Mouse gut microbiomics of short chain fatty acid metabolism and mucosal responses

Floor Hugenholtz

Thesis committee

Promotors
Prof. Dr H. Smidt
Personal chair in the Laboratory of Microbiology
Wageningen University

Prof. Dr M. Kleerebezem
Personal chair in the Host Microbe Interactomics Group
Wageningen University

Other members
Prof. Dr E. Smid, Wageningen University
Dr J. Doré, National Institute for Agricultural Research, INRA, France
Prof. Dr D.J. Reijngoud, University of Groningen
Dr D Bosscher, Cargill, the Netherlands

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

Mouse gut microbiomics of short chain fatty acid metabolism and mucosal responses

Floor Hugenholtz

Thesis
submitted in fulfillment of the requirement for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 23 January 2015
at 13.30 p.m. in the Aula.

Floor Hugenholtz
Mouse gut microbiomics of short chain fatty acid metabolism and mucosal responses, 208 pages

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summaries in Dutch and English

ISBN 978-94-6257-179-2

Summary

The microbiota of the gastrointestinal (GI) tract plays a key role in the digestion of our food. The human gut microbiota can be studied using in vitro and animal models. In this thesis the mouse model is used to study the microbiota interaction with the diet and the host in different regions along the GI tract. These interacting microbes in the GI tract of humans and other mammals yield a wide range of metabolites, among which the short chain fatty acids (SCFA), in particular butyrate, acetate, and propionate, are the most abundant products of carbohydrate fermentation. Fermentable carbohydrates can modify the composition of the gut microbiota and change the SCFA concentrations in the gut. Opportunities for increasing specific SCFA by targeting their producers with carbohydrates are discussed. Five different fibres – resistant starch, inulin, fructooligosaccharides, arabinoxylan and guar gum – are tested for their modification of the mucosal tissue transcriptome, luminal microbiota composition and SCFA concentrations in the murine colon. The fibres inulin, fructooligosaccharides, arabinoxylan and guar gum led to increased SCFA concentrations and induced similar changes in relative abundance of microbial groups as determined by the MITChip, a phylogenetic microarray targeting the 16S ribosomal RNA of mouse intestinal microorganisms. Furthermore, these four fibres induced regulation of overlapping sets of genes in the mouse intestinal mucosa, where the transcription factor PPARγ was predicted to be a prominent upstream regulator of these processes. Multivariate data integration revealed strong correlations between the expression of genes involved in energy metabolism and the relative abundance of bacteria belonging to Clostridium cluster XIVa. Similar analyses were done for the caeca of the same mice, and were complemented with metatransciptome analyses. To comprehensively analyse RNAseq data of complex natural microbial communities, a de novo metatranscriptome assembly pipeline was developed and applied to unravel the activity profiles of the microbiota residing in the mouse cecum. This revealed distinct contributions of bacterial families to the fermentation of fibres into SCFA, involving the *Bifidobacteriaceae*, *Lachnospiraceae*, *Clostridiaceae*, *Bacteroidaceae*, *Erysipelotrichaceae* and *Ruminococcaceae* in some or all stages of the overall fibre fermentation activity. All families expressed genes encoding enzymes involved in the production of SCFA in different ratios. Specifically, butyrate producing bacteria correlated with a set of host genes involved in processes such as energy metabolism, transcriptional regulation and the mucosal immune system.

In addition to complex carbohydrates, amino acids derived from dietary proteins can also serve as substrates for SCFA formation, leading to expansion of the fermentation end-product palet by including branched-SCFA. The long-term effects of high protein-diets on microbial community composition and activity were analysed. The caecal microbiota composition was changed by the high dietary protein. Most of the gene functions detected by metatranscriptomics in these caecal samples were assigned to the *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae*. High protein diets induced a decrease of *Lachnospiraceae* activity, but stimulated the

activity of the *Erysipelotrichaceae*, while the *Clostridiaceae* appeared to express the broadest range of amino acid metabolism associated pathways.
In conclusion, this thesis describes dietary interventions to modulate the mouse intestinal microbiota and mucosa. The data provides expansion of the knowledge on interactions between the diet, microbiota and host. This information can be used to optimize the design and validation on dietary intervention studies in humans.

Table of contents

Introduction 1

Studying the mammalian intestinal microbiome using animal models 11

Modulation of the microbial fermentation in the gut by fermentable carbohydrates. 25

Distinct responses of mucosal transcriptional profiles and luminal microbiota composition and SCFA concentrations to dietary fibers in murine colon. 39

Functional profiling of unfamiliar microbial communities using a validated de novo assembly metatranscriptome pipeline 65

Linking the fate of dietary fibres in the murine caecum to microbial transcriptome patterns 93

Fermentation of dietary milk protein by murine gut microbiota 127

General Discussion 151

References 163

Appendices 187

Chapter 1

Introduction

General introduction

The food we consume every day is largely digested by ourselves, though still an undigested fraction ends up in the large intestine, and comprises mainly of dietary fibres, dietary proteins and host derived products such as mucus (Macfarlane *et al.* 1988, Cummings and Macfarlane 1997, Tasse *et al.* 2010, Flint *et al.* 2012a). These components can, to a large extend, be fermented in the large intestine by the gut microbiota. Human individuals can harbor over 150 different microbial species in their gut, which collectively encode more than 100-fold more non-redundant genes than there are in the human genome (Backhed *et al.* 2005, Ley *et al.* 2006, Qin *et al.* 2010).

Not only humans have microbiota in their gastrointestinal (GI) tract, they are also found in the intestine of all other mammalian species that each harbours a distinct microbial composition (Ley *et al.* 2008a), and based on their microbial community and diet, carnivores, omnivores and herbivores can be grouped separately. These groups can be distinguished by increasing microbiota diversity, respectively, which probably reflects the larger diversity of plant derived carbohydrates in the diet of herbivores that stimulates a larger diversity in the microbiota. The differences in composition and diversity of GI tract microbiota in these animal groups indicate that diet and host collaboratively mediate the bacterial composition (Ley *et al.* 2008a, Ley *et al.* 2008b, Van den Bogert 2011), and underline the importance of the GI tract microbiota in degrading non-digestible plant polysaccharides. These polysaccharides are converted by the bacteria into predominantly short chain fatty acids (SCFA), which the host uses to a large extent as energy source. In herbivores, the SCFA can provide up to 85% of the total energy intake of the animal, whereas in omnivores, especially humans, the fermentation of non-digestible polysaccharides is estimated to contribute only up to 10% of the total energy recovery from the diet (McNeil 1984). The energy of the SCFA is the main energy source of the intestinal epithelium in mammals (Cummings 1981, Cummings *et al.* 1987, Bloemen *et al.* 2009, Louis and Flint 2009, Flint *et al.* 2012b).

The health of the host can be significantly modulated by the different intestinal SCFA, which will be further described in **Chapter 3.**

Microbial ecology of the GI tract

The GI tract microbiota can be defined as a microbial community, since there are a collection of organisms co-occurring in the same habitat – the GI tract. This community is not purely composed of bacteria, but also includes archaea and eukaryote microbes (Scanlan and Marchesi 2008, Rajilic-Stojanovic and de Vos 2014). Archaea are more important in some mammals then in others. For instance in the cow-rumen they represent a relatively large fraction of the overall microbiota community, whereas in monogastrics such as humans they are only present at low abundance. The main role of these Archaea in the intestinal ecosystem relates to their capacity to produce methane from hydrogen and carbon dioxide and in case of

acetoclastic methanogens also from acetate. The eukaryotic microorganisms in the GI tract include mostly anaerobic fungi, yeast and *Blastocystis* – a single-celled eukaryote (Scanlan and Marchesi 2008, Scanlan *et al.* 2014), of which the specific role in the GI ecosystem still needs to be further investigated.
The collective of microbes in our intestine is referred to as the GI tract microbiota, of which in this thesis only the bacterial fraction is studied. These bacteria all live together in the intestinal habitat, where they are sometimes classified by their functional role in the ecosystem – fibre degraders, cross-feeders and mucus degrading bacteria. The intestinal habitat encompasses multiple micro-habitats and bacteria can be mucus associated, free-living in the intestinal lumen, and/or dietary fibre associated. The differences of the communities along the GI tract are described in **Chapter 2**.

Analytical methods to study the GI tract microbiota

Woese and Fox suggested in 1977 to use the 16S ribosomal RNA (rRNA) molecule as a molecular marker for the determination of the phylogenetic classification of the prokaryotes (Woese and Fox 1977). They also proposed that prokaryotes should not be classified into a single domain, but should be divided into two domains of life: the Bacteria and Achaea. The 16S rRNA appeared to be an ideal molecule to use as a phylogenetic marker, for several reasons; i) it has an essential function within the ribosome, therefore all bacteria and archaea encode it; ii) it is large enough to carry the amount of phylogenetic information necessary to distinguish one species from the other; iii) it has conserved and highly variable sequence elements due to the structure and catalytic function of the molecule; iv) it is not exchanged via lateral gene transfer. The use of this single genetic marker has revolutionized microbial ecology (Tringe and Hugenholtz 2008, Pace *et al.* 2012), since it is relatively easy to amplify the 16S rRNA encoding genes from environmental DNA. Nowadays, with next-generation sequencing techniques many microbial environments can be studied in depth, using relatively straightforward procedures.
The microbiota composition in the GI tract has also been studied extensively by using 16S rRNA gene targeted approaches. A variety of such techniques is currently employed to study microbial ecosystem composition: Denaturing Gradient Gel Electrophoresis (DGGE), Cloning & Sanger sequencing, Terminal Restriction Fragment Length Polymorphism (T-RFLP; (Prakash *et al.* 2014), FISH, quantitative PCR (qPCR), and phylogenetic microarrays (Deng *et al.* 2008, Rajilic-Stojanovic *et al.* 2009). The last two techniques have been employed in this thesis and are briefly described. qPCR is a quantitative method used in intestinal microbiota studies either to determine the microbiota community density in a sample or to quantify the abundance of a specific bacterial group or gene within the ecosystem. In this thesis microbiota composition in the GI tract was analysed using a phylogenetic microarray technology that has been developed at the Laboratory of Microbiology of the Wageningen University. These DNA oligonucleotide microarrays target the V1 and V6 variable regions within the 16S rRNA gene sequences of the intestinal microbiota,

allowing the comprehensive profiling of intestinal microbiota composition (Rajilic-Stojanovic *et al.* 2009). In addition to the prototype, the HITChip (Human Intestinal Tract Chip), a platform specific for murine intestinal microbiota has been developed and validated (MITChip) (Geurts *et al.* 2011a, IJssennagger *et al.* 2012, Reikvam *et al.* 2012, El Aidy *et al.* 2013a, El Aidy *et al.* 2013b, Everard *et al.* 2013).

Information about:	Meta-genome	Meta-transcriptome	Meta-proteome	Meta-bolome
Taxonomy				
Functional capacity				
Expressed functions				
Metabolic output				

Table 1.1 Overview of information output from the different omics-technologies.

To address not only microbiota composition, but rather focus on the metabolic potential and actual activity of the intestinal microbiota, we also applied meta-transcriptomic and meta-bolomic analyses. These two methodologies belong to the 'meta-omic' approaches that have emerged during the last decade and are now widely used (Zoetendal *et al.* 2006, Qin *et al.* 2010, Van den Bogert 2011, Zoetendal *et al.* 2012, Fritz *et al.* 2013). Each of the 'meta-omic' approaches provides different information about the functional potential or activity profiles of a microbial community (Table 1.1). Metagenomics is used to determine the members present in a microbial community as well as their functional capacity. Metagenomics was used in the Meta HIT consortium and provided a human microbiome-derived gene catalogue with over 3 million genes, indicating a community of over 150 species in an individual and a 100-fold larger non-redundant gene set compared to the human genome (Qin *et al.* 2010). Metatranscriptomics and metaproteomics are able to provide information about the functions that are expressed by the members of the community. For instance, metaproteomics analysis in rats revealed differential protein patterns in microbiota samples derived from different intestinal locations, indicating aerobic microbial metabolism within the microbiota residing in the mucus layer and anaerobic microbial metabolism in the microbiota derived from the intestinal lumen (Haange *et al.* 2012). Metabolomic approaches are used to detect and quantify the metabolites that are produced by the microbial community. This approach has been suggested to be applicable as a diagnostic tool in diseases that involve aberrations of the intestinal microbiota composition and activity (De Preter and Verbeke 2013). In this

thesis only narrow spectrum metabolomics were used, focussing on the SCFA as the predominant end products of fibre and protein fermentation.

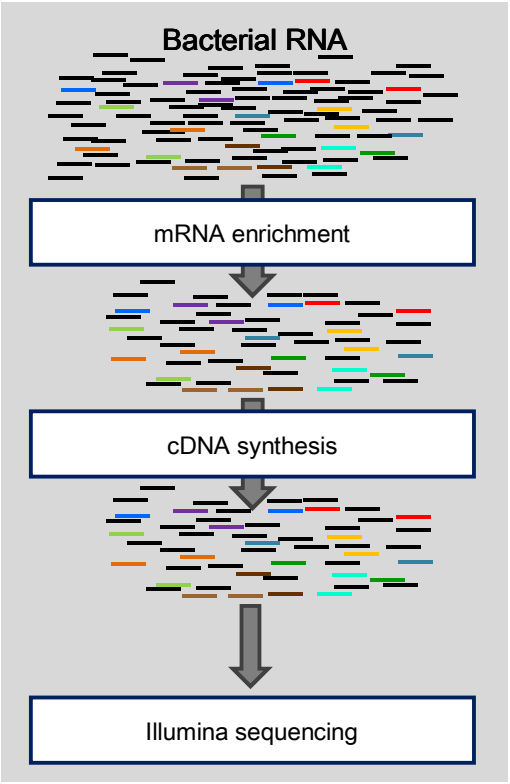


Figure 1.1 Preparation of RNA samples for meta-transcriptome sequencing. Bacterial RNA is enriched for messenger RNA (mRNA) and then synthesized into cDNA before the Illumina sequencing.

In this thesis metatranscriptomic approaches were used (**Chapter 6** and **7**). This method covers the analysis of the transcriptome – i.e. all the genes that are expressed by a microbial community such as the intestinal microbiota. For this analysis bacterial RNA needs to be isolated from environmental samples, i.e., intestinal content (Zoetendal *et al.* 2006, Zoetendal *et al.* 2012, Leimena *et al.* 2013). Notably, total bacterial RNA commonly consists for more than 95% of rRNA (and tRNA) and encompasses less than 5% of messenger RNA (mRNA), which can provide information about the transcription of protein encoding genes. To enlarge the relative fraction of mRNA for metatranscriptome sequencing, enrichment of mRNA can be performed by the selective removal of rRNA, using kits that are relying on oligonucleotide probes that are complementary to the highly conserved rRNA regions allowing post-hybridization removal of the probe-rRNA complexes from the sample. The mRNA enriched RNA sample is subsequently converted to double-strand cDNA, which can be subjected to next generation sequencing (Figure 1.1) (Leimena *et al.* 2013).

Metatranscriptome analysis in the GI tract microbiota enables the elucidation of the specific functional roles microbes have in this complex community. Although initial studies on the human large intestine revealed that different functions are expressed

among individuals, core functions of the microbiota appeared to be consistently expressed in different individuals (Gosalbes *et al.* 2011, Franzosa *et al.* 2014). Moreover metatranscriptome analyses of small intestinal microbiota underpinned the cross-feeding between two dominant members of the small intestinal microbiota, i.e., *Streptococcus* spp. and *Veillonella* spp., where the lactate produced by *Streptococcus* spp. is used as carbon and energy source by the *Veillonella* spp. (Zoetendal *et al.* 2012). Metatranscriptome analysis of the microbiota in humanized mice revealed that mice colonized with the microbiota obtained from a lean human donor displayed higher expression of genes involved in polysaccharide breakdown in propionate and butyrate production as compared to those colonized with the microbiota of an obese human donor (Ridaura *et al.* 2013). These findings imply that metatranscriptomics can provide insight in the differential activity profiles in the intestine microbiota, and enables the reconstruction of the metabolic activity profile of microbial communities.

Dataset integration approaches

The multivariate meta-omics datasets need tools to simplify the datasets and focus on correlations between points of interest, like dietary interventions to the bacterial community or the bacterial community to host responses. Multivariate statistics are used to handle these large datasets and enables quick focus on data of importance (Martin *et al.* 2008b, Martins dos Santos *et al.* 2010, ter Braak and P 2012). The CANOCO 5.0, a tool for multivariate analysis of ecological data, is using ordination methods to analyse communities. Ordination methods are assuming a continue change in the community composition and therefore 'order' species along a gradient (ter Braak 1987). Moreover CANOCO 5.0 enables visually summarising the community patterns with the external variables, such as dietary interventions and or SCFA concentrations. However to determine the effect of these external variables in the microbial community, constrained ordination methods, like Redundancy analysis, can be used to statistically test the impact of a variable on the overall variation in the dataset, which is used in **Chapter 4, 6** and **7**.

As mentioned earlier the microbiota also interacts with the host. The responses in the host are in this thesis measured in the mucosal cell layer. Intestinal tissues are scraped to obtain the mucosal cell layer, of which the RNA is extracted (IJssennagger et al. 2012). The RNA can be used to do real-time PCR or microarray analysis to find responses of the host to the microbes or diet. To address microbe-host interactions, the CRAN R 'mixOmics' library (Le Cao et al. 2009) can be used to first calculate the correlation of two heterogeneous datasets, like microbial composition and host mucosal transcriptome measured in the same samples, and then to visualize these correlations via heatmaps. Additionally the biological interpretation can be added allowing to generate hypotheses and models regarding the biology and corresponding mechanisms underlying interactions between the microbiota and the host.

The project, aims and outline

The interplay between the formation of SCFA, the composition and function of the microbiota, and the functioning of the host has not been evaluated *in vivo*, and so far only community metabolic networks of the microbiota in *in vitro* models have been generated (Kovatcheva-Datchary *et al.* 2009, den Besten *et al.* 2013). In a collaborative project between TI Food & Nutrition and the Netherlands Centre for Systems Biology (NCSB) we studied the SCFAs metabolism of the intestinal microbiota, and evaluated the potential of these microbiome metabolites to modulate the activity in the host's intestinal mucosa and liver, aiming to reconstruct a metabolic and host-microbe interaction framework using systems biology approaches, focussing on short chain fatty acids. The overall project encompassed 4 work packages (WP) (Figure 1.2):

1. Population dynamics of intestinal microbiota in relation to exogenous and endogenous factors
2. Quantification of metabolic activities of the gut microbiota
3. Molecular analysis of host responses, in intestinal mucosa and liver
4. Communication model construction for SCFA-host interaction, comparing the mouse and pig models with human

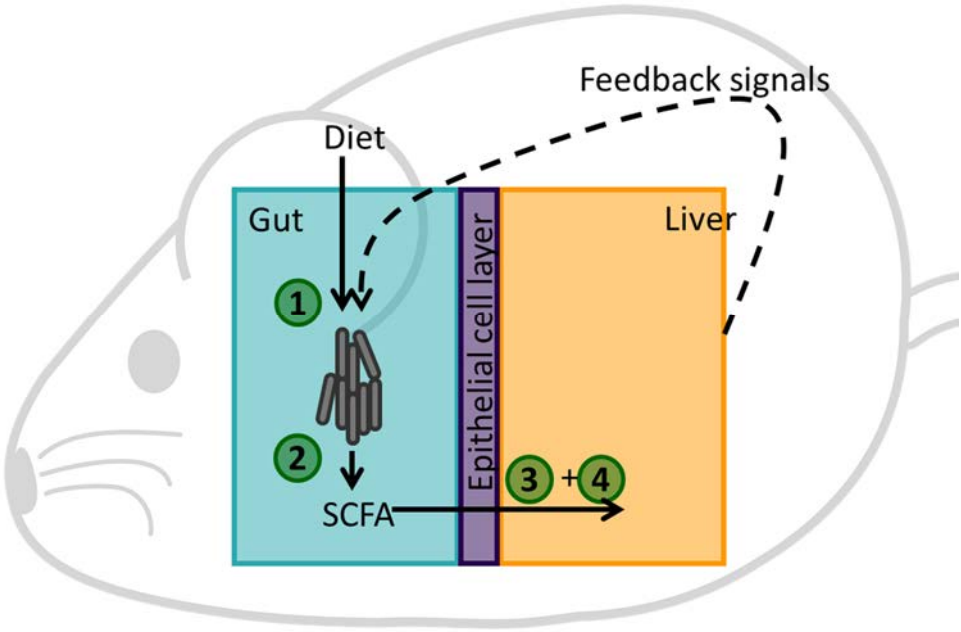


Figure 1.2 Work package overview of the TIFN-NCSB project.

This thesis predominantly encompasses work from WP1, and WP2, and also reports on collaborative efforts with WP3. The aim of the research described in this thesis is to unravel protein and fibre fermentation by the gut microbiota and the corresponding

host responses. Protein and fibre dietary interventions in mice were employed to decipher the metabolic profile changes they elicit in the murine gut microbiota, in correlation with SCFA production and the host responses in the mouse intestine.
In the review provided in **Chapter 2**, an overview is given on human gut microbiota and using in vitro and animal models to study the human gut microbiota. Focus is given to the study of the different regions along the GI tract and the use of rodent and pig models that enable more invasive studies to obtain samples that allow the analyses of the microbiota residing in the more inaccessible regions of the GI tract.
The modulation of the gut microbiota by fermentable carbohydrates is reviewed in **Chapter 3**. Such dietary modulations change the microbiota activity profiles, leading to altered SCFA concentrations in the gut. Focus is given to the opportunities to increase the in situ concentration of specific short chain fatty acids by targeting their microbial producers with specific carbohydrate nutrients.
The impact of five different dietary fibres on the mucosal transcriptome, the luminal microbiota composition and SCFA concentrations in the murine colon was studied in **Chapter 4**. Mucosal gene expression profiling revealed the regulation of fibre-specific, as well as overlapping transcriptional responses in colonic epithelial cells, which appeared to involve transcription factor PPARγ as a prominent upstream regulator of these transcriptional changes. Microbiota composition profiles could discriminate between the different dietary fibre interventions, although the interventions employing inulin, fructooligosaccharides, arabinoxylan and guar gum induced common changes in the abundance of several microbial groups. Multivariate data integration revealed strong correlations between the expression of genes involved in energy metabolism in the mucosa, with the relative abundance of bacteria belonging to the group of *Clostridium* cluster XIVa that includes several known butyrate producing bacteria.
To facilitate effective murine microbiota metatranscriptome RNAseq data processing and interpretation, a generic de novo assembly pipeline for such data was developed in **Chapter 5**. Notably, the pipeline developed is principally applicable for any (complex) microbial community, independent of its composition or the niche it was derived from. Functional-mapping of the caecal mouse microbiota metatranscriptome provided insight in global, and family-specific activity and underpins the potential of this approach to unravel interactions and task division in microbial ecosystems.
The microbiota metatranscriptome activity profiles in the caecum of the mice subjected to the dietary interventions using the different fibres described in **Chapter 4** were studied in **Chapter 6**. This chapter complements the microbiota composition analyses with metatranscriptome analysis of the transcriptional activity of the microbial community, as well as caecal mucosa transcriptome analysis in the host, and steady-state SCFA levels in the caecal lumen. The obtained datasets were independently analysed as well as integrated by multivariate statistical analyses. These analyses revealed distinct activity profiles of specific bacterial families that were associated to the fermentation of the dietary fibres provided in the mouse diet,

and their role in SCFA production. We could distinguish three categories of bacteria, (i) the *Bacteroidaceae*, the *Porphyromonadaceae* and the *Verrucomicrobiaceae* that expressed genes coding for glycosidases, but hardly sugar transporters; (ii) *Bifidobacteriaceae*, *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae* expressed both glycosidase- and sugar transporter-encoding genes; (iii) *Eubacteriaceae*, several *Bacilli* families and the *Ruminococcaceae* hardly expressed genes encoding glycosidases, but did express those coding for sugar transporters.
All these families expressed genes encoding enzymes involved in the production of short chain fatty acids, albeit in significantly different ratios. Furthermore, the relative abundance of bacterial groups involved in butyrate production correlated with transcriptional changes in the host's mucosa that were associated with processes such as energy metabolism, transcriptional regulation and immune system.
The long-term effects of high protein- and/or high-fat diets on the caecal microbiota composition and activity were studied in **Chapter 7**. Determinations of the microbiota composition using phylogenetic microarray (MITChip) technology were complemented with metatranscriptome analyses to unravel microbial activity profiles, and the steady-state caecal concentrations of (branched chain) SCFA. The metatranscriptome data revealed predominant nitrogen metabolism activity in the *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae* bacterial families, irrespective of the diet provided to the mice. The relative activity of *Erysipelotrichaceae* was increased in mice consuming the high protein diets, apparently at the expense of the relative activity of the *Lachnospiraceae* that was suppressed in the caeca of mice that were fed the high-protein diets. Notably, the relative activity of the *Clostridiaceae* appeared relatively insensitive to the dietary modulations, and this bacterial family expressed the broadest range of amino acid metabolism associated pathways.
The final chapter of this thesis, **Chapter 8**, provides a summary of the results presented in this thesis, including by a general discussion of the impact of these studies. This chapter also includes recent data, obtained in a follow-up experiment of **Chapter 4** and **6**, in which intestinal specific PPARγ knock-out mice were subjected to a dietary intervention using inulin. Chapter 8 concludes with a discussion and future directions of research in this field, aiming to expand our knowledge on interactions between the diet, microbiota and the host.

