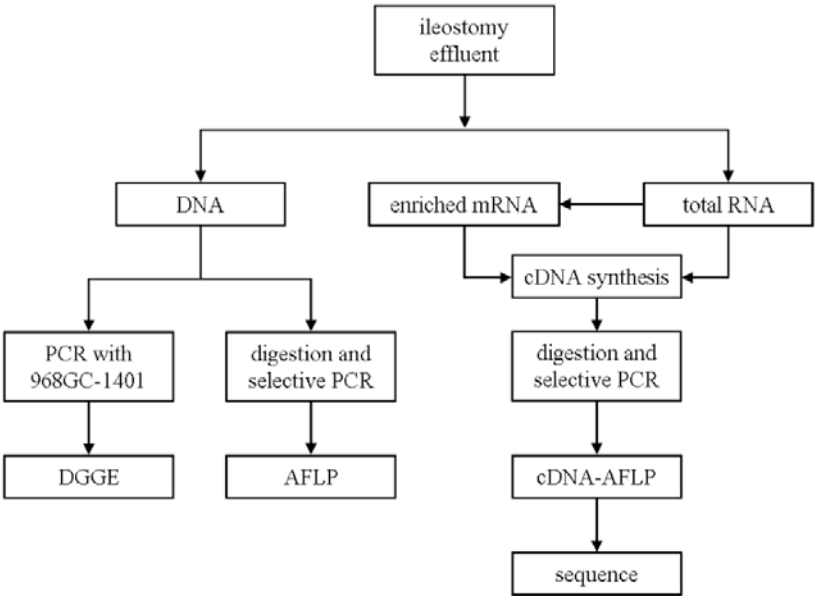


FIGURE 6-S1 Overview of the molecular approaches used to analyse the ileostomy effluent microbiota and its transcriptome.

TABLE 6-S1 Overview of the restriction enzymes and selective nucleic acids used for profiling the nucleotides (DNA, RNA and enriched mRNA) extracted from the ileostomy effluent microbiota. Primer combinations that were used for sequence analysis of the profiled RNA molecules are indicated with an asterisk (*).

Primer codes		Primer1	Primer2
Tr14 / B01	*	TaqI+AT	BfaI+A
Tr14 / B03	*	TaqI+AT	BfaI+G
Tr14 / B04	*	TaqI+AT	BfaI+T
Tr14 / B15	*	TaqI+AT	BfaI+CA
Tr14 / B16	*	TaqI+AT	BfaI+CC
Tr14 / B17	*	TaqI+AT	BfaI+CG
Tr14 / B18	*	TaqI+AT	BfaI+CT
Tr01 / M20	*	TaqI+A	MseI+GC
Tr01 / M23		TaqI+A	MseI+TA
Tr01 / M24	*	TaqI+A	MseI+TC
Tr02 / M17	*	TaqI+C	MseI+CG
Tr03 / M17		TaqI+G	MseI+CG
Tr11 / M20		TaqI+AA	MseI+GC
Tr12 / M20		TaqI+AC	MseI+GC
Tr12 / M23		TaqI+AC	MseI+TA
Tr12 / M24		TaqI+AC	MseI+TC
Tr14 / M20		TaqI+AT	MseI+GC
Tr14 / M23		TaqI+AT	MseI+TA
Tr14 / M24	*	TaqI+AT	MseI+TC
Tr18 / M17	*	TaqI+CT	MseI+CG
Tr20 / M17		TaqI+GC	MseI+CG
Tr14 / M82		TaqI+AT	MseI+TAT
Tr14 / M85	*	TaqI+AT	MseI+TCG
Tr14 / M86		TaqI+AT	MseI+TCT
Tr18 / M55	*	TaqI+CT	MseI+CGA
Tr20 / M57		TaqI+GC	MseI+CGG

TABLE 6-S2 Overview of the transcript sequences obtained from the cDNA-AFLP gels of ileostomy effluent samples from an ileostomy subject that gave highest hits to gene sequences or encoded proteins available in GenBank. Per transcript, information is listed in columns, including sequence length, highest hit with corresponding Blast results and functional category based on COG (clusters A till Z). Sequences marked with I till V refer to transcripts retrieved at both time points (as shown in Figure 6-3).