Chapter 2

Studying the mammalian intestinal microbiome using animal models

Floor Hugenholtz*, Jing Zhang*, Paul W. O'Toole and Hauke Smidt

*These authors contributed equally

Accepted for publication in *Manual of Environmental Microbiology*, 4th Edition, ASM Press.

Introduction

From birth onwards the gastrointestinal (GI) tract of humans and animals is colonized by microorganisms that constitute a community or ecosystem known as the microbiota. These microorganisms, predominantly from the bacterial kingdom, but also including archaea and eukaryotes such as fungi and protozoa, can reach a diversity of at least 160 species per individual, and over 1150 different species were detected in the human gut (Rajilic-Stojanovic *et al.* 2007, Qin *et al.* 2010, Faith *et al.* 2013). This complex ecosystem increases in numbers throughout the length of the GI tract, from 10 to 1000 cells per ml in the stomach, reaching a density of 10^{11} cells per gram of intestinal content (Booijink *et al.* 2007, Walter and Ley 2011) in the large intestine. All three domains of life are present in the large intestine where the bacterial community is dominant as well as the most phylogenetically diverse. At least nine different bacterial phyla have been detected in the large intestine, among which the phyla Bacteroidetes and Firmicutes dominate (Backhed *et al.* 2005, Rajilic-Stojanovic *et al.* 2007, Lozupone *et al.* 2012b). So far, only a minority of the bacteria in the gut have been cultured. Nonetheless molecular techniques that have emerged over the last two decades provided the opportunity to understand this complex GI tract ecosystem much better (Rajilic-Stojanovic *et al.* 2007, Van den Bogert 2011, van den Bogert *et al.* 2011, Fraher *et al.* 2012). The microbial ecosystem differs in the anatomically distinct regions in the intestinal tract, which has been reviewed elsewhere (Booijink *et al.* 2007).

The composition of the intestinal microbiota is driven by external factors such as habitual diet, antibiotic therapy and maternal microbiota, and intrinsic factors such as host species and genotype (Hoskins and Boulding 1976, Thompson-Chagoyán *et al.* 2007, Martin *et al.* 2008a, Claesson *et al.* 2012, Lozupone *et al.* 2012a, Makivuokko *et al.* 2012). Since the intestinal tract is the main point of contact of the host immune system and microorganisms, the role of microbiota in both local and systemic immune function plays an important role in immunity and health (Round and Mazmanian 2009).

Locations and their conditions along the GI tract

pH, transit time and microbial density are just a few of the many factors that are changing along the GI-tract. These differences need to be considered during experiments.

Oral cavity

When food is chewed in the mouth, it will be broken into pieces, moisturized and mixed with digestive enzymes – amylases and lipases – from the salivary glands of the host (Walter and Ley 2011). The most common bacteria found in the mouth are species of the genera *Gemella*, *Granalucatella*, *Streptococcus* and *Veillonella* (Aas *et al.* 2005, Walter and Ley 2011). Furthermore, additional niches exist in the oral cavity, such as supra- and subgingival plaque, which are densely populated by a

large number of different microorganisms, the diversity of which can be similar to that of the intestinal tract (Kolenbrander 2000, Smoot *et al.* 2005).

Stomach

The low pH of 1-2 in the stomach is too acidic for most microorganisms to survive. Until 30 years ago, it was considered to be a barrier for microorganisms, especially pathogens, to enter the body and survive (Savage 1977, Bik *et al.* 2006). However, in 1984, Barry Marshall and Robin Warren isolated a gastric bacterium, which was thought to be linked to gastritis. Later on this bacterium was named Helicobacter pylori and is now known to be present in 50% of human beings, whereas only a minority shows gastritis (Leser and Molbak 2009). Besides the understanding of H. pylori's survival and maintenance in the stomach, not much is known about other species that dwell in the stomach. Bik et al. (Bik *et al.* 2006) found 128 phylotypes from 8 bacterial phyla present in the stomach. This diversity was much higher than expected thus far. Moreover, 50% of the phylotypes were assigned to uncultivated bacteria, and of these 67% were described earlier as bacteria from the mouth. Nevertheless, it remains a question whether these bacteria dwell in the stomach and whether they have adapted to different environments – mouth and stomach. To collect gastric fluid or mucosal tissue, a nasogastric or orogastric catheter can be used, which enters via the nose or mouth, respectively, and passes the oesophagus to enter the stomach.

Small Intestine

The small intestine is considered the first region of the gastrointestinal tract where food meets microbiota. It can be subdivided into the duodenum, jejunum and ileum. This region of the GI tract is hard to access compared to the mouth and large intestine and therefore less well studied. The small intestine can be sampled (like the stomach) by using an intraluminal nasogastric or orogastric catheter, that passes the stomach and part of the small intestine depending on the region where the sampling will occur by peristalsis (van den Bogert *et al.* 2011, Zoetendal *et al.* 2012, van den Bogert *et al.* 2013). The location can be determined using short-interval fluoroscopic control and calculating the distance from the pylorus to the tip (Fraher *et al.* 2012, Zoetendal *et al.* 2012). This method provides an indication of which region the catheter is situated. However, due to considerable differences in the length of the small intestine in different individuals, the precise location cannot be determined. As an alternative, ileostomy subjects provide easier access to small intestinal content, and recently allowed detailed insight into structure and function of small intestinal bacterial communities (Booijink *et al.* 2010, van den Bogert *et al.* 2011, Zoetendal *et al.* 2012). These individuals have their colon removed, and the end of the ileum is surgically attached to an abdominal stoma. Despite not having a colon, ileostomized individuals can have a healthy life, and it could be shown by above-mentioned studies that they provide a suitable in vivo model system that enables analysis of the proximal small intestinal microbiota, rather than the ileum.
The diversity in the small intestine is higher than in the stomach, but smaller than

that found in the large intestine. The proximal small intestine is enriched with *Clostridium* spp., *Streptococcus* spp. and *Veillonella* spp. (van den Bogert *et al.* 2013). In turn, the ileum shows a community dominated by Bacteroidetes and *Clostridium* cluster XIVa and is more similar to the ecosystem of the large intestine (Booijink *et al.* 2010, van den Bogert *et al.* 2011, Zoetendal *et al.* 2012).
In addition to the possibilities outlined above, autonomous, ingestible intestinal sampling devices are being developed (e.g. http://www.micropharma.net), which would allow direct and programmed sampling of luminal as well as mucosal samples from predefined locations along the entire GI tract.

Large intestine
The large intestine can be subdivided into subparts: cecum, proximal, transversal and distal colon (Booijink *et al.* 2007). This region is densely populated by microbiota, the number of which can exceed 10^{11} cells per gram content. The diversity is large and reaches up to 160 bacterial species per individual (Qin *et al.* 2010), of which 90% belong to the Bacteroidetes and the Firmicutes (Ley *et al.* 2005).
The transit time of the intestinal content through the large intestine is much longer than in the other regions of the intestinal tract. Here the more complex food ingredients remain at the end of the GI tract as the sole energy source for the microbiota. Undigested carbohydrates and some fraction of proteins are converted into a broad range of metabolites, of which short chain fatty acids (SCFA), including acetate, propionate and butyrate, are the most abundant. In turn, these metabolites are used by the host as an energy source. This area of the GI tract is almost entirely anaerobic, and many bacteria that inhabit this part of the intestine are (obligate) anaerobic bacteria.
To study the large intestine, usually fresh faeces are collected and analysed. However, the microbial community of faeces is quite different from that residing in the proximal large intestine (Jeffery *et al.* 2012). This part still contains a lot of substrate for microbial growth, concentrations of which decrease towards the distal colon. Additionally, the obligate anaerobic species are much less prevalent in faeces than in the proximal large intestine (Jeffery *et al.* 2012). To obtain samples from the large intestine colonoscopy can be used. However, to actually enter with a colonscope into the colon, patients need to take sedatives and be sober in the last hours. More importantly, in case colonoscopy is performed with prior bowel cleansing, the obtained picture on the remaining microbiota will be affected, although it should be noted that it has recently been shown that colonoscopy doesn't have a lasting effect on faecal microbiota composition (O'Brien *et al.* 2013). Another option is to perform surgery on the patients in the large intestine; a more in depth review on sampling the large intestine is provided by Ouwehand & Vaughan (Kerckhoffs *et al.* 2006). Ingestible sampling devices such as those mentioned above might provide new possibilities also for undisturbed assessment of the proximal colon.

Models of the gut

As described in the previous section, studying the different locations along the GI tract currently requires rather invasive sampling methods. However invasive sampling from large numbers of healthy individuals is not feasible for practical and ethical reasons. In vitro and animal models provide an easier way to collect many (invasive) samples, have multiple comparisons and regulating the genotype background.

In vitro models

Solutions to the challenges explained above for *in vivo* studies can be the use of *in vitro* models, where a broad range of parameters can thus be measured during microbial fermentation. The *in vitro* models used to study the gut microbiota can be classified in batch fermentation models, continuous culture models and the TNO Intestinal Models (TIMs). The set-up and application of these different types of models has been extensively reviewed (Mäkivuokko and Nurminen 2006, Ridaura *et al.* 2013). *In vitro* models are usually inoculated with faecal samples. The faecal sample of an individual can be used in multiple comparisons at the same time, taking care that the different comparisons in the model are all originating from the same individual with the same genotype. By replicating particular conditions found in localized regions of the intestine, the gut microbiota in these models usually shifts to a microbial community more comparable to the corresponding intestinal region, validating to an extent the biological representativeness and value of the model.
In vitro fermentation models are mainly used to study the adaptation of the microbiota and the degradation of food or food ingredients (Kovatcheva-Datchary *et al.* 2009, Van den Abbeele *et al.* 2010). In the continuous culture and TIM models, probiotics are also tested by researchers to study their effects on the community and the washout time for these bacteria (Martinez *et al.* 2011, Martinez *et al.* 2013). Moreover by using membranes and filters, metabolites that are usually taken up by the host can be monitored during the fermentation process. However, mucus associated bacteria will not be present in these models. For this reason the M-SHIME was developed, where mucin-covered microcosms are introduced in the original SHIME model (Van den Abbeele *et al.* 2012b). Probiotics can now be better studied in the M-SHIME model with respect to their adhesion to the mucosal layer and their colonization.

Animal models

In general, *in vitro* models do not allow researchers to study interactions between the host and the microbiota. Animal models and in particular mammalian models provide an alternative way to study the *in vivo* responses to beneficial, commensal and pathogenic microorganisms in the GI tract. The main animals used to study the mammalian GI tract are rodents and pigs. Below we will discuss how they are used. However, to translate the knowledge gained from animal studies to the human

situation, differences in physiology (see Table 2.1, Figure 2.1 a, b and c) and behaviour between animals and humans concerning their GI tract need to be considered (Rodewald 1976, Stevens 1977, Booijink *et al.* 2007).

pH	Human	Rat	Pig
Stomach	1.0-4.4	3.3-5.1	2.2-4.3
Small intestine	5.5-7.5	6.5-7.1	6.0-7.5
Large intestine	5.9-7.0	6.6-7.4	6.3-7.1

Table 2.1 PH of the GI tract of humans, pigs and rodents (Rodewald 1976, Stevens 1977, Booijink et al. 2007).

Rodents

Conventional microbiota rodents

Rodents are often used to study the GI tract and relate this back to the human situation. These animals are relatively small, easy to keep and well known models to study drugs for humans. Like humans, the two main bacterial phyla of the rodent GI tract microbiota are the Bacteroidetes and the Firmicutes. Nevertheless, there are some differences in microbiota composition. For example, in mice bifidobacteria are present, but in lower abundance than in humans. Additionally, in mice, the *phylum Fusobacteria is lacking, while presence of Deferribacteres and Gemmatimonadetes has* been reported (Lozupone *et al.* 2012a). The mouse forestomach is colonized by a biofilm of *Lactobacillus reuteri*, and murine strains of this species are different to those found in humans that produce a vitamin B12 metabolosome (Frese *et al.* 2011). Furthermore mice harbour segmented filamentous bacteria, related to clostridia, that have a profound effect on the maturation of the innate immune system but have been thought to be lacking in humans (Suzuki *et al.* 2004, Gaboriau-Routhiau *et al.* 2009, Ivanov *et al.* 2009). It should be noted that distinct populations of SFB have now also been shown in human infants during the first three years of life, even though no functional studies have yet been performed that would support a similar role in immune maturation as for their murine counterparts (Yin *et al.* 2013).

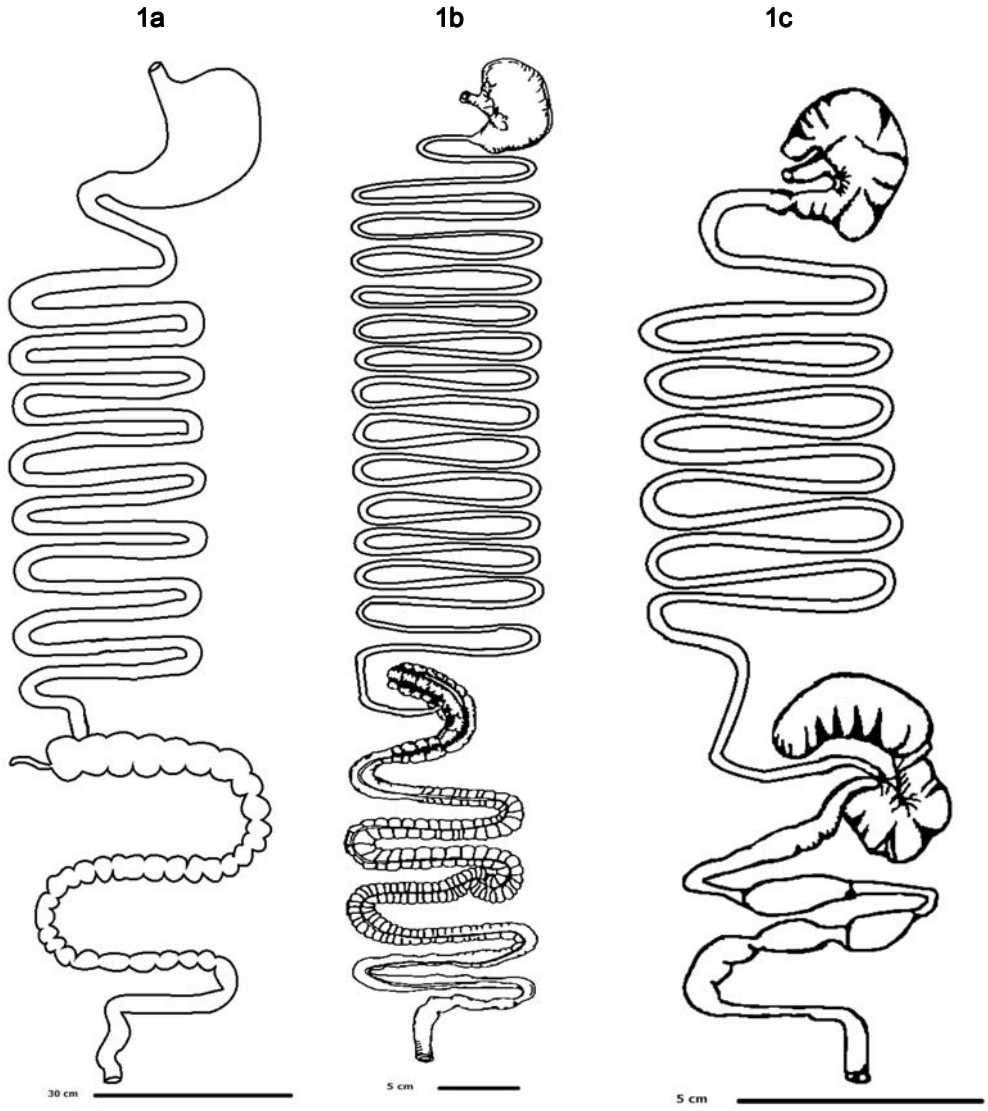


Figure 2.1 a) Gross anatomy of the human GI tract. b) Gross anatomy of the rat GI tract. c) Gross anatomy of the pig GI tract.

The microbiota composition in rodents is usually analysed when diet-microbiota-host interactions are studied. To this end, rodents are usually on or shifted to a specific diet. After a certain time, they are sacrificed in order to collect a range of different types of samples. For the purpose of microbiota analysis, in general two types of samples are collected from different locations along the GI tract, namely intestinal content as well as mucosal scrapings. The latter method allows recovery of the

epithelial cell layer from the intestinal tissue, and samples can be used to extract RNA for the analysis of host responses or to extract RNA or DNA to analyse the mucus associated microbiota.
There are a large number of different strains of mice and rats available. For example, C57Bl/6 mice are generally used in studies related to diet-induced obesity, type 2 diabetes and atherosclerosis (Ley *et al.* 2005, Geurts *et al.* 2011a). Turnbaugh and co-authors (Turnbaugh *et al.* 2008), for example, showed that gut microbiota of obese mice have a more efficient fermentation than lean mice. This fermentation resulted in a higher energy yield for the obese mice than lean mice with the same food. Moreover the obese gut microbiota, with the corresponding phenotypes, could be transferred to germ-free mice.
The choice of an animal model with a certain phenotype, is based on strains that have or are sensitive to this phenotype. For less obvious phenotypes, like response to a change of food ingredients, the choice is more likely to be made for practical reasons, including e.g. the animals are already bred in the facility, they are commercially available, or there is in-house experience with a given strain. However, recent studies have shown that environment and genetic background of mice have a significant impact on the microbial composition (Hildebrand *et al.* 2013), and this must be taken into account when experiments are designed that compare to or proceed from previous data.
Next to these factors, behavioural aspects like eating patterns and coprophagy of rodents need to be considered when setting up experiments, as well as during the experiment itself. Rodents are usually eating around the clock, and have therefore always food in their system along the GI tract. Next to that they practise coprophagy – eating their own faeces, or that of their cage partners – which allows them to extract more nutrients from the food (Sukemori *et al.* 2003). Also in terms of energy excretion coprophagy can have an impact, since the faeces that is eaten contains more water and nitrogen then normal faeces (Kenagy and Hoyt 1980). However, when studying the short-term effects of food intervention studies this can give a negative effect, unless at the start of intervention the faeces is removed.

Germ-free animals

Preferred animals for germ-free and gnotobiotic studies are mice. These animals are used for several purposes that include: to study the effect of colonization with one single bacterial species, a consortium of defined bacterial species or *de novo* colonization with a complex microbiota from animals with a specific genotypic and/or phenotypic background, or from other host species including humans. During conventionalization in mice it takes more than a week for the microbial community to stabilize, which needs to be taken into account in the experimental design (El Aidy *et al.* 2013a). Turnbaugh et al. (Turnbaugh *et al.* 2006) showed that transferring the microbiota of obese or lean mice into germ free mice resulted in a greater adiposity in those animals that received the faeces from obese mice. Another recent example of transferring a phenotype with its microbiota is the ability to transfer the production of testosterone from male mice to female mice. The female mice had higher

testosterone levels when they received via gavage male microbiota (Markle *et al.* 2013).

Humanized rodents

Rats and mice that are born germ-free can be colonized by a slurry of human faeces, often referred to as "humanization" (Mallett *et al.* 1987). Although the difference in physiology can have an effect on the colonization of the human microbiota, these humanized animals are valuable models to study the human microbiota, and can provide information on the interaction between food ingredients, the human microbiota and the host. The microbial shifts occurring in these models due to a certain treatment are likely to take place in humans as well (Van den Abbeele *et al.* 2011). For instance, the microbiota of human twins discordant for obesity was studied in mice, where the impact of the microbiota was linked with the phenotype of the human donor (Ridaura *et al.* 2013). They could furthermore show that specific dietary changes could affect the original phenotypes concomitant with alterations in composition and activity of the microbial community.

Pigs

Pig is an important livestock for human because of their meat production. Therefore, these animals are studied intensively by researchers with the object of production optimization. In recent years, many studies have focused on the GI tract of pigs, because manipulation of the gut microbiota can be used as an alternative of feeding antibiotic to improve the pig health. Additionally, pigs are scientifically important as a result of their high similarities to human beings in physiology, anatomy and nutrition (Douglas 1972, Miller and Ullrey 1987, Guilloteau *et al.* 2010). It makes these animals essential as models for human GI tract studies.

Humanized pigs

The model of human flora-associated piglets (HFAP) was established by Pang et al. (Pang *et al.* 2007) through orally inoculating a whole faecal suspension of a healthy 10-year-old boy to caesarean section derived piglets that were raised in specific-pathogen-free (SPF) conditions. Culture-independent analysis showed that transplantation of human gut microbiota produced a donor-like microbial community in piglet gut with minimal individual variation, and the succession with aging of piglets was similar to that observed in humans. As in humans, the introduction of solid food during weaning altered the gut microbial community, resulting in a decrease in bifidobacteria. This change suggested the HFAP may share similarities with human in the process of microbial colonization, and implied the HFAP could be an attractive model to explore the effect of dietary factors on human gut microbiota. Subsequently, the HFAP model was successfully employed in prebiotic study. In order to assess the effects of short-chain fructooligosaccharides (scFOS) on gut microbiota, Shen et al. (Shen *et al.* 2010) applied this model and confirmed the bifidogenic property of scFOS. They found that the *Bifidobacterium* genus was stimulated consistently except during weaning, however the effect of scFOS on non-bifidobacterial species varied at different developmental stages of the animals.

Gnotobiotic pigs

Gnotobiotic pigs have been used to study various human GI tract pathogens, such as Helicobacter pylori (Nedrud 2006). Recently, gnotobiotic pig has been used as an animal model to study the microbial colonization during early life. Laycock et al. (Laycock *et al.* 2012) used the Altered Schaedler Flora (ASF), a murine intestinal microbiota and a new "Bristol" microbiota containing *Lactobacillus amylovorus* DSM 16698T, *Clostridium glycolicum* and *Parabacteroides* spp (ASF519), to colonize caesarean-derived gnotobiotic pigs prior to their gut closure. The ASF inoculation resulted in unreliable colonization with most (but not all) strains of the ASF. In contrast, the Bristol microbiota reliably colonized the length of the intestinal tract of gnotobiotic piglets. This microbiota can be used to study the consequences of early microbial colonization on development of the intestinal mucosa and immune system, on later colonization by a complex microbiota, and on subsequent susceptibility to disease.

Minipigs

Minipigs are proposed to be good animal models for studying obesity (Johansen *et al.* 2001, Larsen *et al.* 2002, Larsen *et al.* 2005). Pedersen et al. (Pedersen *et al.* 2013) investigated the composition of gut microbiota in relation to diet, obesity and metabolic syndrome in two pig models, Göttingen and Ossabaw minipigs. They found that diet seems to be the defining factor that shapes the gut microbiota as observed by changes in different bacteria divisions between lean and obese minipigs. In the cecum, the lean Göttingen minipigs' had significantly higher abundance of Firmicutes, *Akkermansia*, and *Methanobrevibacter*, while obese Göttingen had higher abundances of the phyla Spirochaetes, Tenericutes, Verrucomicrobia and the genus *Bacteroides*. With respect to the Ossabaw minipigs, the obese minipigs had a higher abundance of Firmicutes in terminal ileum and lower abundance of Bacteroidetes in colon compared with lean minipigs. Overall, the Göttingen and Ossabaw minipigs displayed different microbial communities in response to diet-induced obesity in the different sections of their intestine. This funding also reinforced the notion that the host genotype has to be taken into account when studying the links between microbiota, diet and phenotype.

Piglet model for infant nutrition and development

The piglet has been used extensively in infant nutrition researches. It has been suggested as an appropriate model for human infant because of similarities between piglet and infant in anatomy, physiology and gastrointestinal tract metabolism (Miller and Ullrey 1987, Shulman *et al.* 1988, Darragh and Moughan 1995, Burrin *et al.* 2000, Puiman and Stoll 2008, Guilloteau *et al.* 2010). Moreover, piglet model has also been employed to evaluate the intestinal microbiota of neonates, and preterm and term infants. Development of the intestinal microbiota in neonates and infants is characterized by rapid and extensive changes in microbial abundance, diversity, and composition. These changes are influenced by medical, cultural, and environmental factors such as delivery mode, diet, familial environment, diseases, and therapies

(Matamoros *et al.* 2013). To study the effects of these factors, different piglet models have been developed. These piglet models allow us to generate more information of the dynamics of microbial colonization and its profound influence on intestinal and systemic health throughout life.

Piglet model for investigating effects of environmental microbial association on gut microbiota

To study the effects of environmental association with microbiota on gut health and development in the postnatal period, a model of caesarean derived piglets was designed by Jansman et al. (Jansman *et al.* 2012). In this model, piglets were obtained by caesarean delivery and were equally divided over two treatment groups that were housed in SPF conditions. All piglets received orally the above-mentioned Bristol microbiota consisting of *Lactobacillus amylovorus*, *Clostridium glycolicum*, and *Parabacteroides* spp. on days 1, 2, and 3 after birth. On day 3 and 4 the piglets received either a complex microbiota by providing them with a faecal inoculant of an adult sow (complex association group) or a placebo inoculant (simple association group). By using 16S rRNA gene targeted microarray-based microbiota profiling method, they found faecal microbiota composition was less diverse in the simple association group than the complex association group. The differences of microbiota between treatments persisted for at least three weeks after birth.

Furthermore, other studies reveal that the immediate environment during postnatal development has long-term impact on gut community structure in pigs (Thompson *et al.* 2008). To investigate the extent to which early-life environment impacts on microbial diversity of the adult gut, Mulder and colleagues established a model with genetically-related piglets, which were housed in either indoor or outdoor environments or in experimental isolators (Mulder *et al.* 2009). Analysis of over 3,000 16S rRNA sequences revealed major differences in mucosa-adherent microbial diversity in the ileum of adult pigs attributable to differences in early-life environment. Pigs housed in a natural outdoor environment showed a dominance of Firmicutes, in particular *Lactobacillus*, whereas pigs housed in a hygienic indoor environment had reduced *Lactobacillus* abundance and higher numbers of potentially pathogenic phylotypes. The result revealed a strong negative correlation between the abundance of Firmicutes and pathogenic bacterial populations in the gut, and the microbial composition differences were exaggerated in animals housed in experimental isolators. This study demonstrated strong influences of early-life environment on gut microbiota composition in adult pigs, leading to a follow-up study on the impact of limiting microbial exposure during early life on the development of the gut microbiota (Schmidt *et al.* 2011). In the following study, the outdoor- and indoor-reared piglets, exposed to the microbiota in their natural rearing environment for the first two days of life, were transferred to an isolator facility; and the gut microbial diversity of adult pigs was analysed by 16S rRNA gene sequencing. Although the initial maternal and environmental microbial inoculum of isolator-reared animals was identical to that of their naturally-reared littermates, the microbial succession and stabilization events reported previously in naturally-reared outdoor

animals did not occur. In contrast, the gut microbiota of isolator-reared animals remained highly diverse containing a large number of distinct phylotypes. These results indicated that establishment and maturation of the normal gut microbiota requires continuous microbial exposure during the early stages of life, and this process is compromised under conditions of excessive hygiene.

Piglet model for studying gut microbiota in diseases

For preterm neonates, one of the most serious diseases is the GI inflammatory disorder necrotizing enterocolitis (NEC). The development this disease includes the interplay of nutritional, microbial and immunological determinants. For independent studies of each determinant under clinically relevant conditions, the preterm piglet has been utilized in virtue of its beneficial characteristics compared to other animal models (Siggers *et al.* 2011). With preterm piglets, Sangild et al. (Sangild *et al.* 2006) found NEC pigs showed bacterial overgrowth and a high mucosal density of *C. perfringens* in some but not all animals, however mucosal microbial diversity of healthy pigs remained low and independent of diet. This finding was further confirmed by Cilieborg et al. (Cilieborg *et al.* 2011); their study showed a different microbiota with high *C. perfringens* abundance was observed in preterm pigs with NEC compared with healthy individuals. However, the *C. perfringens* inoculation failed to induce NEC. It indicated *C. perfringens* is more abundant in pigs with NEC but rather as a consequence than a cause of disease. In addition to the above studies, Azcarate-Peril et al. (Azcarate-Peril *et al.* 2011) have used a unique preterm piglet model to characterize spontaneous differences in microbiome composition of NEC-predisposed regions of gut. Their study provided strong support for ileal mucosa as a focus for investigation of specific dysbiosis associated with NEC and suggested a significant role for *Clostridium* spp., and members of the Actinobacteria and Cyanobacteria in the pathogenesis of NEC.

The short bowel syndrome (SBS) piglet model is another application of the piglet model for studying gut microbiota in diseases. The development of a successful SBS model in neonatal piglets provides a possibility for characterizing the colonic microbiota following small bowel resection (SBR). By using 4-week old female piglets that received a 75% SBR, Lapthorne et al. (Lapthorne *et al.* 2013) found a significant level of dysbiosis both two and six weeks post-SBR, particularly in the phylum Firmicutes, coupled with a decrease in overall bacterial diversity in the colon.

Sampling techniques with pig models in kinetic microbiota studies

Small intestinal segment perfusion technique with pig models

The Small intestinal segment perfusion (SISP) technique was developed to study the effects of bacteria on net absorption of fluid and electrolytes, as a more comprehensive and ethical alternative to the ligated loop test in pig models (Nabuurs *et al.* 1993). Recently, this technique has been widely applied to study enterotoxigenic *Escherichia coli* infection (Niewold *et al.* 2005, Niewold *et al.* 2010), *Salmonella typhimurium* invasion (Niewold *et al.* 2007, Veldhuizen *et al.* 2007), and mannose-specific interaction of *Lactobacillus plantarum* with jejunal epithelium

(Gross *et al.* 2008). Furthermore, SISP can be applied in future research to investigate the functional physiological response of probiotics and the crosstalk between probiotics and the host (Van der Meulen *et al.* 2010). In the SISP test, pigs are sedated with azaperone, induced with inhalation anaesthesia and maintained with sevoflurane and nitrous oxide. For pig small intestine, five pairs of segments are prepared and each segment is 20 cm long with inlet tubes at the cranial side and outlet tubes at the caudal side. The segments can be used to study up to 10 perfused ingredients within one pig. All 10 segments are perfused simultaneously either by an infusion system or manually with syringes attached to the cranial tubes for up to 10 hours. Effluent fluid during perfusion and mucosal scraping can be sampled for microbial analysis from each segment.

Cannulation technique with pig models

Cannulation is one of the most frequently applied methods for repeated sampling of digesta from pig gut. Different cannulation techniques, such as simple T-cannula, post-valvular T-cecum cannulation and steered ileocecal valve cannulation, can be employed based on the specific research purpose. Among these cannulation techniques, the simple T-cannula is widely used for evaluating the microbial composition and function in the ileum. A simple T- cannula is normally inserted 10 to 20 cm anterior to the ileocecal valve. It does not transect the small intestine wall, and can maintain a normal physiological state of the intestine (Sauer and De Lange 1992). Currently, surgical procedures for inserting a T-cannula and sampling methods have been established for young pigs (Walker *et al.* 1986, Li *et al.* 1993), growing pigs (Decuypere *et al.* 1977, Gargallo and Zimmerman 1980) and pregnant sows (Stein *et al.* 1998).

Conclusion

Studying the human GI tract can be done with a wide range of methods and technical approaches. Even though each of the methods that we described here have advantages and disadvantages, usually human faeces are used for assessing the impact on the microbiota in intervention studies, whereas animal models are used for more detailed mechanistic studies, including those that aim to see the impact of the host system as well. Nowadays mainly rodents are used to study the human GI tract, while pigs show promising and maybe better comparison to study the human GI tract. Since the optimal system is not yet established, generally to test a certain treatment a combination of methods is used, first in vitro, then in vivo in an animal model, ending up with a human intervention study.

Acknowledgments

Jing Zhang was supported by a fellowship of the China Scholarship Programme, and Jing Zhang and Hauke Smidt were supported by the European Union through the Interplay project (Grant agreement no. 227549) and Cargill's R&D Center Europe (Vilvoorde, Belgium). Cargill had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Research in PWOT's laboratory is funded by Science Foundation Ireland, Dept. Agriculture Food and Marine, and the Health Research Board. Floor Hugenholtz is funded by the Netherlands Consortium for Systems Biology (NCSB) which is part of the Netherlands Genomics Initiative and the Netherlands Organisation for Scientific

Chapter 3

Modulation of the microbial fermentation in the gut by fermentable carbohydrates.

Floor Hugenholtz, Jane Adair Mullaney, Michiel Kleerebezem, Hauke Smidt and Douglas Ian Rosendale

Bioactive Carbohydrates and Dietary Fibre , 2013. Vol 2, issue 2, pages 133-142

Abstract

This review considers fermentable carbohydrates and their role in maintaining health through their availability as fuel for the gut microbiota. The microbiota possesses remarkably diverse function, and is likely modifiable by diet. Therefore a diet rich in varied fermentable carbohydrates such as dietary fibre, glycosylated polyphenolics, glucosinolates and other plant glycans, applied in a sustained fashion may promote microbial diversity leading to improved health. This may be achieved by increasing the flexibility of the microbiota's capability to interact with diverse dietary environments, or via increasing production of short chain fatty acids (SCFAs) from the fermentation of carbohydrates. A higher functional modular complexity is indicative of gut health, whilst SCFAs may reduce the risk of developing gastrointestinal disorders, cancer, and cardiovascular disease.

Keywords: microbiota; fermentable carbohydrates; short chain fatty acids; health

General Introduction

The physicochemical effects of dietary fibre consumption are well recognised by the scientific community, regulatory authorities, food manufacturers and consumers. Generally these “feel the benefit” attributes comprise improvements in laxation: frequency, bulk and transit time. However, there are a host of more subtle health benefits conferred by the gut microbiota which are becoming increasingly recognised. Here we focus on these microbial benefits as conferred by the usage of fermentable carbohydrates. In this review, we consider fermentable carbohydrates to be any carbohydrate source which, for whatever reason, escapes digestion by the host, and passes into the large bowel intact, where it may act as a substrate for the growth and metabolic output of the resident bacteria. These fermentable carbohydrates may not necessarily fall within the current definition of dietary fibre, or prebiotic compounds, but some may loosely fall into that category. Thus we consider not only material which meets Codex definition of dietary fibre, but also other glycosylated compounds which consist of one or more sugar residues attached via glycosidic linkage to non-carbohydrate compounds which collectively do not meet conventional definitions of “carbohydrates” in the strictest sense, but nevertheless are available as microbial fuel to result in the microbial generation of outcomes of interest.

This gut microbiota is diverse, highly abundant, competitive, metabolically active, and largely saccharolytic (Moore and Holdeman 1974, Finegold *et al.* 1983, Qin *et al.* 2010, Tasse *et al.* 2010, Flint *et al.* 2012b). Most members of the microbiota are not reliant on the availability of simple sugars, but are able to derive carbon and energy from the breakdown of sometimes very complex carbohydrates, alone or as a concerted effort. Substrates influencing and fuelling this microbiota include both food which escapes host digestion in the upper intestinal tract, but also endogenous host secretions. For example, it is unlikely to be a coincidence that the sites of highest microbial abundance are the sites of highest mucus abundance – where the main structural component of the mucus is the heavily glycosylated glycoprotein mucin, acting as a barrier to protect the underlying epithelia from damaging agents (including the microbiota) and as a substrate to allow the continued persistence of the microbiota in the absence of any other food source. Also present are other oligosaccharides, peptides, glycosaminoglycans, aliphatic lipids and steroids (Hoskins and Boulding 1976). Products of fermentation of these substrates by the microbiota include short chain fatty acids (SCFA), branched chain fatty acids (BCFA), and a range of other metabolites, such as vitamins, nitrogenous compounds, deconjugated exogenous (phytochemicals) or endogenous (bile) compounds, and others. Many of these metabolic by-products are in turn used by other species (the secondary feeders) such that food webs and food chains develop. Indeed, the stable persistence of a diverse and cooperative microbiota renders it unlikely that any single nutrient is limiting the growth of all the resident microorganisms. Nevertheless, despite the availability of diverse and complex

substrates, it is dietary carbohydrate sources which we may deliberately vary by exercising dietary choice, and thus use to manipulate the microbiota.

The microbiota and health

The diverse, abundant and competitive microbiota largely occupies the distal regions of the intestinal tract. These microorganisms become established in succession from birth onwards (Maynard *et al.* 2012), concurrently driving the maturation of the intestinal tract as a functioning digestive, neuroendocrine and immune organ. The makeup of this microbial consortium is driven by external factors such as food, antibiotic therapy and maternal microbiota (Thompson-Chagoyán *et al.* 2007), and intrinsic factors such as host species (Rawls *et al.* 2006, Martin *et al.* 2008a) and genotype (Hoskins and Boulding 1976, Makivuokko *et al.* 2012). Since the intestinal tract is the main point of contact of the host immune system and microorganisms (Round and Mazmanian 2009), the microbiota in both local and systemic immune function (and dysfunction) play an important role in immunity and health. Immune dysfunction links with metabolic and autoimmune disorders and so deliberately modulating the microbiota with fermentable carbohydrate-based food might permit modulation of systemic immunity and obesity.

Ultimately, there may be in excess of 1000 different species found amongst individual adult humans, with any given individual possessing >100 of these (Qin *et al.* 2010). This complex ecosystem increases in numbers throughout the intestinal tract, from 10 to 1000 cells per mL in the stomach until a density of 10^{11} cells per gram of intestinal contents in the large intestine (Booijink *et al.* 2007, Walter and Ley 2011). These form a complex mixture of resident species, where a variable number and proportion of transients are unable to compete with or displace resident (commensal) organisms in the synergistic associations and food chains which contribute to determining the composition and stability of the microbiota.

So far only 20-46% of the bacteria in the gut have been cultivated. Nonetheless a range of omics-approaches – metagenomics, metatranscriptomics, metaproteomics, metabolomics and fluxomics - of the last few decades have given the opportunity to understand this complex ecosystem through the GI tract much better. For example metagenomic approaches have been used to assess the population and functional diversity of the microbiota, while metabolomic approaches have been used to assess the impact of introducing poorly adapted microbiota across species, which have increased our understanding of the systemic role of this ecosystem (Martin *et al.* 2007, Rajilic-Stojanovic *et al.* 2007, Martin *et al.* 2008a, Martin *et al.* 2009, van den Bogert *et al.* 2011).

Most of the members of the resident gut microbiota can be classified in four phyla: Bacteroidetes, Firmicutes, Actinobacteria and Proteobacteria, although the increasing importance of less abundant phyla such as Verrucomicrobia, or kingdoms such as Archaea, is becoming increasingly recognised (Everard *et al.* 2013). Similarly, yeast, fungi, protozoa and viruses are also present (colonisation or replication without causing disease may define them as commensal) and exert

influence. Indeed, the role of bacteriophage in genetically conferring or stabilising functions within microbiome is also becoming recognised (Reyes *et al.* 2010). Yet it is the key bacterial phyla, particularly the Bacteroidetes and Firmicutes, which are numerically and arguably functionally dominant: associations have been made between phyla ratios and functional or differences in the microbiota, or between relative phyla abundance and host physiology. Interestingly, the carbohydrate degradation machinery of members of these two phyla appears to be polar opposites: the extra- cell-associated machinery of the Bacteroidetes phyla *vs* the extracellular machinery possessed by members of the Firmicutes phyla (Muñoz-Tamayo *et al.* 2011). The latter machinery has been proposed as key to degrading recalcitrant carbohydrates (celluloses and hemi-celluloses). Similarly, the SCFA profiles produced by members of these two dominant phyla differ, with a tendency for butyrate production by members of the Firmicutes phyla, whilst propionate production tends to be dominated by Bacteroidetes. The roles of these and other SCFA in gut and systemic health will be explored in more detail later.

Overall, the microbiota possesses remarkably consistent function across individuals (Qin *et al.* 2010), albeit dependent on gross dietary differences across species (Muegge *et al.* 2011). Recent work employing ecological mathematical principles has shown that the microbiota across populations can be divided on bases of metagenomic complement into different functional modules (how the genes within networks are grouped according to function) of varying complexity (Greenblum *et al.* 2012). Here it appears that the microbiota of lean healthy individuals has a higher functional modular complexity than that of obese or irritable bowel disease (IBD) individuals. Essentially this is simplistically represented as the genetic pathways on the periphery of metabolic networks, notably those featuring the first substrates seen by the microbiota, and the last products produced, are mathematically networked differently. Those of lean healthy individuals have higher numbers of functional modules (increased complexity) each containing less genes/networks, whilst obese or IBD individuals possessed lower numbers of functional modules (decreased complexity) each containing higher numbers of genes/networks. Core metabolic function, shared amongst all members of the microbiota in all individuals (e.g. nucleotide synthesis, cell division etc.), appears not to vary in modularity (Greenblum *et al.* 2012). This variation in complexity of these “peripheral” genes in the network relates to the functional diversity of these microbiota, and the situation is analogous to other systems for which these principles have been applied, e.g. obligate symbionts have very low functional complexity coinciding with adaptation to low diversity environments (Parter *et al.* 2007). This ecological diversity is relevant here for the gut microbiota because those substrate- and product-interacting genes on the periphery of the networks are likely to involve carbohydrates degradation and some SCFA production.

Fermentable carbohydrates

Fermentable carbohydrates are capable of causing favourable changes to the microbiota (van Zanten *et al.* 2012, Haenen *et al.* 2013d). A commonly accepted term to describe this process is “prebiosis”, which is the fermentation of prebiotics. The definition of prebiotics is “non-digestible food ingredients that beneficially affect the host by selectively stimulating the growth and/or activity of one of a limited number of bacteria in the colon” (Gibson and Roberfroid 1995). The term prebiotics and dietary fibre (see below) are sometimes used interchangeably; however they are not the same. Prebiotics stimulate specific bacteria in the colon, while dietary fibres can be fermented by a range of bacteria or not fermented at all (Ouwehand *et al.* 2005). Conversely, a diet low in fermentable carbohydrate (e.g. resistant starch), common amongst westerners, is associated with colonic disorders (Scheppach 1994).

Dietary Fibre

Dietary fibre is, by definition, dietary polysaccharides and oligosaccharides that resist digestion by the human digestive enzymes. It includes the non-starch polysaccharides portion of foods derived from plant cell walls (pectin, hemicelluloses, and cellulose), oligosaccharides such as fructooligosaccharides derived from inulin, digestion-resistant starches, and a range of other non-digestible polysaccharides and oligosaccharides added to food formulations to increase their fibre content. Dietary fibre has been classified into soluble and insoluble fibre. Most, but not all soluble fibres from viscous solution are fermented in the colon. Insoluble fibres are also fermented, but include some, such as cellulose, that are fermented slowly enough to largely survive colonic transit and have a bulking action in the colon. In some cases this is preferred as rapidly fermented fibre can result in uncomfortable physiological effects. Within the human gut microbial metagenome data a wide variety of carbohydrate-degrading enzyme families can be found (Tasse *et al.* 2010). These enzymes are enriched in adults compared to infants, emphasizing the shift to richer mixture of carbohydrates in the gut. This variety of enzymes is necessary to degrade the complex structures present in dietary fibre. The different linkages, with the combination of different mono-sugars, in the fibres require an arsenal of different carbohydrate degrading enzymes (reviewed in (Flint *et al.* 2012a)). Some bacteria, like the *Bacteroides* spp., are well equipped with a range of glycoside hydrolases and are capable of switching between different substrates (Hooper *et al.* 2002). However, these species are more equipped to degrade soluble carbohydrates (Flint and Bayer 2008). In contrast, within the family of *Bifidobacteriaceae* there are some species which are specialised to utilise only certain groups of oligosaccharides (Barboza *et al.* 2009, Riviere *et al.* 2014). In practice, bacterial usage of substrates is also influenced by their ability to adhere to the food matrix within the gut. For example a lawn of *Bacteroides* spp. attached to a food particle and interspersed with islands of *Bifidobacterium* spp. has been observed (Macfarlane and Macfarlane 2006).

Host-derived fermentable carbohydrates

Host carbohydrates (predominantly from the heavily glycosylated mucin glycoproteins that are the main structural component of the mucus layer lining the gastrointestinal tract) were thought sufficient to maintain the large bowel microbiota in its original abundance and diversity in the absence of dietary carbohydrate (Winitz *et al.* 1970, Attebery *et al.* 1972, Bounous and Devroede 1974, Hudson *et al.* 1981, Macfarlane *et al.* 1989), Now, contemporary sequencing methods may reveal community differences previously unobservable through historical microbiological techniques (Gerald Tannock, University of Otago, NZ, pers. comm.). Microbial ecological impact notwithstanding, mucin oligosaccharide forms a major alternate fermentative substrate to the microbiota during a dearth of dietary carbohydrate. A consequence of mucin oligosaccharide utilisation within the mucin layer is that the highly de-glycosylated mucin is rendered less resistant to degradation (Variyam and Hoskins 1983), thus allowing breakdown of the protein scaffold and access to the underlying epithelia. Both dietary and host carbohydrate sources are ultimately catabolised to result in increased microbial biomass and production of the microbial metabolic by-products dominated by SCFAs and, with fermentable protein, BCFAs (Louis *et al.* 2007).