	Sequence	Time point	Length (bp)	Highest hit sequence	Phylum [order]	ID (%)	e-value	GenBank reference	COG	
	4C07	t_0	81	Ribosomal protein S18	Firmicutes [Lactobacillales]	100	6.0E-06	ACA82004	J	
I→	4D06	t_0	63	Glutamate synthase (large subunit)	Firmicutes [Lactobacillales]	80	6.5E-02	EA014133	X	
	4E01	t_0	266	Putative bèta-galactosidase	Firmicutes [Lactobacillales]	61	2.0E-15	ABN43519	I	
	4E05	t_0	201	Acyl carrier Prot. S-malonyltransferase	Proteobacteria [Enterobacteriales]	90	2.0E-25	ZP_02796943	M	R
	4E11	t_0	121	Ribosomal prot. S1	Proteobacteria [Enterobacteriales]	100	8.0E-11	ZP_02794913	P	
	4F11	t_0	237	Put. beta-galactosidase	Firmicutes [Lactobacillales]	61	2.0E-15	ABN43519	X	
	4G01	t_0	186	Hypothetical protein	Firmicutes [Lactobacillales]	78	2.0E-24	AAV59811	X	
	4G09	t_0	283	Sortase B family protein	Firmicutes [Lactobacillales]	78	5.0E-33	ABV11136	X	
II→	4H07	t_0	154	Putative receptor	Proteobacteria [Pseudomonadales]	76	5.0E-16	AAY91513	G	
	4H11	t_0	115	Putative transcriptional regulator	Bacteroidetes [Bacteroidales]	100	3.0E-15	YP_097651	X	
	7E02	t_0	208	(Acyl-carrier-prot.) S-malonyltransferase	Firmicutes [Lactobacillales]	84	1.0E-12	ABJ65720	X	
III→	8B05a	t_0	208	Cons. hypothetical protein	Firmicutes [Lactobacillales]	81	4.0E-24	EDT48500	S	
	8C12a	t_0	450	50S ribosomal protein	Firmicutes [Lactobacillales]	96	5.0E-24	NP_268478	X	
	8D01a	t_0	382	Glutamine ABC uptake transporter	Firmicutes [Lactobacillales]	80	2.0E-47	AAV61104	E	T
	8D07a	t_0	220	Possible cell surface protein	Firmicutes [Lactobacillales]	56	4.0E-10	ABJ61013	X	
	8E05a	t_0	104	Acetoin dehydrogenase complex, alpha subunit	Firmicutes [Lactobacillales]	91	3.0E-11	AAV60703	C	
IV→	8F06a	t_0	253	Ribosomal subunit interface protein	Firmicutes [Lactobacillales]	86	1.0E-32	EAO76428	J	
	8A04b	t_0	211	Aldose 1-epimerase	Firmicutes [Lactobacillales]	70	3.0E-20	ABA47362	G	
	8A06b	t_0	258	Efflux ABC transporter, permease protein	Firmicutes [Lactobacillales]	56	9.0E-12	AcA37582	V	
	8A09b	t_0	128	Magnesium-translocating P-type ATPase	Firmicutes [Lactobacillales]	80	5.0E-50	ABV09442	P	
V→	8A11b	t_0	168	Sugar hydrolase	Firmicutes [Lactobacillales]	70	6.0E-17	ABX75773	G	
	8B04b	t_0	169	ABC transporter permease protein	Firmicutes [Clostridiales]	81	2.0E-16	YP_001692649	V	
	8G10b	t_0	245	Ribulose-5-P 3-epimerase	Firmicutes [Lactobacillales]	93	3.0E-35	AAV61396	G	
	4A05	t_1	179	Sensory box histidine kinase	Firmicutes [Lactobacillales]	73	1.0E-09	AAM99607	T	
	4A07	t_1	159	Succinylglutamate desuccinylase	Proteobacteria [Enterobacteriales]	82	2.0E-17	ABF03663	E	
	4A09	t_1	129	PTS system, IIC component	Firmicutes [Lactobacillales]	94	2.0E-04	AAO81603	X	
	4A12	t_1	119	Integr. membr. prot. possibly inv. in D-ala export	Firmicutes [Lactobacillales]	87	3.0E-14	AAC05775	M	
	4C01	t_1	209	Hypothetical protein	Firmicutes [Lactobacillales]	63	1.0E-16	CAD46652	X	
	4C02	t_1	205	Homoserine O-succinyltransferase	Firmicutes [Lactobacillales]	69	7.0E-23	AAN59124	E	
	4C04	t_1	150	Sensory box histidine kinase	Firmicutes [Lactobacillales]	77	6.0E-06	AAM99607	T	
	4C06	t_1	120	Hypothetical protein	Firmicutes [Lactobacillales]	100	6.0E-04	EDK73304	X	
	4C12	t_1	79	Penicillin-binding protein 2A	Firmicutes [Lactobacillales]	84	3.0E-04	ABP93147	M	
I→	4D05	t_1	63	Glutamate synthase (large subunit)	Firmicutes [Lactobacillales]	85	5.0E-03	EA014133	X	
	4D07	t_1	272	Polysaccharide biosynthesis protein	Firmicutes [Lactobacillales]	96	2.0E-27	EAN10262	X	
	4D10	t_1	76	Transketolase	Firmicutes [Lactobacillales]	100	2.0E-05	ACA35934	G	
	4E02	t_1	276	Put. GTP pyrophosphokinase	Firmicutes [Lactobacillales]	79	4.0E-29	CAA51353	T	K
	4E06	t_1	161	Membr. carboxypeptidase	Firmicutes [Lactobacillales]	51	2.0E-08	ABJ66971	M	
	4E07	t_1	171	Hypothetical protein	Firmicutes [Lactobacillales]	78	2.0E-16	EDT48533	M	
	4F01	t_1	136	Curved DNA-binding protein CbpA	Firmicutes [Lactobacillales]	100	7.0E-18	BAE97513	O	
	4F12	t_1	167	Hypothetical protein	Proteobacteria [Enterobacteriales]	98	5.0E-26	ZP_02813297	S	
	4G02	t_1	178	ATP-dependent protease	Proteobacteria [Enterobacteriales]	98	3.0E-25	EDS92168	O	
	4G07	t_1	324	Macrolide transporter subunit MacA	Proteobacteria [Enterobacteriales]	100	2.0E-51	ZP_02790090	M	
	4G08	t_1	304	Organic radical activating enzyme	Proteobacteria [Enterobacteriales]	100	2.0E-54	ZP_00733569	O	
	4H01	t_1	237	Melibiose carrier protein	Proteobacteria [Enterobacteriales]	98	6.0E-40	ACB16587	G	
	4H02	t_1	205	7,8-Dihydropteroate synthase	Proteobacteria [Enterobacteriales]	98	3.0E-30	ZP_02816510	H	
	4H03	t_1	194	Major facilitator superfamily	Proteobacteria [Enterobacteriales]	100	5.0E-08	ZP_02904139	O	C
II→	4H06	t_1	156	Putative receptor	Proteobacteria [Pseudomonadales]	76	5.0E-16	AAY91513	G	
	4H12	t_1	107	Transposase for insertion seq. element IS712A	Firmicutes [Lactobacillales]	85	7.0E-10	CAL98677	L	
	7A10	t_1	362	Oligopeptide ABC transporter, permease prot.	Firmicutes [Lactobacillales]	89	2.0E-51	EAO74875	EP	
	7A11	t_1	371	Amino acid ABC transporter, permease prot.	Firmicutes [Lactobacillales]	94	4.0E-54	EAO73857	E	
	7B01	t_1	320	Excinuclease ATPase subunit A	Firmicutes [Lactobacillales]	94	5.0E-32	AAV61347	L	
	7B03	t_1	273	Formamidopyrimidine-DNA glycosylase	Firmicutes [Lactobacillales]	77	8.0E-24	AAV62216	L	
	7B06	t_1	233	Hypothetical protein	Firmicutes [Lactobacillales]	66	8.0E-09	EDT4743	X	
	7B07	t_1	147	Bacteriocin self-immunity prot.	Firmicutes [Lactobacillales]	95	2.0E-17	AAV6188	V	
	7B11	t_1	232	Phospho-sugar mutase	Firmicutes [Lactobacillales]	97	2.0E-32	ABJ66418	G	
	7C02	t_1	205	Prot. of unknown function, prob. transporter	Firmicutes [Lactobacillales]	82	6.0E-17	EAN08907	R	
	7C04	t_1	188	Predicted peptidase	Firmicutes [Lactobacillales]	69	6.0E-11	AAF89976	Q	
	7C08	t_1	169	Glucokinase	Firmicutes [Bacillales]	40	7.0E-03	ZP_02595656	K	G
	7C10	t_1	165	Put. replication initiation prot.	Firmicutes [Lactobacillales]	90	2.0E-21	AAM99131	J	
	7E09	t_1	166	Hypothetical protein	Firmicutes [Lactobacillales]	83	3.0E-17	EDT47967	X	
	7E11	t_1	250	Cons. hyp. prot.; possible cobalt permease	Firmicutes [Lactobacillales]	86	9.0E-32	EDT47067	RS	
	7F03	t_1	202	Metal cation ABC transporter ATP-binding prot.	Firmicutes [Lactobacillales]	64	3.0E-16	YP_001129020	P	
	7F05	t_1	174	Transcriptional regulator MarR family	Firmicutes [Lactobacillales]	80	1.0E-18	EAO74740	K	
	7F07	t_1	184	Amino acid transporter	Firmicutes [Lactobacillales]	63	5.0E-16	EDT47270	E	
	7F12	t_1	168	Hypothetical protein (put. membr. prot.)	Firmicutes [Lactobacillales]	95	3.0E-04	EDT48433	M	
	7G01	t_1	147	C-5 cytosine-specific DNA methylase	Firmicutes [Lactobacillales]	72	1.0E-10	EDU35719	X	
	7G12	t_1	250	Metal-dependent hydrolase	Firmicutes [Lactobacillales]	73	9.0E-26	ABF37123	R	
	7H03	t_1	201	Ribosomal subunit interface protein	Firmicutes [Lactobacillales]	82	5.0E-18	EAO76428	J	