Furthermore, the introduction of fermentable dietary carbohydrate to this system results in the redistribution of some of the collective microbial degradative capability away from host carbohydrates towards this additional nutrient resource (Sonnenburg *et al.* 2005), while a fibre-induced decreased transit time combined with increased secretion of mucin ultimately results in faster clearance of the existing microbiota (Tirosh and Rubinstein 1998), and the replenishment of highly sulphated and sialylated mucin sugars (Larsen *et al.* 1993). *Bacteroides thetaiotaomicron* has been characterised in terms of its food and host interactions *in vivo* (Martens *et al.* 2009). In particular, the implications of this organism's complete switch from host-derived to food-derived carbohydrate degradation upon supply of food carbohydrate, illustrates how exposing the collective gut microbiota to carbohydrates could be applied to modify gut health.

Glycosylated bioactives

Many plant compounds are glycosylated. They tend to attract scientific attention upon loss of the sugar group resulting in their aglycone moieties, many of which possess bioactivity due in part to bioavailability: their intrinsic lipophilic properties allowing uptake by cells. Examples are glucosinolates and polyphenols.

Glucosinolates are -D-thioglucoside-(Z)-N-hydroxyiminosulfate compounds that contain an amino acid-derived side chain (Figure 3.1). Glucosinolates such as glucoraphanin from cruciferous vegetables are hydrophilic secondary plant metabolites believed to confer the plant with defence against predation. Glucosinolates themselves appear not to be bioactive, however, upon removal of the glycoside moiety, the resulting aglycones are bitter tasting, lipophilic, host phase II drug metabolism and antioxidant pathways inducing compounds, such as

sulforaphane (Brooks *et al.* 2001).
Polyphenols encompass a broad class of compounds (Manach *et al.* 2004) undergoing extensive modification during digestion where, like glucosinolates, are rendered bioactive. In contrast to glucosinolates which become aglycones, polyphenols are generally found as conjugates of glucuronate or sulfate, with or without methylation of the catechol functional group and consequently have different biological effects from polyphenol aglycones such as those found in green tea catechins (Kroon *et al.* 2004).
Here we are specifically interested in the apparently non-bioactive, glycosylated, hydrophilic form, as they are potential targets for bacterial glycosidases, and perhaps substrates for gut microbial growth.

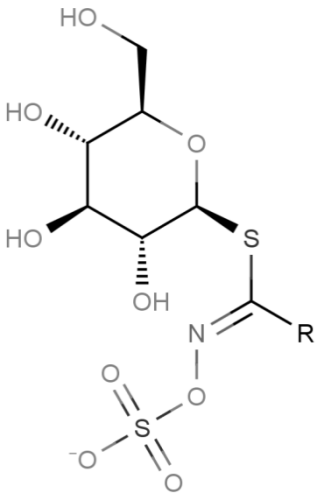


Figure 3.1 General structure for all glucosinolates, the dashed line showing the site where the glucose molecule is cleaved during hydrolysis (Mullaney et al. 2013).

Short chain fatty acids

The products produced by a microbiota include bacterial fatty acids, de-conjugated bile acids, protein putrefaction products and even B vitamins (Stevens and Hume 1998). In the simplest of terms, (SCFAs) are produced by the gut microbiota through their fermentation of carbohydrates. These are metabolised by the microbiota mainly via the glycolytic pathway for hexoses and via the pentose phosphate pathway for pentoses resulting in pyruvate, the main precursor for SCFA (Cummings 1981, Macfarlane and Macfarlane 2003) (Figure 3.2). Anaerobic fermentation in the gut is determined by redox differences between substrates and products (Macfarlane and Macfarlane 2003). This state determines which products can be formed and thus the amount of energy that can be formed. Some of the products, like lactate and butyrate, are also used to get rid of the excess of electrons (Miller and Wolin 1979). Short chain fatty acids (SCFA) are considered to be beneficial fermentation products

in the gut, playing an essential role in the maintenance of colonic integrity and metabolism (Cook and Sellin 1998). SCFAs also exert many other beneficial effects on the host including resistance to disease (Topping and Clifton 2001), have a role in blood pressure regulation (Pluznick *et al.* 2013), and may be protective against cancers by increasing cell proliferation and apoptosis (Scharlau *et al.* 2009). SCFAs act as energy sources (brain, heart, muscle); increase bile salt solubility, mineral absorption, leptin production, leptin regulation which helps to protect against obesity and metabolic disorders (Lin *et al.* 2012); decrease gut pH, ammonia absorption, and inhibit pathogen growth.

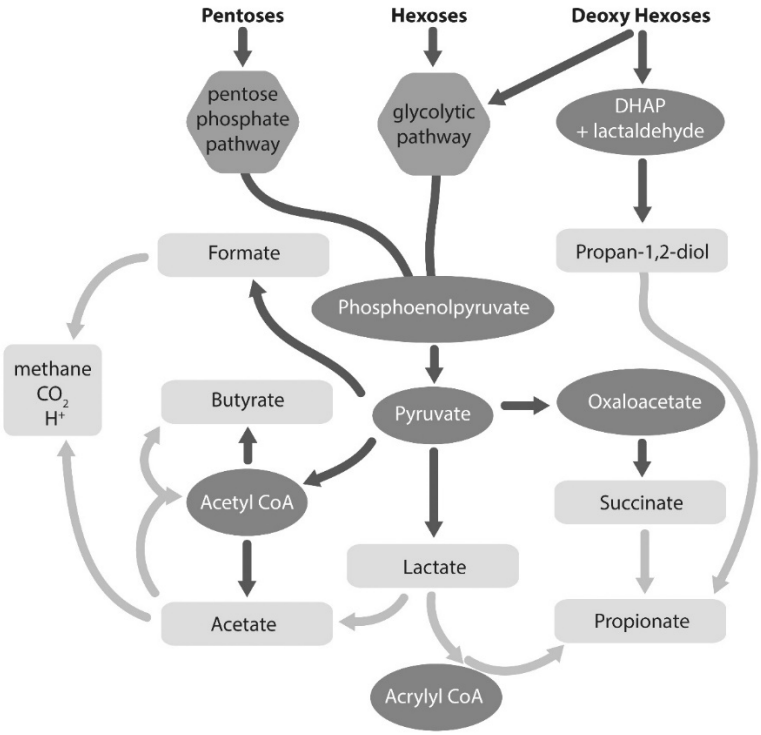


Figure 3.2 Anaerobic sugar fermentation to short chain fatty acids. Metabolites with rounded boxes are intracellular; metabolites with rectangular boxes may be excreted outside of the cell. Black arrows show intracellular pathways; grey arrows indicate where a bacterium may capture and process an excreted metabolite (cross-feeding).

Specific SCFAs may reduce the risk of developing gastrointestinal disorders, cancer, and cardiovascular disease. The major three SCFAs are butyrate, acetate and propionate. Butyrate is the main energy source for colonocytes, propionate is taken up and metabolized by the liver, and acetate is taken up via peripheral circulation for metabolism by peripheral tissues (Wong *et al.* 2006). Acetate, the most highly concentrated SCFA in the colon, has been shown (after absorption) to increase

cholesterol synthesis, while propionate has been shown to inhibit cholesterol synthesis. Butyrate irrigation (enema) was suggested in the treatment of colitis (Scheppach *et al.* 1992), but determined to be ineffective later by the same group (Scheppach *et al.* 1997). Other organic acids (not short chain fatty acids, strictly speaking) include succinate and lactate. The collective functions of the major SCFA and organic acids are summarised in Table 3.1.

Lactate	Acetate	Propionate	Butyrate	Succinate	Biological change
	↑	↑			Energy source (brain, heart, muscle)
			↑		Main energy source (colonocytes)
↑	↑	↑	↑	↑	Decreased gut pH (increased bile salt solubility, increased mineral absorption, decreased ammonia absorption, decreased pathogen growth)
			↑		Anti-cancer (inhibit proliferation, induce apoptosis)
	↑				Lipid metabolism (de novo lipogenesis substrate)
		↑			Lipid metabolism (inhibit HMG-CoA synthase and reductase)
		↑			Increased leptin production
		↑			Decreased acetate absorption and fatty acid synthesis resulting in decreased hepatic lipogenesis
			↑		Decreased inflammation
	↑		↑		Increase associated with obesity in mice

Table 3.1 Summary of biological changes associated with increased SCFA or other organic acids (Fava et al. 2008, Rosendale et al. 2011)

Changes in SCFA production in response to fermentable carbohydrate

The amount and ratio of SCFA can be altered by specific types of fibre (Flint and Bayer 2008, Van den Abbeele *et al.* 2011). Effects of fibres do vary between species (Ferguson *et al.* 2000, McOrist *et al.* 2011). In human faecal samples resistant starch has been seen to specifically increase butyrate (McOrist *et al.* 2011), whereas in a rodent model, different patterns for caecal SCFA level were observed, dependent on the RS type used, while this was independent for colonic concentrations (Ferguson *et al.* 2000). In a simulated human intestinal microbiota population model arabinoxylan has been seen to increase in particular propionate (Grootaert *et al.* 2009). Further changes have been observed with a number of other fibre types (Table 3.2).

In addition, glucosinolates and glycosylated polyphenolics have wrought changes in SCFA profiles. For example, fermentation by-product profiles in the caecum of rats

where the microbiota was primed with supplements of food grade bacteria capable of deglycosylating glucosinolates and further acclimatised to a glucosinolate-supplemented diet had quite different profiles from rats fed basal, un-supplemented diets (Mullaney 2013). Similarly, the SCFA profile and microbial abundance of *in vitro* fermentations in the presence of glycosylated phenolic compounds differed from un-supplemented cultures.

In addition to different carbon sources leading to different SCFA profiles we know that there is significant SCFA profile difference between strains of the same species. However in the context of whole microbiota studies where the microbial information only differentiate at the genus or family level (or above), we frequently cannot consider SCFA production at the strain level. Nevertheless we can attempt to capitalize on known and commonly occurring trends during fermentation by a complex culture.

Cross feeding

Indirectly fibre fermentation alters bacteria that do not ferment dietary fibres, but are using the acetate and lactate produced by others in the gut. These so called 'Cross feeders' are organisms that cannot break down large polymers by themselves but take advantage of the products of other organisms: these products may be polysaccharide fragments, or SCFA resulting from fermentation by the other organisms. This is illustrated by an *in vitro* study where incorporation of a heavy [^{13}C] isotope label from starch into microbial RNA was measured revealed that *Ruminococcus* spp. were the primary starch degraders as indicated by their predominant label incorporation, whilst *Prevotella*, *Eubacterium* and *Bifidobacterium* spp. incorporated lesser amounts of ^{13}C, consistent with a secondary feeding position or crossfeeding upon fermentation by-products of the *Ruminococcus* primary feeders (Kovatcheva-Datchary *et al.* 2009).

Similarly, other studies show different phylogenetic groups are capable of converting lactate or acetate and lactate to butyrate or propionate (Duncan *et al.* 2004, Zoetendal *et al.* 2012) (Figure 3.2). For instance, in the small intestine *Streptococcus* spp. convert simple sugars into lactate (Booijink *et al.* 2007, Zoetendal *et al.* 2012). The lactate can be used by *Veillonella* spp. as a carbon source and converted into propionate and acetate. However in *Clostridium perfringens* and possibly *Bifidobacterium breve* the amount of lactate produced can be dependent on the availability of glucose (Macfarlane and Macfarlane 2003). When there is a surplus of glucose *C. perfringens* produces mainly lactate, since the lactate then functions as an electron donor. If there is a shortage of glucose, *C. perfringens* switches to a high acetate production, where more ATP is formed per glucose molecule.

Dietary fibre	Duration (weeks)	Mammal	SCFA		Molecular profiling methods (16S rRNA)	Bacterial changes detected		Reference
			Overall	*specific*		*up*	*down*	
Resistant starch (Hi Maize)	4	human	up	acetate and butyrate	DGGE; qPCR	*Ruminococcus bromii*		Abell et al. (2008)
Inulin and oligofructose	2.3	human	same	same	qPCR	*Faecalibacterium prausnitzii* and *Bifidobacterium* spp.		Ramirez-Farias et al. (2009)
Inulin (long chain)	3	human	same	same	FISH	*Bifidobacterium* spp., *Lactobacilli* spp. and *Atopobium* spp.	*Bacteroides* spp. and/or *Prevotella* spp.	Costabile et al. (2010)
Inulin	2	human	same	same	FISH	*Bifidobacterium* spp.	*Bacteroides* spp. and/or *Prevotella* spp. and *Clostridium histolyticum*	Kleessen et al. (2007)
Raffinose	3	human	same	same	Sequencing; qPCR	*F. prausnitzii, Bifidobacterium* spp.		Fernando et al. (2010)
Inulin	6	humanized rats	up	propionate (most) and butyrate	HITChip	*Roseburia intestinalis, Eubacterium rectale, Anaerostipes caccae* and *Bifidobacteria longum*	*Akkermansia muciniphila**	Abbeele et al. 2011
Arabinoxylan (long chain)	6	humanized rats	up	propionate butyrate (most)	HITChip	*R. intestinalis, Eubacterium rectale, Anaerostipes caccae* and *B. longum*	*Akkermansia muciniphila**	Abbeele et al. 2011
Arabinoxylan	3	human	same	butyrate	FISH	*Lactobacillus* spp., *Bacteroides* spp., *E. rectale* group, *F. prausnitzii*		Walton 2012
Arabinoxylan-oligosaccharides	3	human	same	butyrate	FISH	*Lactobacillus* spp. and *Bacteroides* spp.		Walton 2012

*= down in cecum, where fermentation occurred, up in faeces.

Table 3.2 A selection of dietary fibre intervention studies and the microbial responses.

In *in vitro* studies some of these species can grow on glucose, and only show lactate utilisation after glucose depletion (Duncan *et al.* 2004). In the *in vivo* situation this might indicate that these lactate-utilising species could switch depending on the dietary availability. However, the amount of monosaccharides in the large intestine is probably not sufficient for lactate-utilizers to switch to monosaccharide fermentation instead of the acetate-lactate fermentation (Cummings and Macfarlane 1991, Duncan *et al.* 2004). Moreover the lactate and acetate utilisation is an important factor for the gut pH homeostasis (Duncan *et al.* 2004, Flint *et al.* 2012b). So far, the identities of the main players in lactate utilisation and what the main SCFA products are, is still being investigated. The main lactate-utilising bacteria might differ when the carbohydrate-metabolising bacteria are different species, or produce different metabolites, depending on the availability and type of the carbon source.

Manipulating the system

Overall, the microbiota possesses remarkably diverse function, and is likely modifiable by diet. In terms of ecological principles, a collective microbiota's higher functional modular complexity, consistent with high diversity environments, appears to correlate with healthy individuals, whilst lower functional complexity consistent with low diversity environments, correlates with dysfunction. If this is indeed the case, then increasing the sugar residue and glycosidic linkage variability and frequency in a sustained manner may be sufficient to promote environmental diversity and ultimately increase functional modular complexity which in turn correlates with gut health. As a consequence of this dietary change, there will be an increase in SCFAs and other related organic acids, which may also confer health benefits.

Finally, substrates targeting (increasing the abundance or activity of) distinct members of the microbiota known to produce specific SCFA species of interest may be a means of addressing particular health concerns. Increasing environmental diversity could be brought about by a diet rich in varied fermentable carbohydrates such as dietary fibre, glycosylated polyphenolics, glucosinolates and other plant glycans, applied in a sustained fashion. We envisage a dietary regime consisting of polymolecular dietary fibre complexes such as cell walls in fruit, vegetables, cereals, nuts and seeds, essentially randomised over time, so that no successive meals are the same. To a large extent, this simulates a normal varied healthy diet.

Increasing SCFA and other beneficial microbial metabolites in a non-specific fashion appears to be simply an outcome of increasing non-specific fermentable carbohydrate consumption. Choices of fermentable carbohydrate then impart a degree of selection over the acids produced. Given sufficient additional information, such as the microbial makeup of the individual's microbiota, we may be able to make informed choices as to which members need to be increased in activity and/or abundance to yield specific results. For example, if increased propionate was the desired response, and a sufficient *Veillonella* population exists, then we could consider increasing lactate production with RS or long-chain inulin in a carbon-rich environment, and rely on the *Veillonella* conversion of lactate to propionate by the

acrylyl CoA pathway (Figure 3.2), whilst in the absence of *Veillonella*, consider arabinoxylans in a low carbon environment being directly fermented to propionate by *Bacteroides* via succinate (Figure 3.2).
The challenge will be to use all the information we have on microbial modulation with dietary fibre: to increase functional modular complexity by using diet to drive increased ecological diversity; or change the SCFA profile to a healthier profile; either increased total SCFA concentrations or enhancing specific acid species concentration.

Acknowledgements

FH is funded by the REINFORCE project (PIRSES-GA-2009-269328) and Netherlands Consortium for Systems Biology (NCSB) which is part of the Netherlands Genomics Initiative and the NWO. Financial support for the preparation and writing of this research article was provided to JAM by a Nga Pae O Te Maramatanga doctoral bridging grant.
We would like to thank John Monro for proof reading the review and useful contributions.

Chapter 4

Distinct responses of mucosal transcriptional profiles and luminal microbiota composition and SCFA concentrations to dietary fibers in murine colon.

Katja Lange*, Floor Hugenholtz*, Henk Schols, Michiel Kleerebezem, Hauke Smidt, Michael Müller and Guido JEJ Hooiveld

*These authors contributed equally

Abstract

Consumption of diets rich in fibers has been associated with beneficial effects on gastrointestinal health. However, detailed studies on molecular effects of fibers in colon are limited. In this study we investigated and compared the influence of five different fibers on the mucosal transcriptome, luminal microbiota and SCFA concentrations in the murine colon. Mice were fed isocaloric diets enriched with fibers that differed in carbohydrate composition, namely inulin (IN), oligofructose (FOS), arabinoxylan (AX), guar gum (GG), resistant starch (RS) or a control diet (corn starch) for 10 days. Gene expression profiling revealed the regulation of specific, but overlapping sets of epithelial genes by each fiber, which on a functional level were mainly linked to transcription/translation and various metabolic pathways including fatty acid oxidation, tricarboxylic acid cycle, and electron transport chain for which PPARγ was predicted to be a prominent upstream regulator. Microbiota profiles were distinct per dietary fiber, but IN, FOS, AX and GG induced a common change in microbial groups. All dietary fibers, except RS, increased SCFA concentrations, albeit to a different extent. Multivariate data integration revealed strong correlations between the expression of genes involved in energy metabolism and the relative abundance of members of *Clostridium* cluster XIVa bacteria.

Introduction

Dietary fibers are complex carbohydrate polymers that escape digestion and absorption in the small intestine (Slavin 1987). Importantly, consumption of fiber-rich diets has been associated with a variety of beneficial health effects, including the improvement of gastrointestinal homeostasis (Aune *et al.* 2011, Balakrishnan and Floch 2012, Eswaran *et al.* 2013, Slavin 2013). A large diversity of types and sources of dietary fibers exist (Slavin 1987, Cummings and Stephen 2007). However, all are fermented in the large intestine by the gut microbiota, mainly generating short chain fatty acids (SCFA) (Nyangale *et al.* 2012, Russell *et al.* 2013). The main SCFA that are produced are acetate, propionate, and butyrate, but their production-ratio is fiber and microbiota composition dependent (Flint *et al.* 2008, Van den Abbeele *et al.* 2011, Van den Abbeele *et al.* 2013b).

Much research has been conducted to investigate effects of dietary fibers on gut health and microbiota composition, however, most *in vitro*, animal or human studies investigated a single fiber at a time. Moreover, data on the genome-wide transcriptional effects of dietary fibers in colonic mucosa are scarce. This is remarkable since there is a major interest in characterizing the genes and networks that are regulated by food components, because this contributes to our understanding of a healthy diet (Müller and Kersten 2003, Afman and Muller 2006). It has only been reported that differential gene expression due to consumption of resistant starch (RS) suggested improvement of structure and function of the gastrointestinal tract in rats (Keenan *et al.* 2012), and induced catabolic but suppressed immune and cell division pathways in the proximal colon of pigs (Haenen

et al. 2013a). In addition, it was shown that oligofructose (FOS) induced expression of genes involved in the TCA cycle, oxidative phosphorylation and proteasome-mediated degradation of intracellular proteins in the rat colon (Rodenburg *et al.* 2008).

Regarding the microbiota, it has been reported that fibers benefit specific groups of bacteria, such as some members of the *Bifidobacteriaceae*, which are specialized to utilize only certain groups of oligosaccharides (Barboza *et al.* 2009, Riviere *et al.* 2014), whereas *Bacteroides* spp. have a range of glycoside hydrolases and are capable of switching between different substrates (Hooper *et al.* 2002). Members in both bacterial groups can degrade starch, which is known to increase butyrate production, although neither of them is known to produce this metabolite (Duncan *et al.* 2004). Bifidobacteria produce mainly lactate and acetate, whereas *Bacteroides* spp. produce several metabolites, including succinate, acetate and propionate (Flint *et al.* 2008). Other bacteria make use of either the sugars released from fibers that are hydrolyzed by these bacteria, or their fermentation products via cross-feeding, leading to the observed increased levels of butyrate (Belenguer *et al.* 2006, Falony *et al.* 2006). As a result dietary fibers modulate the microbiota composition by triggering bacteria that directly feed on them, but also the cross-feeding bacteria that depend on these primary degraders.

The aim of the current study was to comprehensively investigate and compare the effects of five different fibers on the mucosal transcriptome, together with alterations in the luminal microbiota composition in the murine colon. To this end, mice were fed diets enriched with fiber components that differed in carbohydrate composition or a control diet for 10 days. The colonic gene expression profiles and luminal microbiota composition were determined by microarray techniques, and integrated using multivariate statistics. Our data revealed consistent associations between fiber induced enrichment of *Clostridium* cluster XIVa representatives and changes in mucosal gene expression patterns related to energy metabolism. In addition, the latter changes were predicted to involve the nuclear receptor PPAR as an important regulator of these mucosal responses.

Results

Dietary fibers differentially modulate gene expression in colonic epithelial cells

Expression profiling by microarray was performed to assess the genome-wide differences in gene expression in colonic epithelial cells of animals fed different dietary fibers. To compare individual mice on basis of their gene expression profile and diet, sparse PLS-Discriminant Analysis (DA) was performed. In the score plot a clear separation was observed along the x-axis between samples of the CON and RS groups versus the other four fibers, whereas along the y-axis the separation between groups was much less pronounced (Figure 4.1 A). In addition, the 90% confidence ellipses from the FOS, IN and GG groups largely overlapped with each other. These results show that the gene expression profiles were mainly distinguishable based on the chemical composition of the carbohydrate polymers, i.e. the starch vs. the non-starch polysaccharides (NSP). Among the NSP diets, we observed an additional separation between diets consisting of hexose- (FOS, IN, GG) and pentose-polymers (AX). Subsequently, 897 genes were identified that were differentially expressed by any of the 5 fibers (F test $P < 0.01$). A heatmap representing the signal intensities of these genes is given in Figure 4.1 A. In line with the PLS-DA, similar effects of RS and CON diet on gene expression were observed, showing that RS had only minor effects on the transcriptome in colonic epithelial cells as compared to the control diet. The response patterns for IN, FOS, GG and AX were clearly distinct from those for CON and RS. Within the cluster of IN, FOS, AX, and GG samples, the AX samples clustered together, and tended to separate from the majority of IN, FOS and GG samples, although some individual animal variation was observed especially within the FOS group. Based on their expression pattern, genes could be grouped in two main clusters that distinguished RS and CON samples from the other fiber diet samples. Next, the number of significantly ($P < 0.01$) regulated genes per fiber compared to control were determined. The largest number of genes changed was observed for FOS (925), and smallest for RS (287). Only 28 genes were commonly regulated by all fibers, whereas 97 genes were commonly regulated only by the NSP fibers. Taken together, these results showed that RS induced only few, but specific gene expression changes and appeared to be most similar to the control diet. The hexose polymers (FOS, IN, GG) caused relatively similar gene expression changes, while AX induced a quite consistent and specific gene expression pattern that may relate to the fact that this is the only pentose polysaccharide employed in this study.