TABLE 6-S2-*continued*

	Sequence	Time point	Length	Highest hit seq	Phylum [order]	ID (%)	e-value	GenBank reference	COG	
	8A01a	t_1	246	Isoleucyl-tRNA synthetase, class I1	Firmicutes [Lactobacillales]	88	2.0E-28	EAP41136	J	
	8A03a	t_1	240	Alcohol dehydrogenase Zn-binding domain prot.	Proteobacteria [Burkholderiales]	44	3.0E-06	ABX19308	C	R
	8A06a	t_1	211	Tn5252, Orf23	Firmicutes [Lactobacillales]	90	2.0E-10	EAP41017	X	
	8A07a	t_1	174	Phosphate regulon trancr. regulatory prot. PhoB	Firmicutes [Lactobacillales]	74	1.0E-13	EAO75408	T	K
	8B07a	t_1	316	Glycosyl hydrolase, family 65	Firmicutes [Lactobacillales]	95	3.0E-51	EAN10841	G	
	8B09a	t_1	256	Short chain dehydrogenase	Bacteroidetes [Bacteroidales]	70	2.0E-16	CAH07830	IQ	R
III→	8B11a	t_1	203	Cons. hypothetical protein	Firmicutes [Lactobacillales]	83	2.0E-21	EDT48500	S	
	8C01a	t_1	143	Hypothetical protein	Firmicutes [Lactobacillales]	75	5.0E-10	EDT46982	R	
	8E12a	t_1	177	Hypothetical protein	Firmicutes [Lactobacillales]	74	5.0E-18	AAM79548	X	
	8F02a	t_1	143	Put. ABC transporter, ATP-binding protein	Firmicutes [Lactobacillales]	75	1.0E-10	AAN58777	V	
	8G08a	t_1	396	Put. diacylglycerol kinase catalytic domain prot.	Firmicutes [Lactobacillales]	61	1.0E-26	EAO73134	I	R
	8G10a	t_1	323	Hypothetical protein; topoisomerase	Firmicutes [Lactobacillales]	72	4.0E-33	NP_734855	L	
	8G11a	t_1	308	Hypothetical protein	Firmicutes [Lactobacillales]	62	3.0E-06	ZP_02919865	X	
	8H02a	t_1	304	Put. transmembrane protein	Firmicutes [Lactobacillales]	58	8.0E-27	AAG18633	GEP	R
IV→	8H03a	t_1	249	Ribosomal subunit interface protein	Firmicutes [Lactobacillales]	88	9.0E-31	EAO76428	J	
	8H05a	t_1	252	Glycosyl transferase, group 1	Firmicutes [Lactobacillales]	66	1.0E-16	EAP40608	M	
	8H08a	t_1	128	Peptidase M23B:CHAP	Firmicutes [Lactobacillales]	87	3.0E-15	ZP_00874768	X	
	8B10b	t_1	385	Metallo-beta-lactamase superfamily protein	Firmicutes [Lactobacillales]	83	2.0E-52	YP_001127757	R	
	8B11b	t_1	314	Put. transcriptional regulator	Firmicutes [Lactobacillales]	60	3.0E-30	NP_720747	K	
	8C06b	t_1	206	Deblocking aminopeptidase	Firmicutes [Lactobacillales]	79	1.0E-04	YP_001450237	G	
V→	8C07b	t_1	174	Sugar hydrolase	Firmicutes [Lactobacillales]	71	6.0E-18	ABX75773	G	
	8C10b	t_1	163	Cons. hypothetical protein	Firmicutes [Lactobacillales]	88	2.0E-19	EAO77873	S	
	8E10b	t_1	324	Hypothetical protein	Firmicutes [Lactobacillales]	71	5.0E-28	NP_734855	L	
	8F01b	t_1	245	Isoleucyl-tRNA synthetase	Firmicutes [Lactobacillales]	87	8.0E-35	ABP89455	J	
	8F03b	t_1	268	Short chain dehydrogenase	Verrucomicrobia [Verrucomicrobiales]	73	8.0E-19	CAH07830	IQ	R
	8G06b	t_1	130	Hypothetical protein	Firmicutes [Lactobacillales]	86	2.0E-09	EDT47725	H	
	8H07b	t_1	218	Putative NAD(P)H-flavin oxidoreductase	Firmicutes [Lactobacillales]	81	9.0E-21	AAN59244	C	

CELLULAR PROCESSES AND SIGNALLING		%
[D]	Cell cycle control, cell division, chromosome partitioning	0.0
[Y]	Nuclear structure	0.0
[V]	Defense mechanisms	3.4
[T]	Signal transduction mechanisms	3.0
[M]	Cell wall/membrane/envelope biogenesis	6.5
[N]	Cell motility	0.0
[Z]	Cytoskeleton	0.0
[W]	Extracellular structures	0.0
[U]	Intracellular trafficking, secretion, and vesicular transport	0.0
[O]	Posttranslational modification, prot. turnover, chaperones	3.0

INFORMATION STORAGE AND PROCESSING		%
[J]	Translation, ribosomal structure and biogenesis	6.0
[A]	RNA processing and modification	0.0
[K]	Transcription	3.0
[L]	Replication, recombination and repair	4.3
[B]	Chromatin structure and dynamics	0.0

METABOLISM		%
[C]	Energy production and conversion	2.6
[G]	Carbohydrate transport and metabolism	9.3
[E]	Amino acid transport and metabolism	4.5
[F]	Nucleotide transport and metabolism	0.0
[H]	Coenzyme transport and metabolism	1.7
[I]	Lipid transport and metabolism	1.9
[P]	Inorganic ion transport and metabolism	3.2
[Q]	Sec. metabolites biosynthesis, transport and catabolism	1.4

POORLY CHARACTERISED		%
[R]	General function prediction only	6.0
[S]	Function unknown	3.9

NO FUNCTIONAL CLUSTER ASSIGNED		%
[X]	No assignment of functional group possible	36.2

TABLE 6-S3 Overview of 14 primer pairs used to investigate the presence of the encoding genes corresponding to transcripts retrieved by cDNA-AFLP analysis in in the microbiome of ileostomy effluent and faecal samples.

Transcript	Time point	Highest hit sequence (order)	COG	Product (bp)	Forward primer (5' → 3') Reverse primer (5' → 3')	Tm (°C)
4A05	t_0	Sensory box histidine kinase	T	140	GTCACGACGGTTCAAGACAA	59.7
		(Lactobacillales)			AAACCGCGCTACAACAACTC	59.4
4E01	t_0	Putative bèta-galactosidase	I	186	TCCACAAATTTGCAACTGGA	60.1
		(Lactobacillales)			CCTAGGGGAATGATGGTTTG	59.2
4G01	t_0	Hypothetical protein	X	171	CGACTAACTCAGGCTGCTCA	59.3
		(Latobacillales)			AGCTTTGCAAAATGGTACGG	60.1
4G09	t_0	Sortase B family protein	X	201	CATGACGAAGAGCCAAGTCA	60.0
		(Lactobacillales)			GAGGCGACAAACCCTGATAC	59.6
4H11	t_0	Putative transcriptional regulator	X	106	AGCGCTATAACGGGCTTTG	60.3
		(Bacteroidales)			ACCATGAAGGTTTCCCCTTG	61.1
4C01	t_1	Hypothetical protein	X	169	TTGTCAAGTTATCCGTCCAG	58.6
		(Lactobacillales)			GCCACCCTGCATAATCCTAA	59.9
4C02	t_1	Homoserine O-succinyltransferase	E	204	GGGCTAACAACGTCTTGTGAA	60.2
		(Lactobacillales)			TGCTCCTATTGAAACTTTGGATT	59.2
4D07	t_1	Polysaccharide biosynthesis family	X	197	TTTTTGGACAGAACGGCTTT	59.7
		(Lactobacillales)			GACGAAGACGAATCACTGCAT	60.3
4E02	t_1	Putative GTP pyrophosphokinase	T,K	196	CAAATTCCGCTTCAATGTCA	59.7
		(Lactobacillales)			AAGGCATTGGATTACGCAAC	60.0
4G07	t_1	Macrolide transporter	M	196	CACCAATCACCACTTCATCG	60.0
		(Enterobacteriales)			CAGCTCACCGATGTGAAAAA	59.8
4G08	t_1	Org. radical activating enzyme	O	199	TCAGTCAGTGGCAGCAAATC	60.0
		(Enterobacteriales)			ACCAAACACACCTGGGAAAA	60.3
4H01	t_1	Melibiose carrier protein	G	211	CGACCAGAAGGGAATATCCA	59.9
		(Enterobacteriales)			TCCGATTATGGGATGGATTG	60.5
4H02	t_1	7,8-Dihydropteroate syntase	H	180	CAATTCTCTTTTGCGATACCC	58.7
		(Enterobacteriales)			AACGATATCCGCTCCCTTTC	60.4
4H03	t_1	Major facilitator superfamily	O,C	158	AAACGCCGAGGCAAAGAT	60.7
		(Enterobacteriales)			TTAACGATGCAAGCCTGGTC	61.2