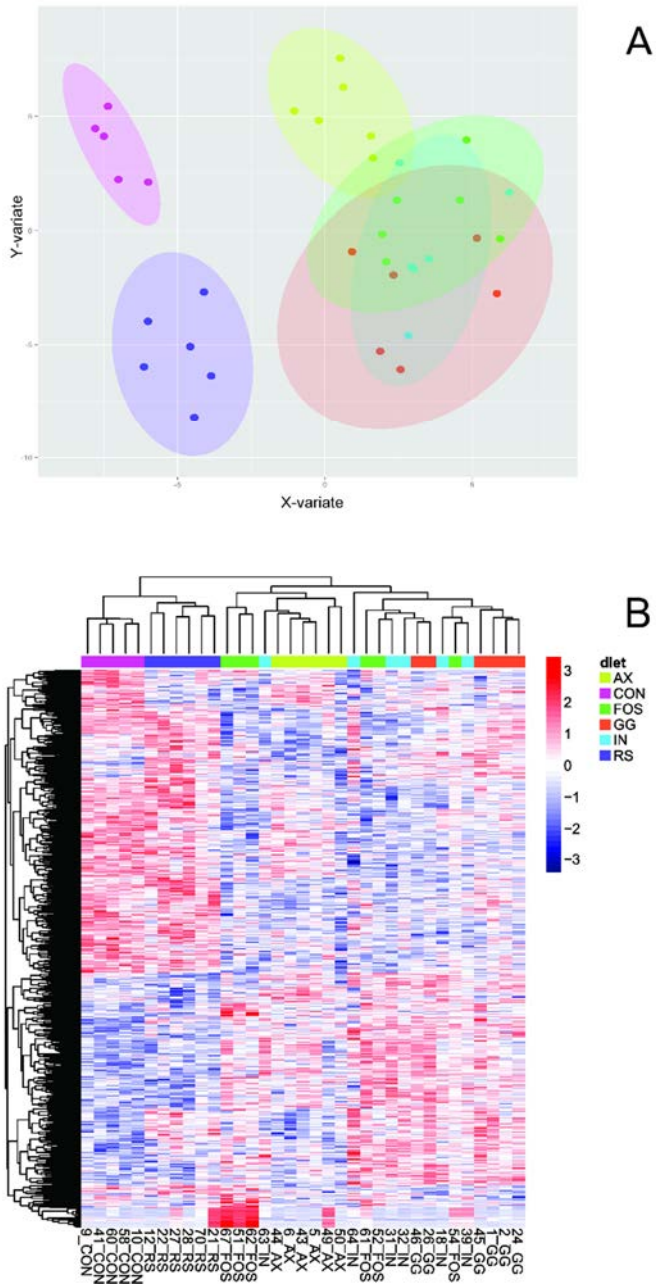


Figure 4.1 Differential effects of dietary fiber on gene expression profiles in colonic epithelial cells. Panel A: PLS-DA score plot of gene expression profiles in colonic epithelial cells of mice fed 5 different fibers or the control diet. In the plot the samples (individual mice) were plotted based on the two main variates. Ellipses indicate 90% confidence intervals of the scores. Panel B: A heatmap was generated to visualize the expression values of the 897 significantly regulated genes across all diet groups (F test P < 0.01). Columns represent the samples

(individual mice), rows represent genes. Both samples and genes were subjected to unsupervised hierarchal clustering based on Euclidean distance. Colors represent row-normalized gene expression values; the color range is continuous and ranges from blue (i.e. low expressed) to red (i.e. high expressed).

Functional implications of differential gene expression in epithelial cells

To gain better insight into the functional changes represented by the regulated genes, gene set enrichment analysis (GSEA) was performed, and significantly changed gene sets (GS) were identified ($P < 0.0001$). The smallest number of changed GS was found for RS (140), whereas for IN (329) the largest number was identified. To better highlight commonalities and differences between fiber diets, an Enrichment Map was created in which all GS were represented that were significantly regulated by at least one of the fibers. Since in this union of regulated GS, inulin modulated the largest number of GS, it was chosen as the fiber to which the effects of all other fibers were compared (Figure 4.2). The resulting Enrichment Map of 604 gene sets displayed two major clusters of functionally related processes, namely energy metabolism and gene transcription/translation (Figure 4.2). Other clusters belonged to immunity, protein-degradation, adhesion dynamics, signaling and morphogenesis. While genes belonging to energy metabolism were induced by IN, FOS, AX and GG, they were not changed or even suppressed by RS compared to CON. In particular, genes belonging to fatty acid catabolism, TCA cycle, and electron transport chain were induced by all fibers except RS. Similarly, target genes of two transcription factors, peroxisome proliferator-activated receptor (PPAR) and nuclear factor (erythroid-derived 2)-like 2 (Nrf2), were induced by IN, FOS, AX and GG. Nevertheless, functional processes similarly regulated by all fiber diets, including RS, were also observed, such as for the category of lipid biosynthetic processes, in which genes belonging to phospholipid and steroid synthesis were induced for all fibers. In addition, most immune-related processes, comprising both adaptive and innate immunity (inflammation, defense response, platelet activation) were modulated by all fiber diets. Within the cluster of innate immunity the gene encoding NFKBIB, which is an inhibitor for NFKB, was commonly yet modestly activated by all fiber diets. While aspects of innate immunity and genes encoding for immunoglobulin domain were consistently activated by all fiber, not all aspects of adaptive immunity were consistently regulated. Within the adaptive immunity gene cluster, the genes belonging to the subcategory of T cell differentiation were suppressed by all fibers, except FOS. Furthermore, we observed that genes belonging to the category adhesion dynamics (cytoskeleton, adhesion function, GTPase signaling) were most significantly activated by the RS diet, while these processes were suppressed (AX, FOS) or unaffected (IN, GG) by the other fiber diets. Genes related to transcription and translation were predominantly suppressed by all fibers. However, within this cluster, genes related to gene-specific transcription were not affected by RS diets and genes related to translation and RNA transport were not affected by FOS. Moreover, we observed a cluster of GS belonging to

protein-degradation (ubiquitin-proteasome system) for which inconsistent regulation by different fibers was observed. While this cluster was most consistently and positively enriched for FOS fed mice, others did not have an effect (AX) or even suppressed theses GS (RS).
Taken together, we found consistent activation of major metabolic pathways and in particular the transcription factor target genes of PPAR and Nrf2 were activated for non-starch fibers. In general, processes related to innate immunity were consistently activated for all fiber whereas transcription/translation was largely suppressed. Other processes such as adhesion dynamics, protein degradation and adaptive immunity were rather inconsistently regulated by the different fiber diets.

Identification of upstream regulators

The underlying mechanisms by which the fibers modulated gene expression changes are not well understood. We therefore aimed to identify potential upstream transcriptional regulators that could explain the observed shifts in gene expression profiles. Next to the canonical involvement of PPAR, the different diets appear to modulate gene sets that are connected to other regulators (Table 4.1). In line with results obtained by GSEA, PPAR, particularly the isoform PPARG, was potentially activated for FOS, AX, GG, but most for IN. In addition, we observed both overlapping and unique sets of PPARG target genes within the fiber specific GS (
Figure S 4.1). Next to PPARG, several other regulators were predicted to play a role in the transcriptional responses elicited by one or more of the fibers. Notably, the transcription regulator KDM5B was uniquely predicted to play a role in control of the RS modulated gene expression profiles, which may at least in part explain the specific gene expression profile induced by RS. KDM5B is a histone demethylase and plays a role in cell fate decisions. In addition, within the FOS-specific transcription pattern changes, NR5A2 and MBD2 were uniquely predicted to play a role in the control of these genes. Among the transcriptional regulators for GG we identified sterol/lipid metabolism related regulator (SREBF1, SREBF2), NR1I2 involved in drug metabolism, and STAT5A and STAT5B involved in cytosolic signaling. For AX, TP63 was specifically identified.
Taken together, we identified potential transcription regulators that may explain commonalities and differences in gene regulation patterns observed for the different fiber diets. In particular, PPARG appears to play a central role in the gene expression response to dietary fiber in colonic epithelia.

Table 4.1 Common and specific potential upstream regulator in colon of mice after feeding different fiber diets as determined by Ingenuity Systems Pathway Analysis Software (Transcriptional regulator and ligand-dependent nuclear receptor which showed an activation z-score ≥ 2 or ≤ -2 and a p-value < 0.05 are displayed)

	Activation score per dietary fiber				
	RS	FOS	AX	IN	GG
PPARG		2.83	2.01	4.23	3.07
HNF4A				2.58	3.50
TP53				2.36	2.82
ATF4		2.61			2.43
PPARGC1A				2.39	2.08
XBP1					2.93
NR5A2		2.61			
SREBF1					2.58
FOXC2				2.43	
SREBF2					2.22
PTTG1				2.21	
NR1I2					2.09
CEBPB				2.02	
KDM5B	2.00				
NCOA2				2.00	
TP63			-2.15		
STAT5B					-2.16
MBD2		-2.23			
STAT5A					-2.36
MYC				-2.63	

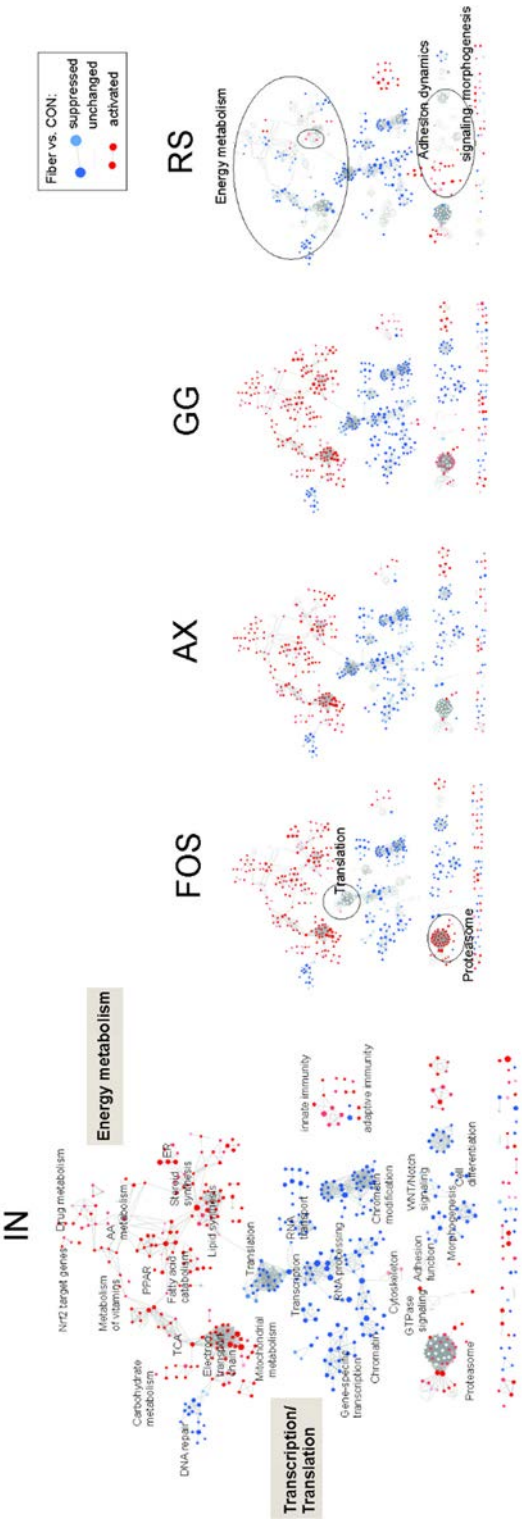


Figure 4.2 Enrichment Map of gene sets that were changed by at least one fiber compared to control. GSEA was performed to identify functional gene sets, i.e. metabolic pathways or signaling transduction routes, that were changed by at least one fiber compared to control (P<0.0001). Nodes represent functional gene sets, and edges between nodes represent their similarity. A red node indicates induction of a gene set, a blue node indicates suppression of a gene set , and a white node indicate no significant regulation of a gene set by a fiber compared to control. Node size represents the gene set size, and edge thickness represents the degree of overlap between 2 connected gene sets. Clusters are manually grouped and labeled to highlight the prevalent biologic functions among related gene sets. As basis layout served the interaction network for IN, because for this fiber most gene sets were changed. Biological processes specific for a fiber are indicated in the respective network.

Dietary fibers differentially modulate microbiota composition

Intestinal content of the colon from four mice per dietary treatment was subjected to microbiota quantification and composition analysis. Although all fibers seemed to increase colonic microbiota density compared to the control diet, no statistical significance was reached (Figure S 4.2). MITChip analysis revealed that all five fibers changed the colonic microbiota composition, except for a single mouse from the RS group that clustered with the mice from the CON group (Figure S 4.3). The microbial diversity, as determined by the Shannon index, did not show any significant differences between different diet groups (data not shown). To relate changes in microbiota composition to the different diets, the hybridization signals of 96 genus-level phylogenetic groups was subjected to redundancy analysis (RDA). Overall, 73.7% of the total variation in microbiota composition was captured within the first two canonical differentiation axes, with diet explaining 34.8% (Figure 4.3). Samples from the RS and control diet clustered separately from IN, AX, FOS and GG. The genus-like groups in the plot that correlated with RS and control diets belong to the Bacteroidetes phylum, and also encompassed specific classes of the Firmicutes phylum (Bacilli, *Clostridium* clusters I, II and IV), and single genus groups of the Actinobacteria, Proteobacteria and Deferribacteres phyla. In the opposite direction groups within *Clostridium* cluster XIVa and a specific genus group of the Bacteroidetes phylum correlated with IN, GG and FOS. The AX diet-group was positioned centrally in the plot, illustrating that this diet did not clearly correlate with changes in any specific bacterial groups. To assess potential differences of individual genus-like groups detected by the MITChip between the different diet groups, all fibers were pairwise compared to the control diet. In total 47 genus-like bacterial groups were significantly different ($P < 0.05$) for at least one of the fibers (Figure 4.4). The main differences included a consistent decrease in members of *Clostridium* clusters I, II and IV and Gammaproteobacteria and an increase in *Clostridium* cluster XIVa genes like groups (except *Ruminococcus obeum* et rel.) for IN, FOS, AX and GG relative to the control diet.

Dietary fibers differentially modulate luminal SCFA levels

Despite differences in fiber type, IN, FOS and GG were highly similar in epithelial cell gene expression responses. Therefore, it was assumed that this might be explained by similar production of fermentation products of these fibers by the microbiota. Moreover, differences between the AX-induced colonic transcriptome patterns and those obtained from mice fed IN, FOS or GG might be, at least partially, explained by the fermentation profiles of the microbiota stimulated by the different fibers. As the main fermentation metabolites of dietary fibers, SCFA were analyzed in the luminal content. Total SCFA concentrations significantly increased in colonic luminal samples obtained from mice that were fed IN, AX, and GG ($P<0.05$), while a similar trend was observed for FOS ($P=0.07$). The highest cumulative SCFA concentrations were observed for mice fed GG, followed by IN.

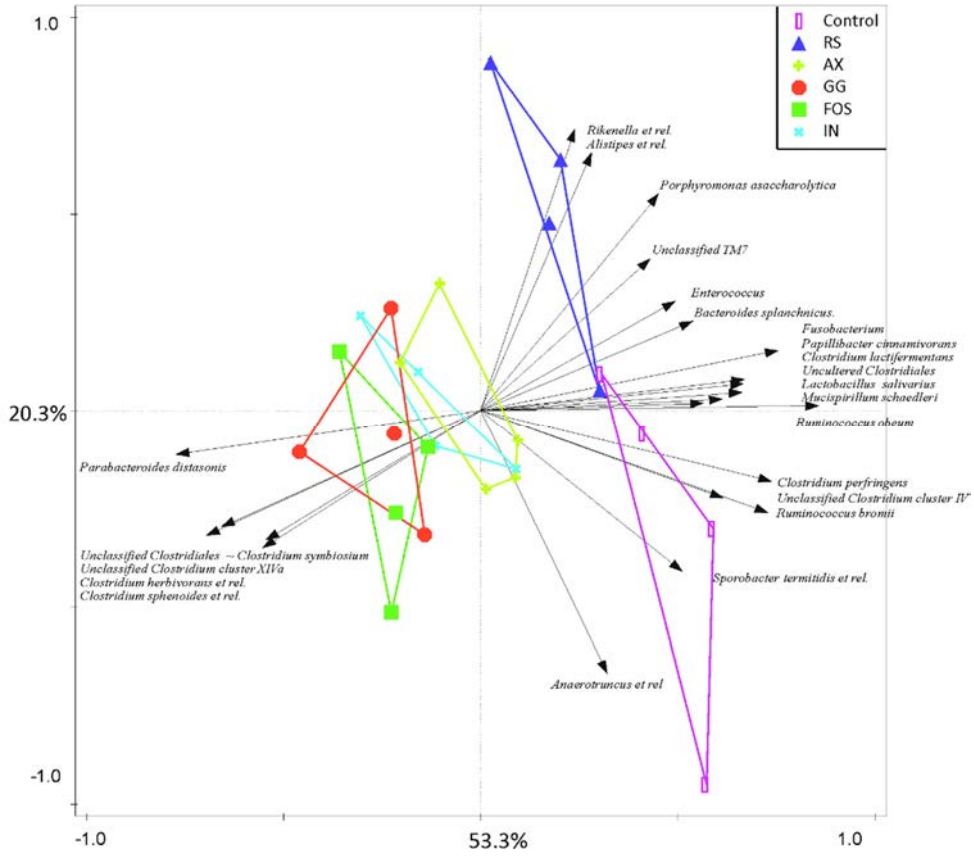


Figure 4.3 Differential modulation of colonic microbiota composition by dietary fibers. Correlation triplot based on a redundancy analysis (RDA) depicting the relationship between colonic luminal microbiota composition and the differences induced by dietary fibers. Dietary fiber, used as explanatory variable, explained 34.8% of the total variation in the microbiota composition, and 73.7% of that variation was explained by the first two canonical axes shown here. Samples are labelled per diet group, and bacterial groups are indicated by arrows. The arrows point in the direction of maximal variation in the species abundances, and their lengths are proportional to their maximal rate of change. Long arrows correspond to species contributing more to the data set variation. Right-angle projection of a sample dot on a species arrow gives approximate species abundance in the sample.

		RS	FOS	AX	IN	GG
Bacteroidetes	*Bacteroides distasonis* et rel.		*	*	*	*
	Bacteroides splanchnicus et rel.					*
	Bacteroides vulgatus et rel.					*
	Porphyromonas asaccharolytica et rel.					*
	Rikenella et rel.	*				
Deferribacteres	*Mucispirillum schaedleri* et rel.			*		*
Fibrobacteres	*Fibrobacter succinogenes* et rel.					*
Bacilli	*Staphylococcus aureus* et rel.					*
	Streptococcus intermedius et rel.					*
Clostridium cluster I	*Clostridium perfringens* et rel.					*
Clostridium cluster II	*Clostridium lactifermentans* et rel.		*	*	*	
	Unclassified *Clostridiales* II		*	*		
Clostridium cluster IV	*Anaerotruncus* et rel	*		*		*
	Papillibacter cinnamivorans et rel.			*		*
	Ruminobacter amylophilus et rel.					*
	Ruminococcus bromii et rel.	*	*	*	*	*
	Ruminococcus callidus et rel			*		
	Sporobacter termitidis et rel.	*	*			
	Unclassified *Clostridiales* IV			*		*
	Uncultured *Clostridiales*		*			*
Clostridium cluster IX	*Peptococcus niger* et rel.			*	*	*
Clostridium cluster XI	*Anaerovorax* et rel.			*		
	Clostridium difficile et rel.					*
Clostridium cluster XIVa	*Clostridium herbivorans* et rel.		*	*		*
	Clostridium sphenoides et rel.	*				*
	Clostridium symbosium et rel.			*		
	Dorea et rel.	*				*
	Lachnospira pectinoschiza et rel.			*	*	
	Roseburia intestinalis et rel.		*			
	Ruminococcus obeum et rel.		*	*	*	*
	Unclassified *Clostridiales* -Close to *Clostridium symbiosium* et rel.		*		*	
	Unclassified *Clostridiales* XVI					*
	Solobacterium moorei et rel.					*
Fusobacteria	*Fusobacterium*		*	*		*
Gammaproteobacteria	*Vibrio* et rel.					*

*Figure 4.4 Heatmap of phylogenetic groups that were significantly affected by any of the dietary fibers. All fibers were pairwise compared to the control diet to identify genus-like bacterial groups that were affected by any of the fibers. In total 47 genus-like bacterial groups were significantly different ($P < 0.05$). Changes in abundance are depicted in colors; red: increased, blue: decreased. * indicate statistical significance as determined by Mann-Whitney U-test.*

In contrast, the SCFA concentrations in samples obtained from mice fed RS were comparable to those obtained from the control fed animals (Figure 4.5 B). In all diet groups, the acetate concentration was highest. ANOVA analysis revealed that mean concentrations of both acetate and propionate were significantly different between any of the fibers (P <0.05) (Figure 4.5 A). Specifically, acetate concentrations in samples obtained from mice fed AX, IN and GG were significantly higher compared to samples obtained from mice fed CON and RS diet, while propionate concentrations were significantly higher in samples from mice fed FOS, AX, IN and GG compared to CON. Butyrate concentrations were not significantly different between the diet groups but showed a trend (P =0.061). No significant difference was observed for iso-butyrate, valerate or iso-valerate . Taken together, these results show that the different fibers increased total colonic luminal SCFA to a

different extent. The highest SCFA concentrations were found for IN and GG.

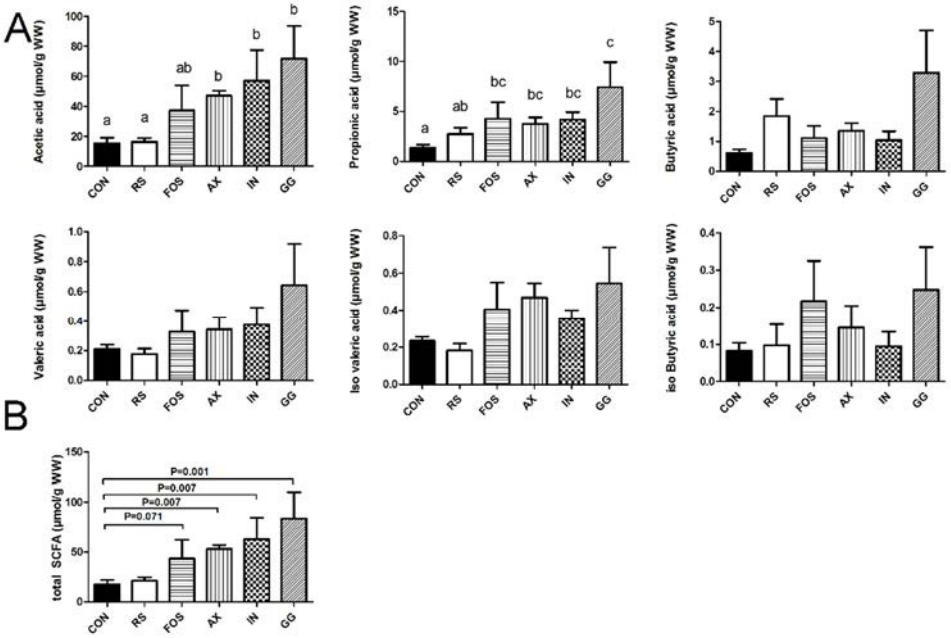


Figure 4.5 Dietary fibers differentially modulate luminal SCFA levels. Colonic luminal SCFA concentrations in µmol/g colonic content were measured with gas chromatography. The mean ± SEM for A) individual SCFA and B) total SCFA concentrations for each diet group is represented in a bar plot. Different letters indicate statistical difference between the diet groups as tested with ANOVA and Tukey post hoc test.