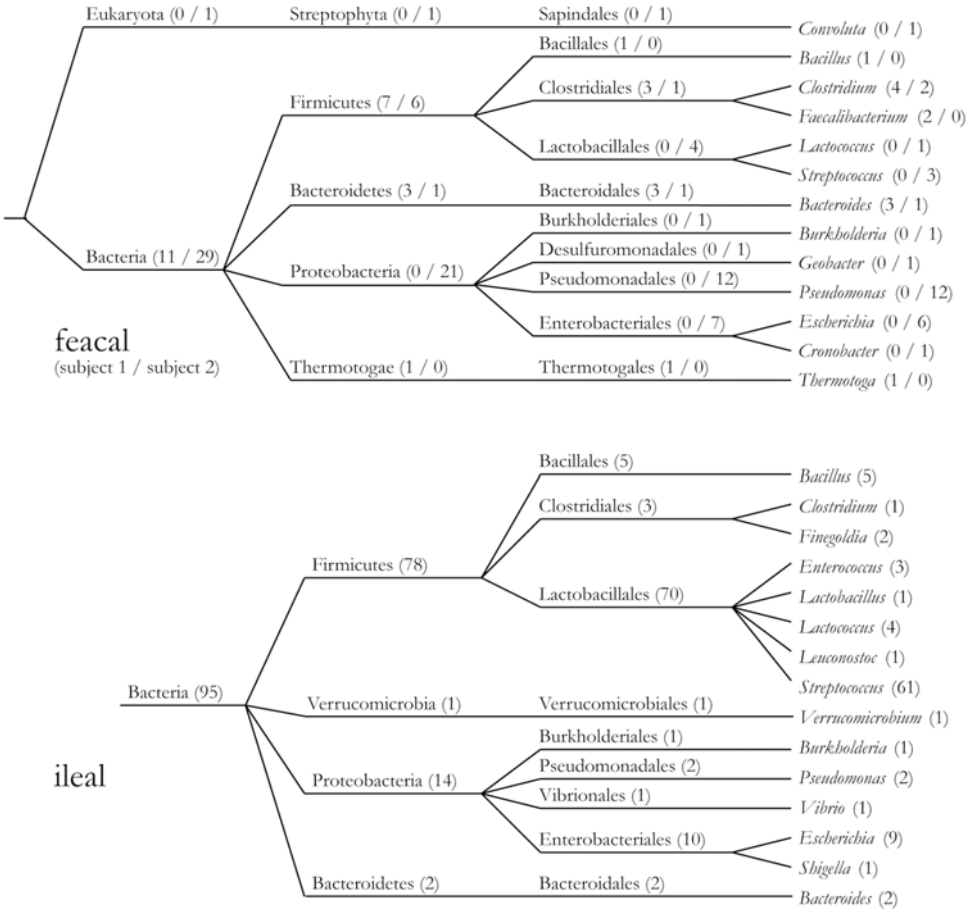


FIGURE 6-S2 The distribution of the transcript sequences retrieved from the mRNA enriched cDNA-AFLP profiles of faecal samples of two subjects (upper cladogram) and of transcript sequences from ileostomy effluent samples of a subject with an ileostomy (lower cladogram) among different phylogenetic groups (kingdom, phylum, order and genus) based on best hit.

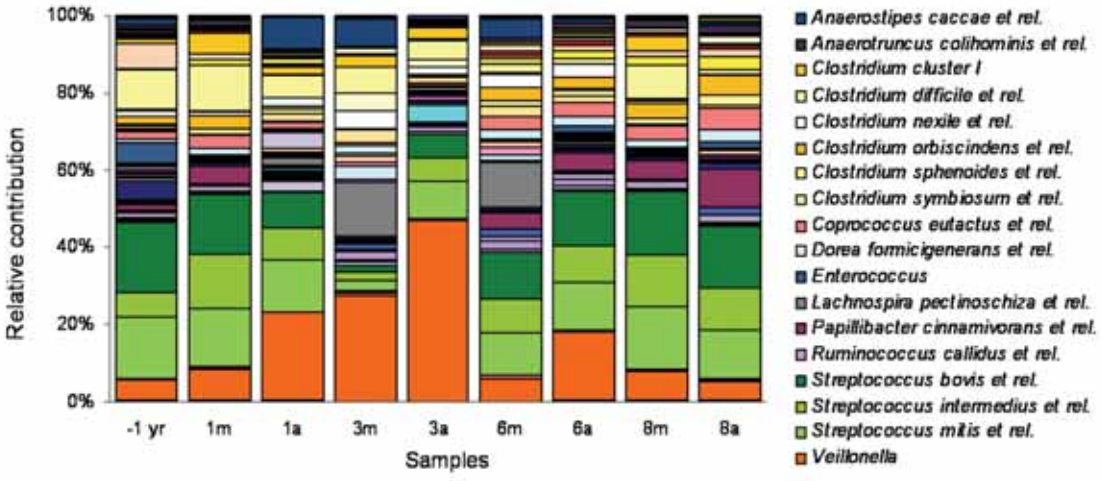


FIGURE 8-2 Relative contribution of the detected phylotypes with HITChip in nine ileostomy effluent samples collected over time of a healthy ileostomy subject (i4). The day of sampling is indicated under the bars, whereas m = morning and a = afternoon. In the legend, phylotypes that contribute at least 1% to the majority of the profiles are indicated.

CO-AUTHOR AFFILIATIONS

Willem. M. de Vos[1,2,3]
Michiel Kleerebezem[1,2,4]
Erwin G. Zoetendal[1,2]
Hauke Smidt[1,2]
Sahar El-Aidy[1,2]
Mirjana Rajilić-Stojanović[1]
Hans G.H.J. Heilig[1]
Freddy J. Troost[2,5]
Eline S. Klaassens[6]
Jos Boekhorst[7]
Joël Doré[8]
Marion Leclerc[8]
Fabienne Béguet[8]

[1] Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, the Netherlands
[2] Top Institute Food and Nutrition (TIFN), PO Box 557, 6700 AN Wageningen, the Netherlands
[3] Department of Basic Veterinary Sciences, PO Box 66, FN-00014, University of Helsinki, Helsinki, Finland
[4] NIZO food research, Health and Safety Department, PO Box 20, 6710 BA Ede, the Netherlands
[5] Department of Internal Medicine, div. of Gastroenterology-Hepatology, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands
[6] CSIRO Livestock Industries, 306 Carmody Road, St Lucia Queensland 4067, Australia
[7] Bioinformatics, Department of Biology, Utrecht University, PO Box 80.056, 3508 TB Utrecht, the Netherlands
[8] INRA, Unit for Ecology and Physiology of the Digestive Tract, Domaine de Vilvert, F-78352 Jouy-en-Josas, France

DUTCH SUMMARY - NEDERLANDSE SAMENVATTING

De microbiota en de mens

De focus van het onderzoek beschreven in dit proefschrift is op de bacteriën die leven in de dunne darm van volwassen mensen. De dunne darm begint meteen na de maag, en gaat na ongeveer zeven meter bij de blinde darm over in de dikke darm. Bij de start van mijn promotie-onderzoek richtte veel onderzoek aan de microbiota* van het menselijke maagdarmkanaal zich op de dikke darm, waarbij feces gebruikt werd als uitgangsmateriaal om de bacterie-samenstelling te onderzoeken. Met dank aan deze onderzoeken is inmiddels bekend welke groepen micro-organismen de dikke darm bevolken. Daardoor weten we dat elk volwassen persoon zijn eigen unieke microbiota heeft, welke vrij stabiel is over de tijd. De functies die de microbiota in de dikke darm vervullen zijn zeer divers, zoals berscherming tegen invasie van pathogene bacteriën, de vertering van voor de mens onverteerbare voedselresten, de produktie van vitamine K en de produktie van aminozuren. Bij aanvang van dit project was de samenstelling van de microbiota in de dunne darm nauwelijks beschreven. Dit komt met name doordat het erg lastig is om representatieve monsters te krijgen uit de dunne darm. Via de mond moeten eerst de slokdarm en de maag gepasseerd worden, terwijl er via de anus een lange weg door de dikke darm afgelegd moet worden. Daardoor is het verkrijgen van materiaal uit de dunne darm erg invasief voor de vrijwilliger.

Personen met een ileostoma

Om monsters uit de dunne darm te verkrijgen heb ik mensen met een ileostoma benaderd. Bij deze mensen is de dikke darm operatief verwijderd, vaak vanwege ontstekingen (bijv. ziekte van Crohn) of vanwege kanker aan de dikke darm. Deze mensen kunnen na een dergelijke operatie dankzij een ileostoma volledig herstellen, zonder gebonden te zijn aan permanente medicatie. De inhoud van de dunne darm kan in dit geval het maagdarmkanaal verlaten via een stoma in de buikwand, waarna het zogenaamde effluent opgevangen wordt in een stoma-zakje. Dit effluent heb ik gebruikt om de bacterie-samenstelling in de dunne darm te bepalen en daarnaast om de functionaliteit van de bacteriën te onderzoeken. Het grote voordeel van het gebruik van ileostoma effluent is dat dit materiaal erg makkelijk verkregen kan worden zonder dat de persoon in kwestie daar hinder van ondervindt. Bovendien is het mogelijk om over een langere periode monsters te verzamelen.

Identificatie van bacteriën

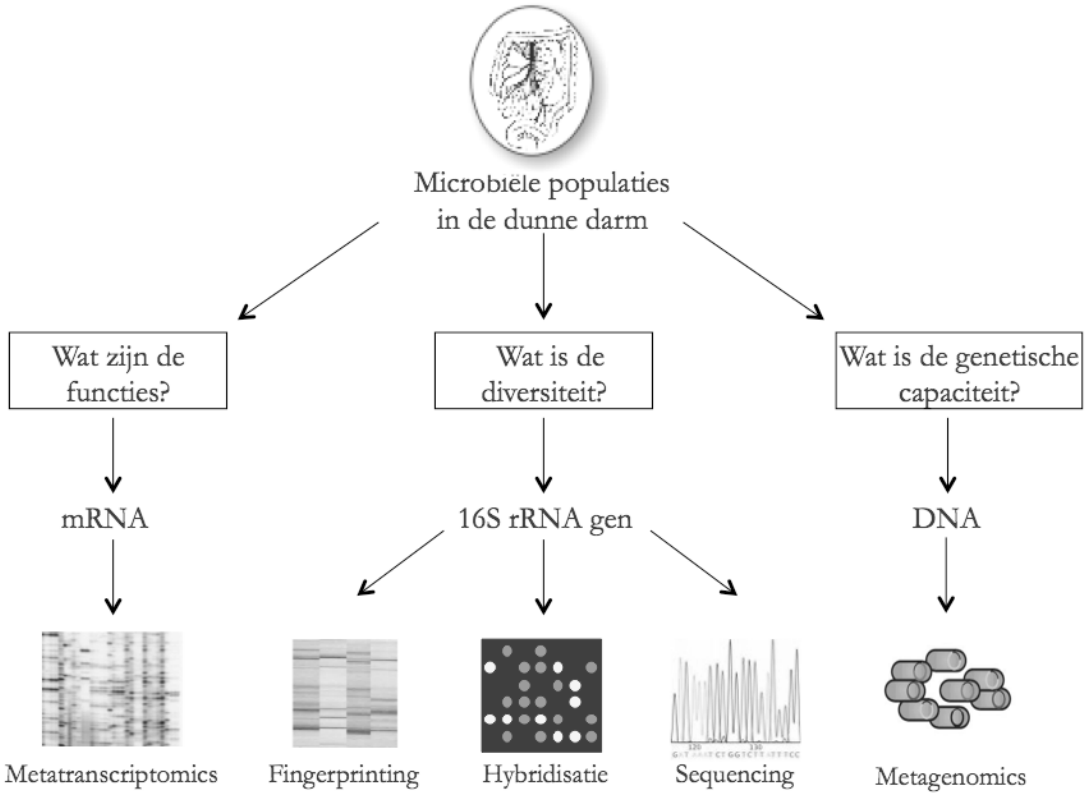
De bacterie-samenstelling in de dunne darm kan op verschillende manieren worden bepaald. Voor de intrede van moleculaire technieken werd dit onderzoek veelal uitgevoerd door bacteriën te kweken, de zogenaamde klassieke microbiologie. Een voorwaarde om een bacterie-soort te kweken is dat zijn groeibenodigdheden bekend zijn. Daar zit het grote probleem bij het bestuderen van complexe bacteriële populaties zoals aanwezig in de dunne darm: het is onbekend wat de individuele groeibenodigdheden zijn van de vele verschillende bacterie-soorten. Bovendien wordt vermoed dat niet alle soorten te kweken zijn. Met behulp van kweektechnieken wordt dientengevolge slechts een incompleet beeld verkregen van de bacterie-samenstelling in de

*Ofschoon van oudsher de term microflora veel gebruikt wordt, is het feitelijk juister om over microbiota te spreken om te verwijzen naar alle kleine organismen die niet tot het planten- of dierenrijk behoren.

dunne darm. Bij de moleculaire identificatie van bacteriën, waarbij wel een compleet beeld wordt verkregen, wordt gebruik gemaakt van verschillen in het erfelijk materiaal (ook wel DNA genoemd) tussen bacterie-soorten. Vaak wordt hier DNA voor gebruikt dat codeert voor de kleine subeenheid van een ribosoom, het 16S rRNA gen.

Microbiële diversiteit in de dunne darm

De samenstelling van de bacteriën in de dunne darm hebben we vooral bestudeerd aan de hand van het hierboven genoemde 16S rRNA gen. We hebben drie verschillende methodes gebruikt om de microbiële diversiteit in de dunne darm in kaart te brengen (Figuur A). Bij fingerprinting wordt gekeken naar een stukje uit het 16S rRNA gen dat verschilt tussen soorten, waarbij elk bandje van de fingerprint een bacterie-soort representeert. Deze methode is vooral geschikt om de dominante bacterie-soorten in verschillende monsters met elkaar te vergelijken. Het toepassen van fingerprinting op de ileostoma effluent monsters liet zien dat er bij individuele personen vrij grote fluctuaties optreden in de microbiota over de tijd.



FIGUUR A Overzicht van de verschillende onderzoeksvragen bij aanvang van mijn project, met de daarbij gebruikte benaderingen om deze te beantwoorden.

Vervolgens hebben we hybridisatie toegepast; hierbij zijn stukjes van het 16S rRNA gen van bekende bacteriën uit de darm aan een objectglaasje gekoppeld. Op deze glaasjes worden stukjes van het 16S rRNA gen van een onbekend monster gebracht, in dit geval van de ileostoma effluent monsters. Alle soorten die aanwezig zijn in de onbekende monsters kunnen vervolgens gedetecteerd worden doordat we de 16S rRNA stukjes in deze monsters een fluorescent label

hebben gegeven. Met deze techniek kan de microbiële samenstelling in een monster veel gedetailleerder in kaart worden gebracht dan met behulp van fingerprinting. Uit de hybridisaties bleek dat de *Streptococcus*- en *Veillonella*-soorten dominant aanwezig zijn in ileostoma effluent.