Integrative analysis of changes in colonic epithelial cell gene expression and luminal microbiota composition

To get insight in the interaction between changes in gene expression and microbiota composition, and to generate new hypotheses about potential mechanisms explaining host transcriptional responses to fiber fermentation by the microbiota, we investigated how changes in microbiota composition correlated with changes in colonic epithelial cell gene expression. The correlation pattern between microbiota and gene expression across 22 samples was visualized in a heat map. This clustered heatmap revealed five clusters of genes and three main clusters of discriminating bacterial groups (Figure 4.6) (Gene lists are available upon request) for all genes and bacteria per cluster). The strongest correlations were found for bacteria in cluster C, which positively correlated with genes in cluster 1, but negatively correlated with genes in cluster 5. This cluster C contained known butyrate-producing bacteria belonging to *Clostridium* cluster XIVa. These bacteria co-clustered with saccharolytic bacteria such as *Parabacteroides distasonis* (Sakamoto and Benno 2006) in cluster C, which suggests a potential cross-feeding relationship. Genes in cluster 1 were involved in metabolic, energy-generating and oxidative processes, whereas genes in cluster 5 were involved in adhesion dynamics and signaling. While these

processes were positively correlating with bacteria from *Clostridium* cluster XIVa, *Turicibacter* et. rel and *Clostridium perfringens* showed negative correlation with these sets of genes. Correlation of the three main SCFA, acetate, propionate and butyrate with host gene expression showed strongest correlation for acetate and propionate (data not shown). The strongest correlation was found for acetate correlating with genes related to the functions adhesion dynamics (Cmah) and immunity (Trim15, Nos2, Duoxa2, Atf3). Thus, multivariate analyses revealed strong correlations between gene expression changes and relative abundance of bacteria, among which members of *Clostridium* cluster XIVa stood out because of strong association with mucosal gene expression patterns.

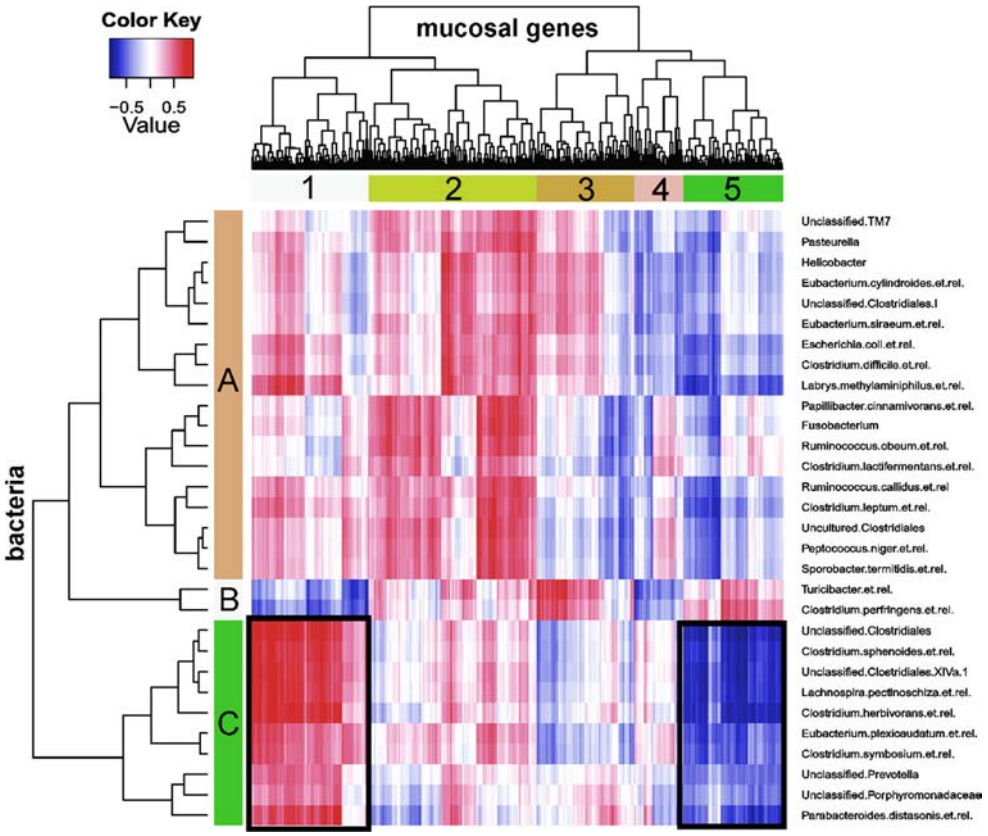


Figure 4.6 Integration of epithelial cell gene expression with luminal microbiota composition. Sparse PLS canonical correlation analysis was performed to integrate gene expression values with relative abundance data of bacteria for individual mice. The heatmap represents the correlation structure of both dataset; red: positively correlated, blue: negatively correlated. The more intense the color is, the higher the correlation value. Correlation values were subjected to unsupervised hierarchal clustering based on Euclidean distance for both genes and microbial groups. Three main gene clusters, and five main bacterial clusters were identified.

Discussion

In the current study five different dietary fibers were fed to mice to comprehensively study the effects on epithelial cell gene expression, luminal microbiota composition and SCFA concentrations. We found that each fiber provoked a specific, but overlapping response with respect to these different aspects.

PPARγ is proposed to be a central regulator of transcriptional responses to fiber-diets.

Dietary fibers have diverse effects on the microbiota and host metabolism, of which the latter is most likely mediated by the SCFA that have been shown to elicit diverse effects on colonic gene expression patterns (Vanhoutvin *et al.* 2009, Alex *et al.* 2013). To the best of our knowledge, however, it has not yet been determined how different dietary fibers can modulate colonic gene expression in the same mouse model. At the start of this study, we anticipated that different dietary fibers would induce differential responses in terms of colonic epithelial gene expression. Our results showed that many of these transcriptional responses are conserved for several fibers on the level of functional implications, but that specific genes are more variable. Among the functional implications, target genes of the nuclear receptor PPAR were found to be commonly regulated, albeit to a different extend, i.e. different targets were regulated by non-starch fiber diets and in different magnitude (Figure S 4.1), which may be indicative for differences in SCFA fluxes across the colonic epithelium. Furthermore, the main separation of the diet-associated transcriptome profiles was obtained for mice on non-starch fiber diets compared to starch diets. This separation coincided with changes in the cumulative SCFA levels in the colonic lumen. SCFA have recently been identified as modulators of PPARγ *in vitro* (Alex *et al.* 2013). Therefore, we conclude that PPARγ appears to play a prominent role in regulation of the *in vivo* responses to altered fermentation activity and increased SCFA levels in the colon. It has been shown that PPARγ mainly regulates genes involved in metabolism, in particular lipid metabolism, but also affects signaling, motility and cell adhesion (Su *et al.* 2007). In our study, especially the IN fed mice displayed an increased expression level of genes involved in lipid metabolic processes. Taken together we propose that fiber fermentation by the colon microbiota is leading to activation of PPARγ in the colon epithelia, probably through increased levels of SCFA exposure. Nevertheless, the variability of gene response magnitude as well as specific gene expression patterns that may be related to PPARγ activation and the role of SCFA production patterns warrant further research. Activation of PPARγ is also of interest because this transcription factor has been demonstrated to coordinate the expression of anti-inflammatory properties in inflammatory bowel disease (Dubuquoy *et al.* 2006) and appears to play a pivotal role in the interplay between metabolism and immune function regulation (Hou *et al.* 2012).

Clostridium cluster XIVa correlates with epithelial cell metabolic pathways

We showed that dietary fibers yielding increased SCFA concentrations in the colonic lumen commonly increased the transcription of genes involved in metabolic processes associated with energy metabolism. The data also demonstrate relationships between the microbiota composition and these changing host gene expression patterns. In particular bacterial groups within *Clostridium* cluster XIVa correlated with genes involved in energy metabolism. This bacterial group is known to encompass many secondary fermenters, of which several have been shown to produce butyrate as their metabolic end product (Louis and Flint 2009). In addition, *Clostridium* cluster XIVa adherence to the mucosal layer has been reported using *in vitro* models, a feature which may facilitate the delivery of butyrate directly to the epithelial cells (Van den Abbeele *et al.* 2013a). Several of these known butyrate producers were found to positively correlate with the gene expression changes associated with energy metabolism. Unfortunately, and analogous to many other studies, the data presented here are based on single time-point measurements, and thereby fail to represent actual production or absorption rates of SCFA. Such flux data could give a considerable refinement to our understanding of the rate of production of butyrate by these bacteria and the actual levels of butyrate flux experienced by the colonic epithelia, respectively. Interestingly, the *Clostridium* cluster XIVa bacteria co-clustered with *Parabacteroides distasonis*, for which it has been shown that its oral administration to mice can reduce inflammation in DSS induced colitis, which was proposed to be exerted by specific immune regulatory mechanisms as well as changes in the colonic microbiota (Kverka *et al.* 2011). This observation further supports the anti-inflammatory potential of dietary fibers, which may be linked to their stimulatory effects on specific bacterial groups that produce anti-inflammatory antigens and/or metabolites.

Specific effects of dietary fiber

The transcriptional effects observed in colonic epithelia in mice fed RS or AX was quite distinct from those obtained for mice fed GG, IN and FOS. The transcriptomes associated with AX clustered separate from the mixed transcriptome clustering of the IN, GG and FOS samples, which may be related to the particular chemistry of AX. This fiber is a polymer consisting of arabinose and xylose moieties, and thus consists of pentose- rather than hexose-polymers, the latter of which are the building blocks of IN, FOS, GG and RS. The transcriptome patterns obtained from mice fed the RS diet were clearly distinct from any of the other fiber diets, which coincided with a clear separation of the RS diet on basis of colonic luminal SCFA concentration and microbiota composition data. Preliminary analyses of mono-, di- and oligosaccharides in the intestinal lumen samples revealed that RS fed mice contained large amounts of non-degraded starch in their colonic lumen as compared to the other diets (Figure S 4.4). Hence, the deviating effects of RS might be explained by difference in fermentation process, possibly leading to production of other microbial metabolites. Next to SCFA, there are a variety of other metabolites

formed by the intestinal microbiota (Russell *et al.* 2013).

Conclusion

Using comprehensive transcriptome and microbiome analysis we showed common and specific effects of dietary fibers on gene expression patterns in the colonic epithelia and microbiota and SCFA compositions in colonic lumen. The common regulation of genes involved in energy metabolism by some dietary fibers is proposed to involve the increased colonic SCFA level that can modulate the transcription factor PPARγ. The transcriptional regulation of epithelial metabolic processes was strongly correlated with the abundance of members of *Clostridium* cluster XIVa, among which were several known butyrate producers. Taken together, the data illustrate that dietary fibers that can induce increased total SCFA concentrations in combination with the abundance of *Clostridium* cluster XIVa may have a beneficial effect on colonic tissue homeostasis by targeting PPARγ. This modulation of the immuno-metabolism within the mucosa, leading to increased anti-inflammatory characteristics of the colonic tissue, which to a certain extent appeared to be independent of the fiber-type used. The fiber-specific differences, however, may also be of importance, as they may have location specific implications due to differences in microbiota composition and activity, differences in the secretion of microbe-derived factors and the related host-response.

Acknowledgments

We would like to thank Dr Diederick Meyer (Sensus) and Dr Hans van der Saag (BioActor) for their kind gifts of inulin and oligofructose, and arabinoxylan, respectively. This work was (co)financed by the Netherlands Consortium for Systems Biology (NCSB), which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.
The authors have declared no conflict of interest.

Methods

Ethics statement

The institutional and national guidelines for the care and use of animals were followed and the experiment was approved by the Local Committee for Care and Use of Laboratory Animals at Wageningen University (DRS code: 2010167).

Animals, diets, design and sampling

Male C57BL/6J mice were purchased from Charles River Laboratories (Maastricht, the Netherlands) at 6 weeks of age. Mice were housed in pairs in a light- and temperature-controlled animal facility of Wageningen University (12 hour light-dark cycle; light on from 11h PM to 11h AM, 21 °C). Mice had free access to water and food throughout the entire experimental period. Upon arrival, mice were fed standard lab chow (RMH-B, Arie Blok, Woerden, the Netherlands) for 3 wks. Subsequently, all mice were adjusted to the control diet, a standard semi-synthetic low fat diet containing corn starch, for 2 wks. To achieve similar weight distribution among the diet groups, mice were stratified according to their body weight to one of the six diet groups (n=10 per diet group), i.e. control (CON), inulin (IN), oligofructose (FOS), arabinoxylan (AX) , guar gum (GG) or resistant starch type 3 (RS). The diets enriched in fiber were identical to the control diet, except that 10% (w/w) of corn starch was replaced by each fiber (20% for RS, see below). Inulin (brand name Frutafit IQ) and oligofructose (brand name Frutalose OFP) were a gift of Dr Diederick Meyer (Sensus, Roosendaal, the Netherlands); arabinoxylan (brand name NAXUS) was a gift of Dr Hans van der Saag (BioActor, Maastricht, the Netherlands); guar gum (brand name Viscogum) and resistant starch (brand name ActiStar) were obtained from Cargill R&D Centre Europe (Vilvoorde, Belgium). According to the supplier, the RS was only 50% resistant to digestion in the small intestine, and was therefore included in the diet at a 20% (w/w) level, i.e. double the amount of the other fibers. Diets were prepared by Research Diet Services (Wijk bij Duurstede, The Netherlands). Detailed composition of the diets is presented in Table S 4.1. Mice were fed the fiber or control diets for 10 days. On the day of sections, mice were fasted for 4hrs (starting at 5 AM). Mice then received a calibrated meal of 1g of their habitual diet to reduce the inter-individual variation in physiological state at time of tissue collection. Four hours later mice were anaesthetized with isoflurane, and the colon was excised. The adhering fat around the colon was carefully removed, and the colon was cut open longitudinally. The luminal content was sampled and the tissue was rinsed with ice-cold phosphate buffered saline. Subsequently, the epithelial lining of the colon was scraped off. Luminal content and scrapings were collected in tubes, which were immediately snap frozen in liquid nitrogen and stored at -80°C.

RNA isolation, Affymetrix microarray processing and analysis.

Colonic scrapings (n=6 per diet group) were subjected to genome-wide expression

profiling. In brief, total RNA was isolated from epithelial scrapings and were hybridized on Mouse Gene 1.1 ST arrays (Affymetrix). Packages from the Bioconductor project (Gentleman *et al.* 2004), integrated in an online pipeline (Lin *et al.* 2011), were used for quality control and statistical analysis of the array data. Due to insufficient quality, 1 array from the control group had to be excluded from further analysis. The dataset was filtered to only include probe sets that were active (i.e. expressed) in at least 5 samples using the universal expression code (UPC) approach (UPC score > 0.50) (Piccolo *et al.* 2013). This resulted in the inclusion of 8,831 (42%) of the 21,187 probe sets. Differentially expressed probe sets were identified by using linear models and an intensity-based moderated t-statistic (Smyth 2004, Sartor *et al.* 2006). Probe sets that satisfied the criterion of $P < 0.01$ were considered to be significantly regulated. Array data have been submitted to the Gene Expression Omnibus under accession number GSE59494. Detailed information on microarray processing and data analysis can be found under Supplemental Methods.

Functional analysis of array data

Changes in gene expression were related to functional changes using gene set enrichment analysis (GSEA) (Subramanian *et al.* 2005). The Enrichment Map plugin for Cytoscape was used for visualization and interpretation of the GSEA results (Merico *et al.* 2010). Upstream Regulator Analysis in IPA (content version 18030641 released 2013; Ingenuity Systems) was used to identify the cascade of potential upstream transcriptional regulators that may explain the observed gene expression changes in the data set, and whether they are likely activated or inhibited. Functional annotation of selected genes, e.g. identified by multivariate correlation analysis, was performed in Enrichr (Chen *et al.* 2013a).

DNA isolation, microbiota MITchip processing and analysis

Total bacterial DNA was extracted from colonic luminal content samples (n=4 per diet group) using the repeated bead beating plus column method (Yu and Morrison 2004). Quantification of the overall bacterial community density was performed by qPCR targeting the 16S rRNA gene, whereas the microbial community composition was analyzed using the Mouse Intestinal Tract Chip (MITChip) (Geurts *et al.* 2011b) (for further details also see Supplemental Methods). The relative abundance of 96 genus-level bacterial groups detected on the MITchip was determined by the Robust Probabilistic Averaging algorithm (Lahti *et al.* 2013). Pairwise statistical testing for differences for microbial groups between each fiber diet and the control group was done using the Mann-Whitney U-test. Groups that satisfied the criterion of $P < 0.05$ were considered to be significantly affected. To assess the correlation of the microbial groups with all diets groups, multivariate redundancy analysis (RDA) was performed as implemented in Canoco for Windows 4.5 (Lepš and Šmilauer 2003). The Monte Carlo Permutation test was used to assess the significance of the variation in the dataset in relation to the diet.

Short-chain fatty acid analysis in colonic luminal content

Luminal samples (n=3-5 per diet group) were analyzed for SCFA concentration by gas chromatography as described before (Haenen *et al.* 2013c).

Multivariate integration and correlation analysis.

To get insight into the interactions between changes in gene expression and microbiota composition, the datasets were combined using the linear multivariate method partial least squares (PLS) (Boulesteix and Strimmer 2007). This analysis ignores diet group membership. For 15 mice both gene expression and microbiota composition data was available, but to increase power the dataset was expanded with 7 measurements performed in mice housed in the same cage. Since we did not want to make any 'a priori' assumption on the relationship between the two sets of variables that were analyzed, the canonical correlation framework of PLS was used (Le Cao *et al.* 2009b). Both datasets were log2 transformed before analysis, and the correlation matrices were visualized in clustered image maps (Gonzalez *et al.* 2012). Analyses were performed in R using the library mixOmics (Le Cao *et al.* 2009).

Supplemental information

Supplemental Figures

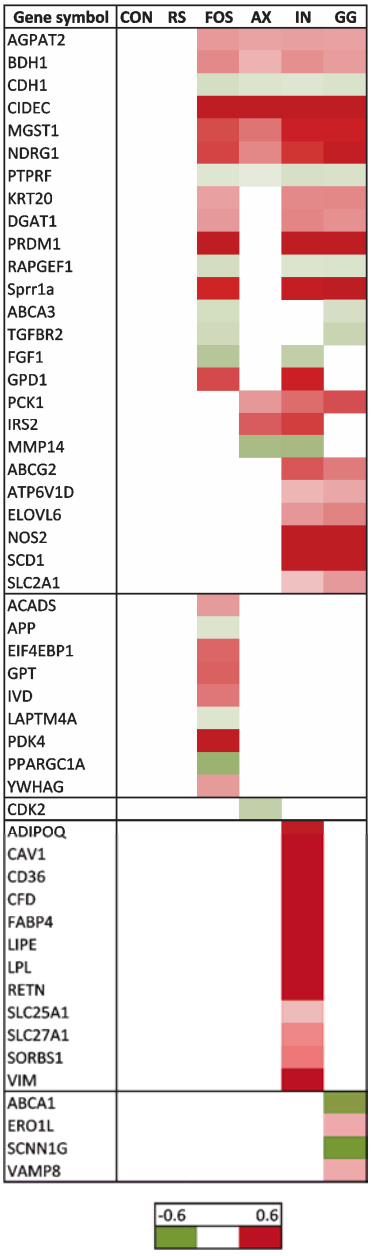


Figure S 4.1 Upstream regulator analysis. PPARG target genes were determined by Ingenuity Pathway Analysis. A heatmap represents the relative gene expression values for each fiber diet compared to control. Red indicates increased expression, while green indicates decreased expression. The relative gene expression values were log2 transformed, i.e. a fold change of 0.6 means 1.5 fold upregulation, while -0.6 means -1.5 downregulation with fiber compared to control.

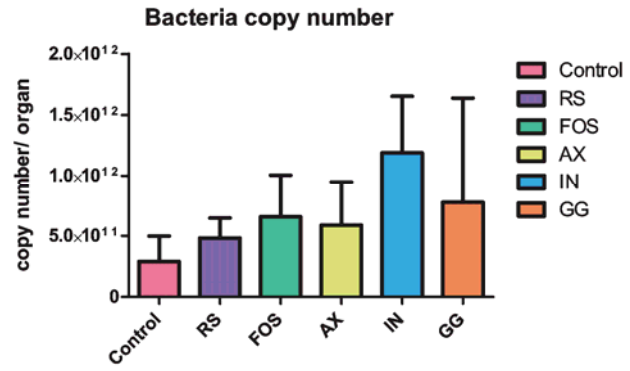


Figure S 4.2 Quantitative PCR on total bacteria. 16S rRNA gene-targeted qPCR was used to assess total bacterial numbers. The copy number per 16S rRNA gene was calculated back to total copy number per organ weight.

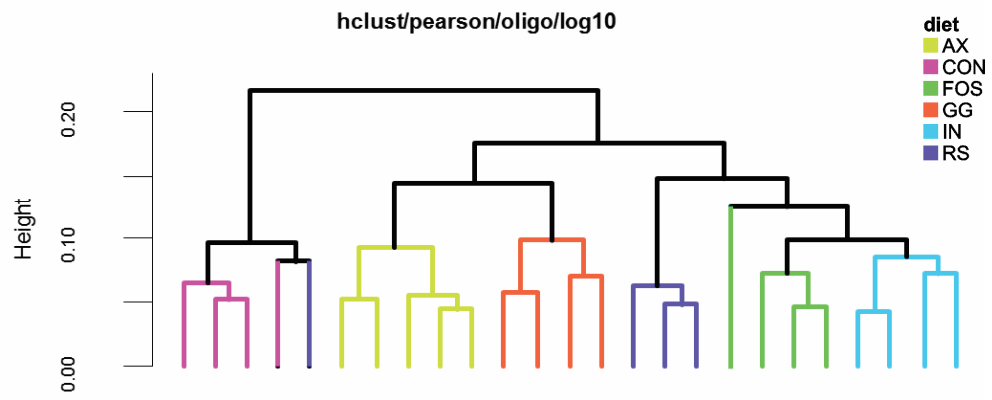


Figure S 4.3 Clustering of MITChip profiles at the probe-level.
Pearson distance-based clustering of the samples on log10 transformed probe level data of the MITChip.