Ook hebben we sequencing van 16S rRNA genen toegepast. De verkregen sequenties werden vergeleken met databases die tezamen ruim 800.000 16S rRNA sequenties bevatten, om te bepalen van welke bacterie-soort de verkregen 16S rRNA sequenties afkomstig waren. Hiermee hebben we de resultaten van de hybridisatie techniek bevestigd en daarnaast aangetoond dat 23% van de bacterie-soorten aanwezig in het ileostoma effluent tot op heden onbekend is. Bij het vergelijken van de bacteriële samenstelling van de monsters van de vijf personen met een ileostoma bleek dat er in alle monsters een groep van acht dezelfde bacterie-soorten aanwezig is. Tot deze groep behoren onder andere de *Streptococcus*- en *Veillonella*-soorten. Dit suggereert dat deze groepen een stabiele kerngroep vormen in de dunne darm van volwassenen.

Microbiële functionaliteit in de dunne darm

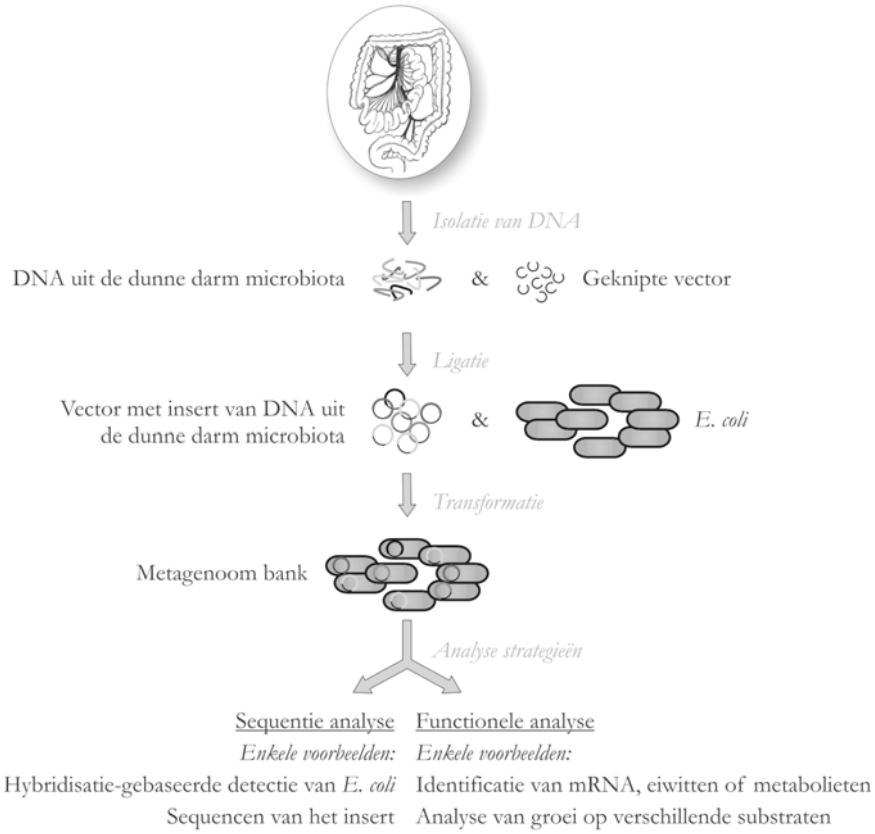
De functionaliteit van bacteriële populaties hebben we onderzocht op mRNA niveau. mRNA staat voor "messenger RNA" en is letterlijk de boodschapper tussen DNA en eiwit. Met andere woorden, het mRNA wordt afgelezen van het DNA, waarbij elk gen een ander mRNA oplevert. Dit mRNA wordt vervolgens gebruikt om een eiwit te maken. In hoofdstuk 6 hebben we onderzocht welke bacteriële mRNAs aanwezig zijn in de ileostoma effluent monsters uit de dunne darm. Vooral mRNAs die coderen voor eiwitten betrokken bij de opname en afbraak van koolhydraten werden veel aangetroffen. Deze mRNAs bleken voor een groot deel (70%) afkomstig te zijn van verschillende soorten behorend tot het genus *Streptococcus*, wat erop duidt dat deze soorten zeer actief zijn in de dunne darm. Een mogelijke verklaring voor die activiteit zijn de omgevingsfactoren van de dunne darm waarin deze streptococcen moeten overleven. De stroomsnelheid in de dunne darm is vrij hoog, wat het noodzakelijk maakt om voedingsstoffen snel op te nemen. En juist streptococcen zijn heel snel in het opnemen van suikers.

Metagenoom analyse

Een "microbieel metagenoom" betekent letterlijk: alle genetische informatie uit een microbieel ecosysteem. Een dergelijk metagenoom kan onderzocht worden met behulp van een metagenoom bank. Deze bank wordt gemaakt door het DNA te isoleren uit een bacterie-populatie van interesse. Een stapsgewijze weergave van de daaropvolgende constructie van een metagenoom bank, alsmede de analyse hiervan, is weergegeven in Figuur B.

Ik heb een metagenoom bank onderzocht die gemaakt is van de bacteriën uit de slijmlaag (of mucuslaag) van de dunne darm van een volwassen gezond persoon. Aangezien we geïnteresseerd waren in de soorten die in staat waren mucus als voedingsbron te gebruiken, hebben we de hele metagenoom bank getest voor groei op mucus. 1,1% van alle *E. coli* cellen uit de metagenoom bank bleek te kunnen groeien op mucus, terwijl *E. coli* dit normaal niet kan. Twee *E. coli* cellen die groei vertoonden op mucus hebben we verder geanalyseerd. We hebben gekeken naar de snelheid waarmee het mucus werd omgezet, alsmede naar de DNA-volgorde van het stuk metagenoom. Hieruit bleek in het eerste geval dat de *E. coli* cel DNA van *Bacteroides thetaiotaomicron* bevatte, waaronder twee genen die mogelijk betrokken zijn bij de groei

op mucus. De andere *E. coli* cel bevatte een stuk DNA van *Enterococcus faecalis*, een soort waarvan niet eerder is aangetoond dat deze mucus kan gebruiken als voedingsbron. Hiermee hebben we laten zien dat de gebruikte functionele analyse geschikt is om mucus-groeiers op te sporen, waarbij tevens de betrokken genen geïdentificeerd kunnen worden.



FIGUUR B Overzicht van de constructie en analyse-mogelijkheden van een microbiële metagenoom bank.

De belangrijkste conclusies

Met de gekozen aanpak hebben we laten zien dat *Streptococcus* en *Veillonella* de dominante soorten in de microbiota van de dunne darm zijn. Bovendien blijkt *Streptococcus* zeer actief te zijn, met name in het opnemen en omzetten van suikers. Verder treden er grote fluctuaties op in de microbiota van de dunne darm, zelfs binnen één dag. Deze fluctuaties lijken echter rondom een stabiele kerngroep te schommelen, welke in ieder van de vijf onderzochte personen aanwezig is.

Met de gekozen strategie voor de functionele analyse van de microbiota in de dunne darm hebben we veel sequenties verkregen die we niet terug konden vinden in databases waarin alle genetische informatie van bekende bacteriën is opgeslagen. Momenteel zijn er grote projecten gaande die meer dan 100 bacterie-soorten uit de darm volledig sequencen. Bovendien worden de sequencers steeds sneller, efficiënter en goedkoper. De verwachting is dat met de groeiende hoeveelheid sequentie-informatie de functionele analyse van complexe populaties, zoals in de dunne darm, makkelijker zal worden. Daardoor zal het in de toekomst wellicht mogelijk worden op basis van de aanwezige microbiota in de darm individuele personen van voedingsadviezen te voorzien.

ACKNOWLEDGEMENTS - DANKWOORD

Zonder de hulp van een heel aantal mensen was het boekje er niet in deze vorm uit komen te zien, en ik wil graag van de gelegenheid gebruik maken om eenieder hier te bedanken. Willem, bedankt voor het in mij gestelde vertrouwen om dit project tot een goed einde te brengen, en daarnaast voor de nuttige feedback die je op cruciale momenten hebt gegeven. Met je ongeëvenaarde energie en daadkracht ben je een ware inspiratiebron. Michiel, ofschoon van afstand heb je veel bijgedragen aan de inhoud van het uitgevoerde onderzoek en wist je goed knopen door te hakken als er beslissingen genomen moesten worden. Daarnaast wil ik je bedanken voor de kritische feedback op mijn manuscripten. Erwin, ik wil je bedanken voor de nuttige adviezen aangaande het practische werk, en je snelle feedback op mijn manuscripten. Vooral tijdens de eindsprint van mijn promotie-onderzoek bleek je een onmisbare schakel. Hauke, bedankt voor je niet aflatende interesse in mijn project, en alle tijd en energie die je er aan besteed hebt. Het is erg fijn om een vakgroepleider te hebben die op zo'n enthousiaste en positieve manier sturing geeft aan de groep. Elaine, dankzij jou heb ik een vlotte start van het project gehad. Tenslotte wil ik Cindy als opvolgend teamleider van het WCFS-project C007 bedanken voor de getoonde interesse in zowel de inhoudelijke als de persoonlijke aspecten van de uitvoering van mijn proefschrift.

Eline, je was met je uitgesproken karakter en doortastendheid duidelijk aanwezig in de groep en met dank aan je organisatie-talent hebben we heel wat leuke borrels en feestjes meegemaakt. Veel plezier en vooral gezondheid down under! Mirjana, thanks for your patience during the analysis of the HITChip results, but more importantly, for sharing usefull advises concerning thesis writing as well on the personal level. Muriel, thanks for sharing your knowledge about mucin, as well as about the French culture and language. Hans, jij bent van onschatbare waarde voor het lab, en hebt me vaak met raad en daad bij gestaan. Ook heel erg bedankt voor je bijdrage aan het derde hoofdstuk. Sahar, I am very happy that you choose to perform your internship under my supervision, resulting in essential contributions to two chapters in this thesis. I feel honoured that you are my paranymph, and wish you good luck with your project! Petia, thanks for being my neighbour in the office and for the exchange of many thoughts and ideas, also concerning motherhood. I am very happy that you are next to me during my defense and wish you all the best with your own defense. I would also like to thank all other "2012" roomies for the nice atmosphere: Kaouther, Maaike, Meta, Odette, Susana and Paula. Jeroen, bedankt voor je bijdrage aan mijn project tijdens je stage. Furthermore, I would like to thank all other MolEco's, including Alcina, Anne, Anna, Carmelina, David, Diana, Erja, Farai, Haisa, Hermien, Hoa, Ineke, Jana, Jelena, Joan, Joana, Kathrin, Kees, Lidia, Mark and Mark, Mahmut, Mariana, Markus, Martha, Milkha, Neslihan, Pauline, Pavel, Prapha, Rina, Rocio, Rozalin, Sara, Sebastian, Sergey, Stephanie, Suzanne, Thomas, Tom, Toshio, Vesela, Wilma and Ye for the nice time spend. Tot slot wil ik ook graag Anja, Ans, Francis, Jannie, Joop, Nees, Renee en Wim bedanken voor alle ondersteuning die het uitvoeren van experimenten bij Microbiologie mogelijk maakt.

Alle collega's betrokken bij het WCFS-project C007 wil ik bedanken voor de nuttige feedback tijdens de projectbijeenkomsten: Andrea, Bart, Freddy (ook voor je input bij hoofstuk drie), Jolanda, Maaike, Maria, Michiel W. (en voor je bioinformatica bijdrage), Peter van B.,

Roger, Peter B., Roland (ook voor de bioinformatica), Sonia en Saskia. Hetzelfde geldt voor de collega's van WCFS-project C012: Koen, Albert, Annet, Bart, Daisy, Henrike, Johanne, Petia, Saed, Steven en Tao. Verder wil ik de collega-WCFS AIO's Johan, Kerensa, Neleke, Sandra, Wendy en Ynte uit het Top Forum bedanken voor de gezellige vergaderingen inclusief etentjes bij Unitas. Jos, bedankt voor je bio-informatica bijdrage en de bijbehorende eet- en schaats-uitjes.

Il y a plusieurs personnes dans le labo d'INRA à qui je voudrais dit merci: Joël, pour ta confiance en moi et des propositions bien pensées afin d'améliorer l'étude metagenomique. Marion, une personne très optimiste et professionelle comme toi est très plaisante pour collaborer. Aussi dans le labo, j'ai de bons souvenirs des heures de "shlouping" et je suis très heureuse que tu aies voulu participer à ma soutenance afin de discuter encore avec moi. Fabienne, pour ton aide, tes explications des mots français et le bon diner chez toi. Hervé, pour des bonnes tasses de café et notre excursion pour acheter des vélos. Michèle, pour ta patience pour m'écouter quand j'ai eu besoin de quelque chose. Nelly, il y avait les plaques, les peignes et des litres LB innombrables que tu as autoclavés pour moi. Grâce à Catherine, Florence, Françoise, Julie, Karine, Marlène, Omar, Pascale, Rafael, Robèr (ook bedankt voor de NMR-profielen), Sandra et Violaine pour les bon temps pendant les déjeuners, les apératifs et les barbecues. Aussi tous les gens connectés au "Labhealth", je voudrais remercier pour les merveilleux neuf mois que j'ai passés à Jouy-en-Josas: Emanuelle, Maarten, Véronique, Aleksandr, Fabrizio, Daniel, Jasna, Michael, Olivier, Oriana, Stathis et Silvia.

De acht anonieme vrijwilligers met een ileostoma of pouch die belangeloos mee wilden werken toen ik ze benaderde voor monsters wil ik hartelijk bedanken. Jullie bijdrage was essentieel voor de totstandkoming van dit proefschrift.

Ik heb in mijn vrije tijd heel wat uurtjes aangenaam (ont)spannend doorgebracht bij het spelen van spelletjes met Eline, Meta, Kees, Bernd, Anne-Meint, Mahmut, Mark en Diana. Verder wil ik Adrien, Anneke, Elise, Frederiek, Hajo, Hanneke, Ina, Jeroen, Karin, Kobus, Louisa, Mieke, Maaike, Mara, Marc, Margreet, Moses, Neeli, Nicole, Sandra, Stanly, Tessa, Thies, Wiebe en Wouter bedanken voor de gezellige en waardevolle tijd die ik met jullie heb door gebracht. Ook wil ik alle Graspiepers (Barbara, Javier, Michiel, Miranda, Ronald, Sandra, Silvia, Vinije en Vivian) bedanken, onder meer voor de vele gezellige uren die we hebben doorgebracht bij het prepareren en consumeren van allerlei gerechten.