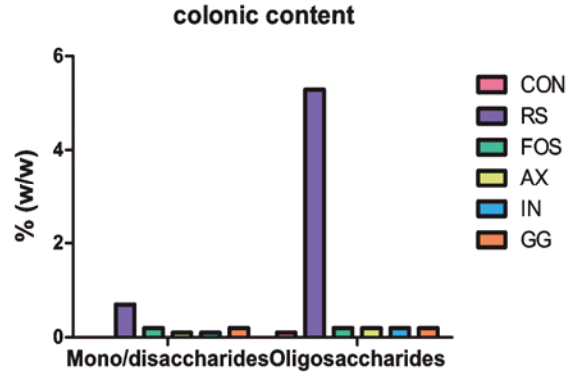


Figure S 4.4 Carbohydrate content analysis.
Carbohydrate content was analyzed in luminal content of mice (n=1; per group) using HPAEC (High-performance Anion Exchange Chromatography). The mono/disaccharides were calculated based on glucose, and the oligosaccharides were calculated based on FOS.

Supplemental Tables

	CON		IN		FOS		AX		GG		RS	
Based on formula #	**D12450B**		**10% INULIN**		**10% FOS**		**10% NAXUS**		**10% GUAR GUM**		**20% Resistant**	
Ingredient	**gm**	**% (w/w)**	**gm**	**% (w/w)**	**gm**	**% (w/w)**	**gm**	**% (w/w)**	**gm**	**% (w/w)**	**gm**	**% (w/w)**
Casein, lactic	200.0	*19.0*	200.0	*19.0*	200.0	*19.0*	200.0	*19.0*	200.0	*19.0*	200.0	*19.0*
L-Cystine	3.0	*0.3*	3.0	*0.3*	3.0	*0.3*	3.0	*0.3*	3.0	*0.3*	3.0	*0.3*
Corn Starch (control fiber)	427.2	*40.5*	321.7	*30.5*	321.7	*30.5*	321.7	*30.5*	321.7	*30.5*	216.2	*20.5*
Maltodextrin	100.0	*9.5*	100.0	*9.5*	100.0	*9.5*	100.0	*9.5*	100.0	*9.5*	100.0	*9.5*
Sucrose	172.8	*16.4*	172.8	*16.4*	172.8	*16.4*	172.8	*16.4*	172.8	*16.4*	172.8	*16.4*
Experimental fiber	nvt		105.5	*10.0*	105.5	*10.0*	105.5	*10.0*	105.5	*10.0*	211.0	*20.0*
Cellulose, BW200	50.0	*4.7*	50.0	*4.7*	50.0	*4.7*	50.0	*4.7*	50.0	*4.7*	50.0	*4.7*
Soybean Oil	25.0	*2.4*	25.0	*2.4*	25.0	*2.4*	25.0	*2.4*	25.0	*2.4*	25.0	*2.4*
Palm oil	20.0	*1.9*	20.0	*1.9*	20.0	*1.9*	20.0	*1.9*	20.0	*1.9*	20.0	*1.9*
Mineral Mix S10026*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*
DiCalcium Phosphate	13.0	*1.2*	13.0	*1.2*	13.0	*1.2*	13.0	*1.2*	13.0	*1.2*	13.0	*1.2*
Calcium Carbonate	5.5	*0.5*	5.5	*0.5*	5.5	*0.5*	5.5	*0.5*	5.5	*0.5*	5.5	*0.5*
Potassium Citrate, 1 H2O	16.5	*1.6*	16.5	*1.6*	16.5	*1.6*	16.5	*1.6*	16.5	*1.6*	16.5	*1.6*
Vitamin Mix V10001**	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*	10.0	*0.9*
Choline Bitartrate	2.0	*0.2*	2.0	*0.2*	2.0	*0.2*	2.0	*0.2*	2.0	*0.2*	2.0	*0.2*
FD&C Yellow Dye #5	0.1	*0.0*	0.1	*0.0*	0.1	*0.0*	0.1	*0.0*	0.1	*0.0*	0.1	*0.0*
FD&C Red Dye #40	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*
FD&C Blue Dye #1	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*	0.0	*0.0*
Total	**1055.1**	*100.0*	**1055.1**	***100.0***	**1055.1**	***100.0***	**1055.1**	***100.0***	**1055.1**	***100.0***	**1055.1**	

Table S 4.1 Diet composition

**Mineral mix S10026 contains the following (g/kg mineral mix): 41.9 magnesium oxide, 257.6 magnesium sulfate•7H2O, 259 sodium chloride, 1.925 chromium KSO4•12H2O, 1.05 cupric carbonate, 0.035 potassium iodate, 21 ferric citrate, 12.25 manganous carbonate, 0.035 sodium selenite, 5.6 zinc carbonate, 0.20 sodium fluoride, 0.30 ammonium molybdate•4H2O, 399.105 sucrose.*

***Vitamin mix V10001 contains the following (g/kg vitamin mix): 0.80 retinyl palmitate, 1.0 cholecalciferol, 10 all-rac-a-tocopheryl acetate, 0.08 menadione sodiumbisulfite, 2.0 biotin*

(1.0%), 1.0 cyancocobalamin (0.1%), 0.20 folic acid, 3.0 nicotinic acid, 1.6 calcium pantothenate, 0.70 pyridoxine-HCl, 0.60 riboflavin, 0.60 thiamin-HCl, and 978.42 sucrose.

Supplemental Methods

Microarray processing and data analysis

Total RNA (100 ng) was used for whole transcript cDNA synthesis by using the Ambion WT expression kit (Life Technologies) and subsequently labelled using the Affymetrix GeneChip WT Terminal Labeling Kit. Samples were hybridized on Mouse Gene 1.1 ST arrays (Affymetrix), washed, stained, and scanned on an Affymetrix GeneTitan instrument. Detailed protocols for array handling can be found in the GeneChip WT Terminal Labeling and Hybridization User Manual (P/N 702808, Rev. 7; Affymetrix). Packages from the Bioconductor project (Gentleman *et al.* 2004), integrated in an online pipeline (Lin *et al.* 2011), were used to analyze the array data. Various advanced-quality metrics, diagnostic plots, pseudoimages, and classification methods were used to determine the quality of the arrays before statistical analysis (Heber and Sick 2006). The probes on the Mouse Gene 1.1 ST array were redefined using current genome information (Dai *et al.* 2005). In this study, probes were reorganized on the basis of the gene definitions available in the NCBI *Mus musculus* Entrez Gene database based on the mouse genome build 38 patch release 1 (GRCm38.p1) (custom CDF v17). Normalized gene expression estimates were obtained from the raw intensity values using the robust multiarray analysis preprocessing algorithm available in the library 'AffyPLM' using default settings (Irizarry *et al.* 2003). Differentially expressed probe sets (genes) were identified by using linear models, applying moderated t-statistics that implemented empirical Bayes regularization of SEs (Smyth 2004). To adjust for both the degree of independence of variances relative to the degree of identity and the relation between variance and signal intensity, the moderated t-statistic was extended by a Bayesian hierarchical model to define an intensity-based moderated t-statistic (Sartor *et al.* 2006).

Microbiota analysis-*MITChip*

This phylogenetic microarray was designed using criteria of the Human Intestinal Tract Chip (HITChip) (Rajilic-Stojanovic *et al.* 2009). The MITChip consists of 3,580 different oligonucleotides specific for the mouse intestinal microbiota (Rajilic-Stojanovic *et al.* 2009, Geurts *et al.* 2011a, Reikvam *et al.* 2012). The array targets the V1 and V6 regions of bacterial 16S rRNA genes. The 16S rRNA genes were amplified from twenty nanogram of intestinal extracted DNA with the primers T7prom-Bact-27-F and Uni-1492-R (Table S 4.2). PCR products were then transcribed, and RNA was labelled with Cy3 and Cy5 dyes and fragmented as described previously (Rajilic-Stojanovic *et al.* 2009, Geurts *et al.* 2011a, Reikvam *et al.* 2012). Finally the samples were hybridized on the arrays at 62.5^{o}C for 16 hours in a rotation oven (Agilent Technologies, Amstelveen, The Netherlands). After washing and scanning of the slides, data was extracted with the Agilent Feature

Extraction software, version 9.1. The data was normalized and analysed using a set of R-based scripts in combination with a custom-designed relational database, which operates under the MySQL database management system.

Quantification of bacterial community

Quantification of the bacterial 16S rRNA gene was done by a qPCR assay using the primers developed by (Suzuki *et al.* 2000). The qPCRs were performed in 384-well plates (BioRad) sealed with a film (Microseal B film, Bio-Rad) using a MyIQ cycler with MyIQ software (version 1.0.410, Bio-Rad). The reactions were carried out in a total volume of 12.5 µl consisting of 1x IQ SYBR green Supermix (Bio-Rad), 200 nM of the forward and reverse primer and 2 µl of template DNA, and the cycling program and melting curve analysis as previously described (van den Bogert *et al.* 2011). The standard curve consisting of a 8-fold dilution series was a 16S rRNA gene PCR product of Escherichia coli top10.

Primer name	Sequence	Application
T7prom-Bact-27-F	5'-TGA ATT GTA ATA CGA CTC ACT ATA GGG GTT TGA TCC TGG CTC AG–3'	MITChip
Uni-1492-R	5'-CGG CTA CCT TGT TAC GAC-3'	MITChip
PROK1492R	5' -GGW TAC CTT GTT ACG ACT T-3'	QPCR
BACT1369F	5'-CGG TGA ATA CGT TCY CGG-3'	QPCR

Table S 4.2 List of primers (Suzuki et al. 2000, Rajilic-Stojanovic et al. 2009)

Chapter 5

Functional profiling of unfamiliar microbial communities using a validated de novo assembly metatranscriptome pipeline

Mark Davids*, Floor Hugenholtz*, Vitor Martins dos Santos, Hauke Smidt, Michiel Kleerebezem and Peter J. Schaap

*These authors contributed equally

Abstract

Background

Metatranscriptomic landscapes can provide insights in functional relationships within natural microbial communities. Analysis of complex metatranscriptome datasets of these communities poses a considerable bioinformatic challenge since they are essentially non-restricted with a varying number of participating strains and species. For the RNA-Seq data type a standard approach is to align the generated reads to a set of closely related reference genomes. This only works well for microbial communities for which a near complete catalogue of reference genomes is available at a small evolutionary distance. In this study, we focus on the development of a validated *de novo* metatranscriptome assembly pipeline to obtain functional profiles from the caecal microbial communities of four C57BL/6J mice fed on a high-fat high protein diet.

Results

A *de novo* assembly pipeline was developed using RNA-Seq data from an *in silico* generated mock community and further validated using *in vivo* RNA-Seq data from a restricted microbial community taken from an inbred non-obese diabetic (NOD) mouse model colonized with Altered Schaedler Flora (ASF). Precision and recall of gene expression, functional and taxonomic profiles obtained were compared to those obtained with a standard alignment method. The validated pipeline was subsequently used to generate expression profiles from non-restricted caecal communities of four C57BL/6J mice fed on a high-fat high-protein diet spiked with an RNA-Seq data set from a well-characterized human small intestinal sample. The spike-in control was used to estimate precision and recall at assembly, functional and taxonomic level of non-restricted communities. Functional and taxonomic annotation of genes expressed in the mouse microbiome provided insight in global, and family specific activities and illustrated the potential of this approach to study interactions and task division in unfamiliar microbial ecosystems.

Conclusions

A generic de novo assembly pipeline for metatransciptome data analysis was developed for microbial ecosystems, which can be applied for microbial metatranscriptome analysis in any chosen niche.

Background

High throughput metagenomics have revolutionized our knowledge of microbial communities such as those that populate the human and animal gastrointestinal (GI) tract. Complementing 16S ribosomal RNA gene-based compositional analyses, metagenome sequencing of these communities provided a broad description of the genetic content and relative abundance of individual members. The human enterotypes, for instance, have been defined using comparative metagenomic analysis of the human gut microbiomes of 39 individuals (Arumugam *et al.* 2011). Metagenomics, however, does not provide insights in the functional interactions within a complex microbial ecosystem and how these interactions may change in response to an ever-changing environment, including diet. RNA transcript profiling can serve as a proxy for ecosystem responses to environmental cues. Recent advances in massive parallel sequencing of mRNA-derived cDNA sequences (RNA-Seq) have led to an exponential increase of such transcriptome profiling studies. While most RNA-Seq based expression studies focus on a single species, in a number of cases RNA-Seq has been used to profile complex natural microbial communities in marine, soil and human and other mammalian GI tract environments (Frias-Lopez *et al.* 2008, Gilbert *et al.* 2008, Urich *et al.* 2008, Turnbaugh *et al.* 2010, Baldrian *et al.* 2011, Xiong *et al.* 2012, Leimena *et al.* 2013). Analysis of these large complex datasets poses a considerable bioinformatic challenge since natural microbial communities are usually non-restricted with a varying number of participating strains and species. A standard approach is to align the generated RNA-Seq reads to a set of closely related reference genomes or well-annotated metagenomes. This approach works well for well-studied microbial communities that have a nearly complete catalogue of reference genomes at a small evolutionary distance available (Leimena *et al.* 2013). However, at a larger evolutionary distance, the extensive sequence diversity at nucleotide level between the sample and the reference database significantly reduces the mapping efficiency of the alignment method and increases the probability of spurious assignments.

To overcome these problems a *de novo* assembly method can be used. *De novo* assembly of RNA-Seq reads into contigs increases the information content and therefore grants a more reliable annotation of the expressed genetic content of an unknown microbial community. Subsequently the newly assembled contigs can be directly used as target sequences in an mRNA-read mapping approach to obtain gene expression data. Currently a whole class of de Bruin graph based assemblers have been developed for *de novo* assembly of Illumina sequencing data (Nagarajan and Pop 2013, Yang and Smith 2013). Most of them have been designed to work with genomic data from a single species and assume that reads are uniformly sampled along a length of a single genome. As such they cannot efficiently deal with the existence of many co-linear genomic regions in the genomes of strains and species encountered in a non-restricted natural microbial community. Sequencing errors, exacerbated by authentic micro diversity caused by the coexistence of

syntenic strains of the same species in a community and strong sequence conservation of genes common to many species in the community thus can lead to assemblies with a relatively high rate of small contigs and to ambiguous chimeric contigs. To assess the correctness of *de novo* metagenome assemblies' several statistical methods are available. These methods involve testing each of the larger contigs for a uniform read coverage and for a uniform distribution of k-mer scores (Clark *et al.* 2013). Due to the limited size and strong variations in read coverage, however, these statistical analysis methods will not reliably work for RNA-seq derived contigs.

The microbiome of the human GI tract of healthy individuals fulfils a variety of beneficial functions for human health (O'Hara and Shanahan 2006). Numerous studies have linked an altered gut microbiome to disorders in energy and metabolic homeostasis including obesity and diabetes, as well as immune aberrations and excessive inflammation diseases (Backhed *et al.* 2004, Musso *et al.* 2010, Diamant *et al.* 2011). For a systematic study of the influence of diet, environmental factors and host genotype on the microbial diversity and function in the GI tract animal models provide an indispensable tool, and the mouse model has emerged as one of the preferred model systems. Mouse intestinal microbial communities have been mapped using 16S rRNA gene-based community profiling, and many microbial mouse intestinal commensals have been identified and categorized. Although the phylogenetic make-up of the GI tract microbial communities in human and mouse appear to be similar at phylum level, zooming in to genus and species resolution reveals a large difference in bacterial composition (Salzman *et al.* 2002, Ley *et al.* 2005). The large evolutionary distance of the microorganisms, combined with a strong bias towards human microbiome sequences in the current GI gene catalogs results, as we will show here, in low-resolution outcomes of the analysis of mouse metatranscriptome data with standard alignment methods (Xiong *et al.* 2012). This led us to design and implement a *de novo* assembly method that provides better gene assignment results, and evaluate sensitivity, reliability and validity of the method for the function analysis of complex metatranscriptome data. The generic *de novo* assembly method developed enabled the reliable functional profiling and taxonomic binning of unfamiliar microbial ecosystems and was validated by using metatranscriptome datasets of community-restricted samples, and samples obtained from the mouse GI, and included a spike-in human control sample.

Results and Discussion

Workflow, Samples and Data Filtering

A generalized metatranscriptome assembly and analysis pipeline was designed (Figure 5.1). Briefly the workflow consisted of filtering RNA-Seq reads for low quality and non-informative reads such as reads derived from ribosomal RNA (rRNA) followed by assembly of the remaining putative 'mRNA' reads into contigs, ORF calling and gene-function annotation, taxonomic classification and estimation of gene expression levels by using read-frequency analyses.

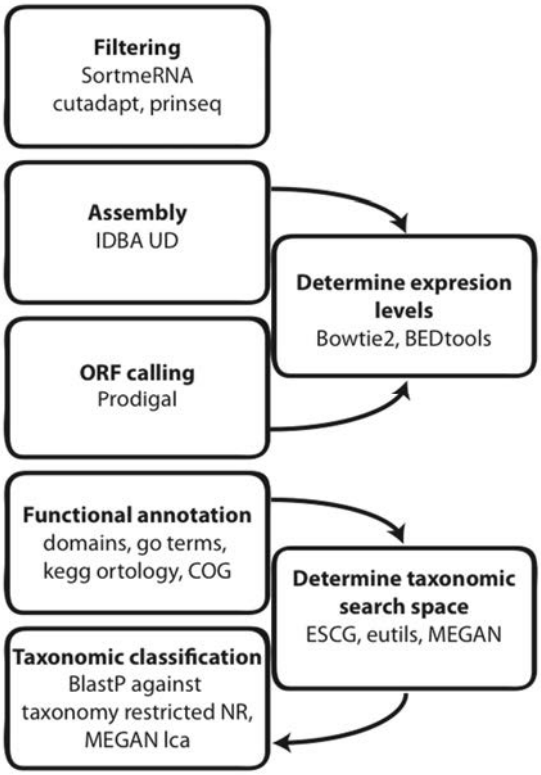


Figure 5.1 Metatranscriptome analysis workflow. Details of the programs used are described in the methods section

In order to validate the proposed pipeline, two single-end RNA-Seq metatranscriptome datasets were used. The first RNA-Seq dataset, obtained from an *in silico* generated mock community was used for a primary evaluation of the entire workflow. The second RNA-Seq dataset used to further validate the workflow resulted from an *in vivo* study of a restricted intestinal microbial community of limited complexity obtained by the colonization of inbred non-obese diabetic (NOD) mice with Altered Schaedler Flora (ASF) obtained from Xiong *et al.,* 2012 (Xiong *et al.* 2012). Both datasets were used to estimate precision and recall of the metatranscriptome assembly procedure at sequence, functional and taxonomic levels. Finally, the pipeline was used to generate functional and taxonomic community profiles of RNA-Seq datasets from natural microbial caecal communities

of four C57BL/6J mice that were fed a high-fat high protein diet (Schwarz *et al.* 2012). To estimate precision and recall of *de novo* assemblies obtained from this complex non-restricted mouse community, RNA-Seq data from a well-characterized sample from the human small intestine (Leimena *et al.* 2013) was used as spike-in control. Technical replicates were available for one of the mouse samples and for the human small intestine sample and included to establish the technical reproducibility of the procedure. Furthermore, PCR followed by Sanger sequencing of the amplified product confirmed the correctness of the sequence of a number of randomly selected mouse derived transcript assemblies.

An essential first step is data filtering (Figure 5.1). By far the largest fraction of the RNA content of a microbial cell is composed of ribosomal RNA (rRNA) and although standard mRNA enrichment protocols have been applied to all samples, rRNA derived reads remain present in significant amounts. SortmeRNA was employed to remove rRNA derived reads (Kopylova *et al.* 2012). Subsequently adapter sequences introduced during sequence-library preparations were removed with Cutadapt (Martin 2011), and nucleotides with low quality Phred scores were removed by trimming the reads with PrinSeqLite (Schmieder and Edwards 2011). Finally, only reads with a minimal length of 50 nucleotides and a minimal mean Phred quality score of 30 were kept and used as input for assembly. For all samples filtering details are provided in Table S 5.1.

*Table 5.1 Performance of de Bruin graph assemblers on metatranscriptome data. *mRNA reads obtained from the four individual mouse samples were merged with a human small intestine spike-in control sample before assembly.*

Assembly tool	Mock Community			Mixed Mouse and Human sample*		
	contigs	total length	Assembled reads (%)	contigs	total length	Assembled reads (%)
IDBA-UD	8,943	10,282,975	94.67	23,926	20,967,684	68.95
IDBA-HYBRID	9,003	10,282,926	92.82	23,921	20,965,864	68.83
IDBA-TRAN	9,739	11,908,188	91.68	31,708	30,735,244	62.86
Trinity	14,841	12,972,081	90.52	34,324	24,283,850	71.31
IDBA	9,069	10,245,202	82.73	24,413	20,645,762	53.99
SOAP	10,080	9,377,817	66.06	26,386	19,297,513	40.63
MetaVelvet	7,871	8,059,412	58.13	25,013	16,354,900	23.06
Velvet	11,705	6,579,171	22.47	23,806	11,618,506	9.76

Assembler performances

Using filtered mock and mouse caecal metatranscriptome datasets as input, the performance of various assemblers was tested; including the IDBA class of assemblers (Peng *et al.* 2010), Trinity (Haas *et al.* 2013), SOAP (Luo *et al.* 2012), (meta)Velvet (Zerbino and Birney 2008, Namiki *et al.* 2012) and Oases (Schulz *et al.* 2012) (Table 5.1). Although IDBA-MT (Leung *et al.* 2013) has been specifically designed for metatranscriptome assembly it was not included as it requires paired-end read information, and the datasets employed in this study comprised of single-end reads. The main criterion for assembler selection was the fraction of mRNA reads represented by the *de novo* assembled contigs. We found that both Trinity and IDBA-UD assemblies represented most of the reads for both the mock community and the C57BL/6J mouse cecum samples. However, with Trinity we found that many reads could be mapped to multiple contigs (75%) which hinders the determination of gene expression levels based on read abundance in the next step of the procedure due to non-specific mapping. IDBA-UD was therefore selected as the most appropriate assembler for single read metatranscriptome data and was subsequently used for assembly of all samples.

Table 5.2 Composition of the mock metatranscriptome RNA-Seq dataset and assembly results

Species	# filtered reads	Relative read abundance (%)	Assembled reads**	Sample ID
Streptococcus agalactiae	3,224,516	35.0	(98.0%)	SRR922307 (Richards *et al.* 2013)
Clostridium beijerinckii	1,586,292	17.2	(95.2%)	SRR988002 (Wang *et al.* 2013)
Pediococcus claussenii	1,371,187	14.9	(95.8%)	SRR647762 (Pittet *et al.* 2013)
Streptococcus pneumoniae	1,235,598	13.4	(97.2%)	SRR1009263
Enterococcus faecium	667,246	7.2	(89.8%)	SRR922448 (Chang *et al.* 2013)
Lactobacillus casei	500,000	5.4	(90.0%)	SRR616266
Streptococcus thermophilus	396,951	4.3	(87.6%)	SRR390316
Clostridium difficile	239,138	2.6	(62.7%)	ERR406251*
Overall	9,220,928		(94.7%)	

**Pre submission data taken from http://www.sanger.ac.uk/datasharing/*

***Assembly results presented were obtained by using the IDBA-UD assembler*

Metatranscriptome assembly of a mock community

An *in silico* mock metatranscriptome was built by merging RNA-Seq data of eight single species transcriptome profiling experiments downloaded from public repositories (Table 5.2). For each of the selected species a high quality reference genome was available and the selected RNA-Seq datasets were generated with the Illumina HTS platform using 68 to 107 sequencing cycles. From paired-end datasets an arbitrarily selected single-end dataset was selected. To capture some of the complexity of a true unrestricted community three closely related species from the genus *Streptococcus* where chosen and mixed with five species at a larger evolutionary distance. Furthermore, in this mock community the number of mRNA reads of each of the eight members was varied mimicking a high variation in species abundance (Table 5.2).