Tenslotte wil ik mijn schoonouders bedanken voor de ondersteuning en het begrip als ik weer eens niet kwam opdagen op een verjaardag, vooral tijdens de laatste drukke maanden. Anja, Marthijn, Magnus en Joek, elk bezoek van of aan jullie is een feestje, fijn om zo'n energieke schoonfamilie te hebben. Berto en Johanneke, ver weg in het noorden, bedankt dat jullie altijd geïnteresseerd bleven in mijn bacterie-verhalen en succes bij de uitvoering van jullie Nieuw-Zeeland plannen. Lieve pap en mam, bedankt voor alle steun bij elke keuze die ik tot dusverre heb gemaakt. Fijn om te weten dat jullie altijd voor me klaar staan.

Lieve Hendrik, wat ben ik blij dat je de afgelopen tijd zoveel van je geduld hebt laten zien en me zelfs tot en met het formuleren van de tiende stelling hebt geholpen. Dat liefde geen afstand kent hebben we in Canada en Frankrijk mogen ondervinden en ik kijk nu al uit naar het volgende land dat we samen zullen bezoeken! Lieve Kyra, wat een vrolijk en ondernemend meisje ben je en wat een voorrecht dat ik je mama mag zijn.

VLAG PhD EDUCATION FORM

Discipline specific courses

Ecophysiology of the gastrointestinal tract, VLAG, Wageningen	2003
Bioinformatics technology I, Wageningen	2003
Programming in Python, RU Nijmegen	2004
Safe handling with radioactive materials and sources, Wageningen	2005
Metabolomics, VLAG/EPS, Wageningen	2005

General courses

Time planning and project management, PE&RC, Wageningen	2003
PhD week, VLAG, Bilthoven	2004
Scientific writing, CENTA, Wageningen	2004
Career perspectives, Meijer & Meijaard, Wageningen	2007

Meetings

Darmendag Wageningen	2003
Darmendag, Utrecht (poster presentation)	2004
Mini-symposium on nutrition and gut flora, Windermere, UK (oral presentation)	2004
Scientific Spring meeting NVvM, Arnhem (oral presentation)	2005
Symposium LAB-8, Egmond aan Zee (poster presentation)	2005
RRI INRA symposium on Gut Microbiology, Aberdeen, UK (oral presentation)	2006
ISME-11, Vienna, Austria (poster presentation)	2006
GutHEALTH meeting, Amsterdam	2007
SSA GutImpact 3rd Platform Meeting on Foods for Intestinal Health, Haikko, Finland (poster presentation)	2007
Darmendag, Wageningen (keynote speaker)	2007
Seminar TU Delft (invited speaker)	2007
Seminar LabMET UGent, Belgium (invited speaker)	2008

Other activities

Preparing PhD research proposal	2003
WCFS project meetings C007	2003-7
PhD/Postdoc meetings, Laboratory of Microbiology, Wageningen	2003-7
Member of WCFS PhD Top Forum	2003-5
PhD trip, VLAG, Japan	2004
Journal club, Wageningen	2004-5
INRA, Jouy-en-Josas, France (9 months)	2006

CURRICULUM VITAE

Carien Booijink werd geboren op 30 september 1976 te Almelo. Na het doorlopen van de HAVO in Nijverdal, werd in 1995 het VWO diploma behaald op het Pius X college te Almelo. In datzelfde jaar begon ze met de studie Biologie aan de Universiteit Groningen, waar ze in 2000 afstudeerde. Gedurende het eerste deel van haar doctoraalfase deed ze onderzoek aan de functionele karakterisatie van genen in *Bacillus subtilis*, waarvan de genoom-annotatie juist voltooid was. Dit onderzoek werd uitgevoerd bij de vakgroep Moleculaire Genetica te Haren, onder begeleiding van dr. Caroline Eschevins en prof. dr. Gerard Venema. Tijdens het tweede afstudeeronderzoek werd de microbiële diversiteit van organische bodems onderzocht met behulp van t-RFLP, bij de vakgroep Microbiële Ecologie onder begeleiding van drs. Theresa Gomèz en prof. dr. Larry Forney. In 2000 begon Carien aan de postdoctorale opleiding tot docent biologie te Groningen, waarbij ze les gaf aan de Regionale Scholengemeenschap in Steenwijk. Vervolgens werkte ze als docent op Saxion Hogeschool te Deventer. In 2003 begon ze als promovenda bij de leerstoelgroep Microbiologie aan de Wageningen Universiteit onder supervisie van prof. dr. Willem de Vos, waar ze werkte aan de microbiota in de dunne darm. Voor haar project werkte ze gedurende negen maanden met een Marie Curie beurs in Frankrijk bij INRA te Jouy-en-Josas. De resultaten van het promotie onderzoek, dat uitgevoerd werd binnen onderzoeksprogramma drie van Top Institute Food and Nutrition, staan beschreven in dit proefschrift. Sinds het voorjaar van 2008 is Carien werkzaam als docent bij de opleiding Biologie en Medisch Laboratoriumonderzoek op Saxion te Deventer.

PUBLICATIONS

Te Biesebeke, R., Boesten, R., Klaassen, E.S., Booijink, C.C.G.M., de Vries, M.C., Derrien, M., Cohen, D.P.A., Schuren, F., Vaughan, E.E., Kleerebezem, M. & de Vos, W.M. (2004) Microbial functionality in the human gastrointestinal tract. *Microbes & Environments* **19**: 276-280

Zoetendal, E.G., Booijink, C.C.G.M., Klaassens, E.S., Heilig, H.G.H.J., Kleerebezem, M., Smidt, H. & de Vos, W.M. (2006) Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat. Protocols* **1**: 954-959

Zoetendal, E.G., Heilig, H.G.H.J., Klaassens, E.S., Booijink, C.C.G.M., Kleerebezem, M., Smidt, H. & de Vos, W.M. (2006) Isolation of DNA from bacterial samples of the human gastrointestinal tract. *Nat. Protocols* **1**: 870-873.

Booijink, C.C.G.M., Zoetendal, E.G., Kleerebezem, M. & de Vos, W.M. (2007) Microbial communities in the human small intestine: coupling diversity to metagenomics. *Future Microbiol.* **2**: 285-95

Booijink, C.C.G.M., El-Aidy, S., Rajilić-Stojanović, M., Heilig, H.G.H.J. Troost, F.J., Smidt, H., de Vos, W.M., Kleerebezem, M. & Zoetendal, E.G. Population dynamics and diversity of the human ileal luminal microbiota. *Submitted for publication.*

Booijink, C.C.G.M., Boekhorst, J., Zoetendal, E.G., Smidt, H., Kleerebezem, M. & de Vos, W.M. Gene expression profiling of the human GI tract microbiota. *Submitted for publication.*

Booijink, C.C.G.M., Leclerc, M., Béguet, F., Zoetendal, E.G., de Vos, W.M. & Doré, J. Functional analysis of metagenomic libraries of the human ileal mucosa. *Manuscript in preparation.*

Booijink, C.C.G.M., El-Aidy, S., Zoetendal, E.G., Smidt, H., de Vos, W.M. & Kleerebezem, M. Linking *in situ* activity in the human ileum to microbial identification. *Manuscript in preparation.*

Booijink, C.C.G.M., Lepercq, P., Juste, C., Kemperman, R., Beauvallet, C., Kleerebezem, M., de Vos, W.M., Doré, J. & Leclerc, M. Riboflavin production by a human distal ileum metagenomic clone, related to the *Firmicutes* species, *Faecalibacterium prausnitzii*. *Manuscript in preparation.*

El-Aidy, S., Booijink, C.C.G.M., Zoetendal, E.G. & Kleerebezem, M. Evaluation of RING-FISH for linkage between genes and microorganisms in complex niches. *Manuscript in preparation.*

COLOFON

Printed by: GVO drukkers & vormgevers B.V. | Ponsen & Looijen, Ede, the Netherlands
Cover design: Hetty Booijink-Otten