Since the mRNA reads in the mock community dataset originate from a specific set of known genomes the output of the *de novo* assembly workflow can be directly compared with results obtained from a standard alignment procedure. In total 8943 contigs were obtained with the IDBA-UD assembler and their precision at sequence level was assessed by aligning these sequences to the reference genomes. For 86% of these contigs an unambigous high quality full lenght sequence alignment without

insertions or deletions to a reference genome was obtained (Table S 5.1). The remaining contigs showed a varying degree of mainly small nucleotide mismatches at the 5'- and 3'-end of the contig sequences. Since most of these discrepancies were limited, the majority of these remaining contigs could still be used for further functional profiling (see below). Only 2% of the assemblies were recognized as a cross-species assembly. Manual inspection showed that these assemblies aligned to sequences that showed high levels of conservation among the most closely related species (Table S 5.1) suggesting that sequence micro-diversity does not have a major impact on the assembly performance, and that a taxonomic classification of metatranscriptome assemblies at genus level and above should be possible. mRNA reads that were not used in the assemblies (Table 5.2) were analysed by a direct alignment with the corresponding reference genomes. In many cases these reads mapped to genes with a small open reading frame (< 100 amino acids) and transcripts of low abundance (data not shown). On basis of our analyses, we estimated that a threefold coverage of a gene is required for at least a partial transcript assembly (Table S 5.1).

Functional expression profiles obtained by a direct genome alignment and by *de novo* assembly were compared. For both the assembled sequences and reference genomes ORFs were predicted using the Prodigal ORF prediction tool (Hyatt *et al.* 2010), and candidate coding sequences were annotated with a KEGG ontology using Kaas (Moriya *et al.* 2007). Gene expression values were obtained by mapping mRNA reads on the contigs and the genomes using Bowtie2 followed by extraction of gene expression levels using BEDTools (Quinlan and Hall 2010) (Figure 5.1). Expression data of proteins with an identical KEGG ontology identifier were lumped. Comparison of the functional profiles obtained from both procedures revealed a high congruency (Pearson correlation > 0.99) for the mock community as a whole as well as for each of its individual members (Figure 5.2). For functional prediction of assembly assigned reads we calculated a precision score of 0.97 with a recall of 0.94 under the assumption that such an assignment is a true positive if both methods agree, a false negative if the assembly method failed to assign a function and a false positive in case of a different assignment.

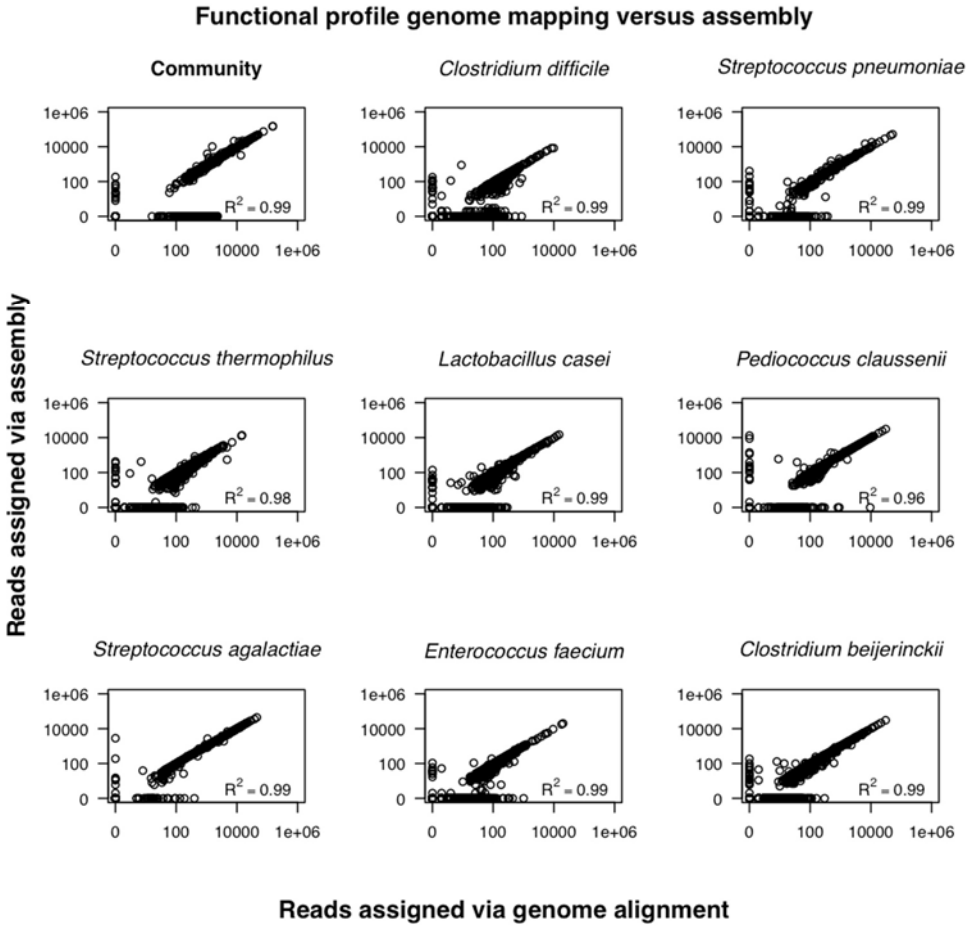


Figure 5.2 Comparison of functional profiles of an eight species mock community metatranscriptome. Reads assigned via direct genome alignment method (x-axis) and de novo assembly with IDBA-UD (y-axis). Each dot represents a specific KEGG orthologous function.

Metatranscriptome assembly of Altered Schaedler Flora from the intestines of a NOD mouse model.

For further validation of the pipeline a published RNA-seq dataset obtained from a defined *in vivo* mouse intestinal community was used (Xiong *et al.* 2012). In this study RNA-Seq data was obtained from four inbred non-obese diabetic (NOD) germ-free mice colonized with a defined mixture of eight commensal bacteria (Altered Schaedler Flora; ASF). Twelve caecum and colon samples were prepared using multiple RNA-extraction protocols and sequenced using the Illumina HTS platform. At the time of publication the complete set of ASF genomes was not known, and to bridge the evolutionary distance to known species a peptide-based alignment procedure (blastx) was used for functional profiling. Using this procedure the authors were able to link 16% of the sequence reads to a known bacterial gene. Recently, the draft genome sequences of all eight bacteria in the ASF community have been determined (Wannemuehler *et al.* 2014), providing the opportunity to validate the metatranscriptome assembly pipeline with a restricted *in vivo* RNA-Seq dataset from a mouse intestinal community.

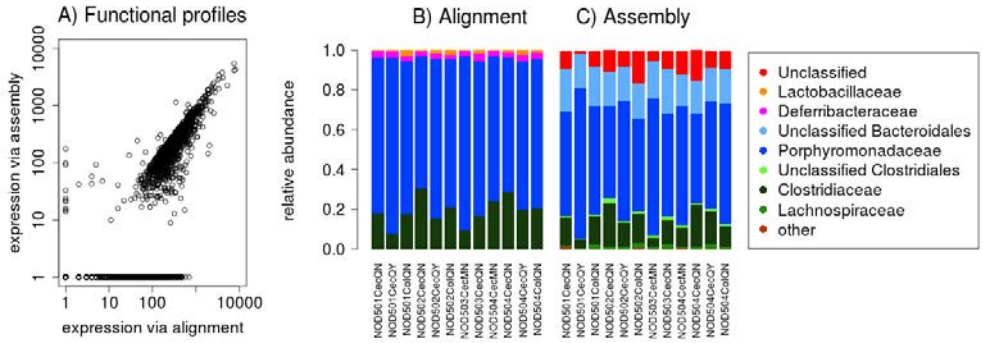


Figure 5.3 Comparison of functional and taxonomic profiles of the Altered Schaedler Flora from the intestine of a NOD mouse model (Xiong et al. 2012). A) Alignment vs assembly functional profiling; x-axis, direct genome alignment; y-axis, de novo assembly. Taxonomic profiles of mRNA reads obtained by direct genome mapping (B) and by using the de novo assembly method (C). Sample labels were taken from Xiong et al., 2012.

ASF reads were taken from the SRA repository (SRP012007) (Xiong *et al.* 2012), and approximately 5,4 million reads of high quality passed the filtering stage of the pipeline. 1,7 million of these reads could be directly mapped on the recently sequenced draft ASF genomes while 1,5 million mapped to the host genome. The origin of the remaining 2,2 million reads could not be established. A blastx search of those unmapped reads against the NR database did not return any significant results. All 5,4 million reads were used in an assembly with IDBA-UD to yield 7160 contigs. The majority (6638) of these contig could be accurately aligned with one of the eight ASF draft genomes and 256 contigs were derived from the host leaving 266 of the contigs (4%) unaccounted for. The ASF-mapped contigs captured 8688 ORFs that were functionally annotated and compared with a direct mRNA read

alignment to the annotated ASF draft genomes following the procedures described above. The result showed a strong correlation (Pearson correlation > 0.9) between the expression values obtained from the assembly and the direct-alignment method (Figure 5.3).
The precision and recall for functional annotation was 0.97 and 0.58 respectively. The precision was therefore comparable to what was observed for the mock community, whereas the recall is much lower, which is probably due to a much lower sequencing depth of ASF samples (Table S 5.1).
Additionally we explored the performance of a taxonomic classification of de novo assembled sequences. For a taxonomic classification of the ASF microbiome a two-step procedure was developed. First a subset of 103 proteins derived from essential single copy genes (ESCG) (Dupont *et al.* 2011) was identified and used to determine the boundaries of the taxonomic search space by aligning them with the entire NR protein database and further classification using the MEGAN processing pipeline (Huson *et al.* 2007) (see methods section for details). In order to mimic the lack of good reference genomes ASF derived protein sequences were excluded from this search. Sixty-seven of the 103 marker proteins were classified as proteins belonging to the order of *Bacteroidales* while 33 belonged to the *Clostridiales*. Only three marker proteins could not be classified at the order level and therefore this taxonomic rank was used to restrict the reference database. With the taxonomic boundaries set to the orders of *Clostridiales* and *Bacteroidales*, next non-ESCG proteins were classified by aligning them with all *Clostridiales* and *Bacteroidales* proteins present in the NR database and classification with MEGAN again while the ASF proteins were excluded from the search (Figure 5.3). The use of ESCG to restrict the search space drastically reduced the computational time to classify the full set of proteins. However the low abundant species of different orders may be missed. In this case the presence of *Lactobacillales* and *Defferibacterales,* which made up six percent of the mRNA reads in total, was not detected.
With large sets of mRNA reads assigned to a taxonomic rank via genome mapping we can estimate the precision and recall for the assignments via assembly. We assumed that the taxonomic association of a read obtained by a direct genome alignment to an ASF gene summarized in Figure 5.3 is true. If via the *de novo* assembly method a different taxonomic association was obtained it was considered to be a false positive association and in case such a read was not incorporated by the *de novo* assembly method or ended up in a taxonomically unclassified contig, it was considered to be a false negative association. For the correctness of a taxonomic classification of ASF proteins via the *de novo* assembly method we estimated a precision of 0.95 for taxonomic ranks down to genus level. The recall score for phylum to order level was 0.66 but was reduced to 0.38 for lower taxanomic ranks (genus & family) (Table S 5.1).
Xiong *et al.*, 2012, used a BlastX based alignment procedure to taxonomically assign mRNA reads to a known bacterial gene (Xiong *et al.* 2012). Due to the small read length this method can give rise to many ambiguous assignments. This can be

substantially improved by using a *de novo* assembly approach. Relative abundance of community members could be estimated with relative accuracy and were in very good agreement with those obtained by direct genome mapping (Figure 5.3).

Metatranscriptome assembly of mouse cecum samples and validation of assemblies by PCR and Sanger sequencing

High protein diets are suggested as effective weight loss regimes and therefore would fit in a successful strategy to achieve a long-term weight loss for a positive effect on health and to decrease obesity and associated metabolic disorders (Westerterp-Plantenga *et al.* 2012). To study the effects of such diets on a GI tract community, four C57BL/6J mice were fed on a high-fat high-protein diet for 12 weeks (Schwarz *et al.* 2012). At the end of the intervention, caecal content was obtained, and the microbial activity present in these samples was analysed by RNA-Seq. Initially, analyses of the mRNA reads employed a previously developed direct alignment approach (Leimena *et al.* 2013). However, due to the large evolutionary distance between mouse and human microbiome sequences, and the much more limited availability of mouse microbiota associated reference genomes, this procedure resulted in functional and taxonomic information of low resolution (Figure S 5.1). To increase the functional and taxonomic resolution the here developed assembly workflow was applied. To monitor precision and recall in this complex and mostly unfamiliar microbial dataset, an RNA-Seq dataset from a well-characterized human small intestinal community (Leimena *et al.* 2013) was used as a spike-in control and co-assembled with the mouse RNA-Seq data. This led to a *de novo* assembly of 24077 contigs and allowed for the prediction of 36012 partial and full length ORFs within these contigs. Of these ORFs, 25897 were solely derived from the mouse datasets, whereas 9707 were assembled exclusively from the human dataset. For the mouse-derived sequences virtually identical results were obtained when the mouse RNA-Seq data was separately assembled (results not shown). A total of 407 hybrid ORFs were assembled consisting of reads obtained from both data sources.

For taxonomic classification the here developed two-step procedure was used. From the full set of proteins ESCG were identified as described above and used to limit the taxonomic search spaces. Alignment of the full set of translated proteins against a thus restricted NR database suggested low levels of sequence identity between mouse GI bacterial proteins and NR proteins (Figure 5.4). This notion was further confirmed in the taxonomic classification of proteins from the two environments, where 93% of the human small intestine protein sequences could be classified at family level, while only 48% of the mouse cecum bacterial protein sequences could be classified at this rank.

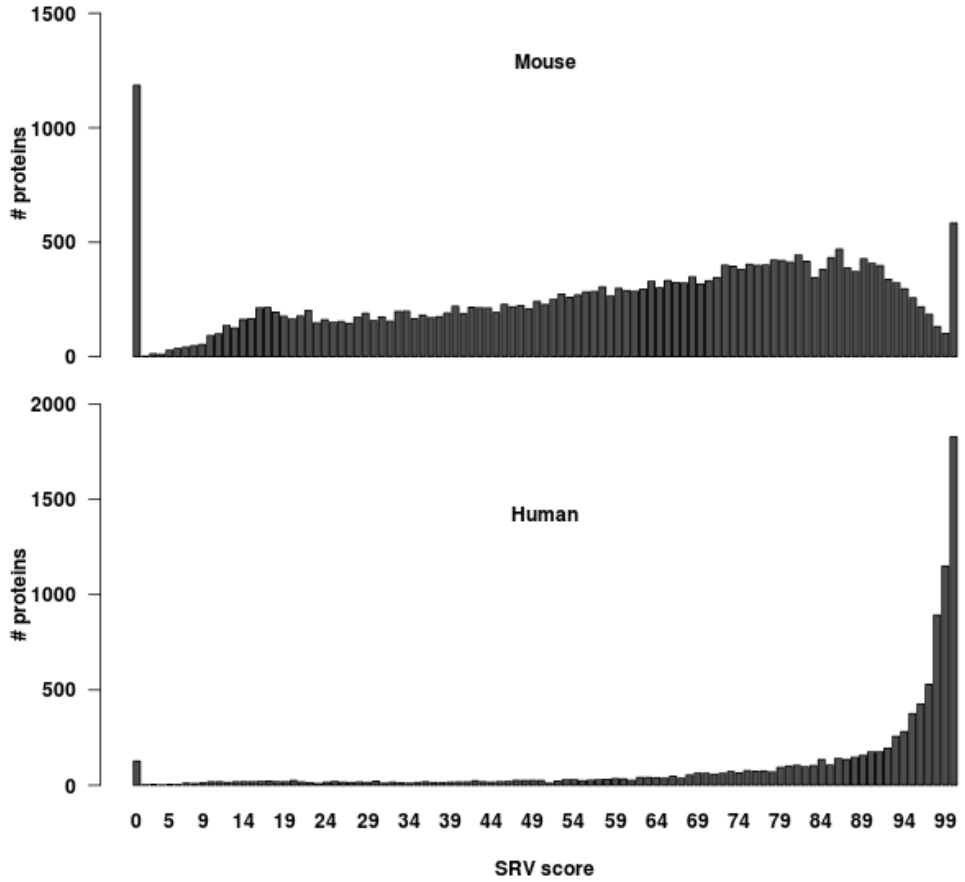


Figure 5.4 Similarity score distributions of predicted Mouse and Human microbial community proteins to known proteins. Translated proteins were aligned to the NCBI nr protein database and binned according to their SRV score. The SRV score represents the bit-score of the best hit divided by the maximum obtainable bit-score (Lerat et al. 2003).

Sanger sequencing of genomic sequences was used to further validate the correctness of the mouse metatranscriptome assembly. Twelve transcript assemblies of at least 800 nucleotides were randomly selected for having high and low read coverages. For each of these assemblies two sets of specific primer pairs were designed, and in each case two partially overlapping fragments of the correct size could be amplified using caecal microbial DNA of mouse 2 as template (Table S 5.2). For ten assemblies Sanger sequencing of the amplified DNA returned nucleotide sequences that aligned with 98%-99% identity with the corresponding assembly and thus confirmed the existence of corresponding DNA sequences in the bacterial metagenome. Two PCR products were shown to be a mixture of amplicons originating from isogenic genes.

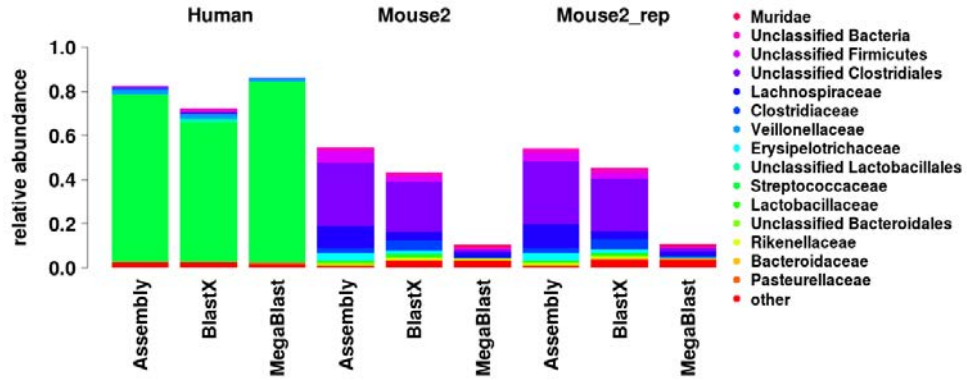


Figure 5.5 Taxonomic composition of the transcriptome using three different methods. Reads for three samples were assigned to family level using de novo assembly, BlastX and MEGABLAST.

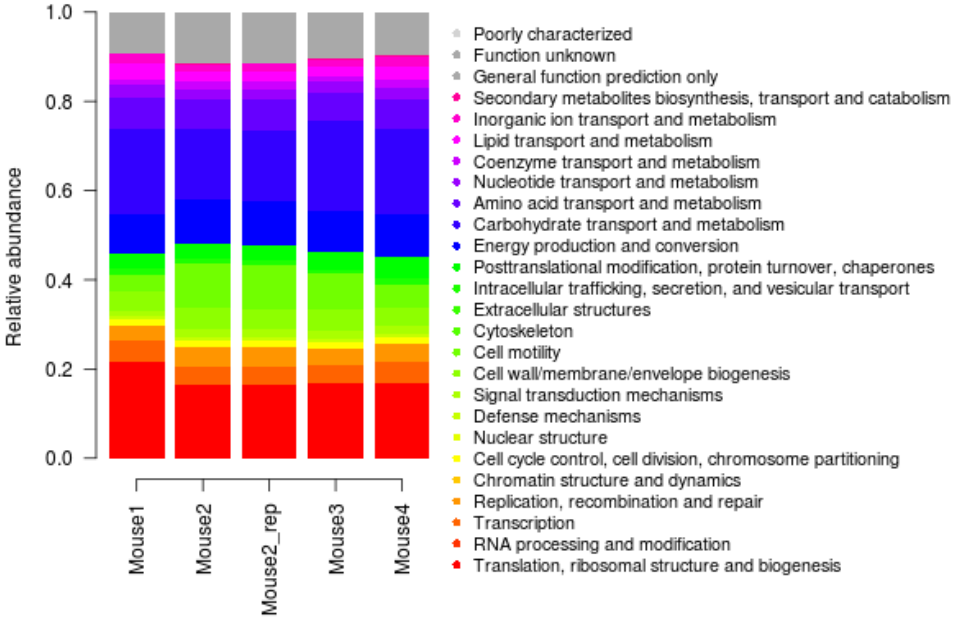


Figure 5.6 Distribution of COG functional categories of the mouse caecal metatranscriptome.

For one mouse sample and for the human small intestine sample a technical replicate of higher sequence depth was available, and these were employed to establish the technical reproducibility of the procedure and to assess the level of noisiness in the data analysis pipeline. In the mouse duplicate sample, essentially the same results were obtained for transcripts of high abundance even though sequence depth between the two replicates differed more than 20 fold. Moreover, transcripts of low abundance results were consistent with an increase in noise, and the human small intestine sample a similar pattern was obtained (Figure S 5.2).

Of the RNA-seq mRNA reads derived from the human small intestine sample, 85% was captured in an assembly whereas this fraction was 61-71% for the mRNA reads derived from the four individual mouse cecum samples. The fraction of the reads that was assigned to an ORF was 80% and 45-55% for the human small intestine sample and mouse cecum samples, respectively (Figure 5.5) (Figure S 5.1).

Functional analysis of the mouse caecal community

Using the pipeline, metatranscriptome profiles of caecal microbial communities were obtained from four C57BL/6J mice fed on a high-fat high protein diet (Schwarz *et al.* 2012). For those four individual communities, the relative expression of COG functional categories was compared. Proteins were labelled according to the COG ontology system (Tatusov *et al.* 1997) and expression levels of proteins belonging to the same category were extracted and lumped (Figure 5.6). Although the taxonomic profiles of the caecum samples of the four individual mice showed clear differences in relative abundance and distribution of phylotypes the overall COG activity profiles of the four communities were highly similar (0.984 -0.999 Pearson correlation, Figure 5.6).

To determine whether bacteria from different families fulfilled different roles, homologous proteins of the samples from the C57BL/6J mouse cecum belonging to the four bacterial families with highest numbers of assigned transcripts, namely the *Clostridiaceae*, *Lachnospiraceae*, *Erysipelotrichaceae* and *Lactobacillaceae*, were annotated using KO identifiers (Moriya *et al.* 2007) and their cumulative family-specific expression levels were mapped using iPATH (Yamada *et al.* 2011) (Figure S 5.3). For the four families the results showed distinct differences in their metabolic profile. For example *Lachnospiraceae* seemed to be active in propionate formation and vitamin B12 biosynthesis, while *Erysipelotrichaceae* appeared to be active in butyrate formation (Figure 5.7). The *Lactobacillaceae* metabolic activity was found to be mainly oriented towards the production of acetate and lactate, a well-established metabolic feature of this bacterial family. Finally, members of the *Clostridiaceae* family did not display a very clear metabolic activity pattern but compared to the other three families appeared to consistently express amino acid degradation pathways at a higher level (Figure S 5.3). Further studies should give insight into dietary effects on community composition and microbial metabolic activity.

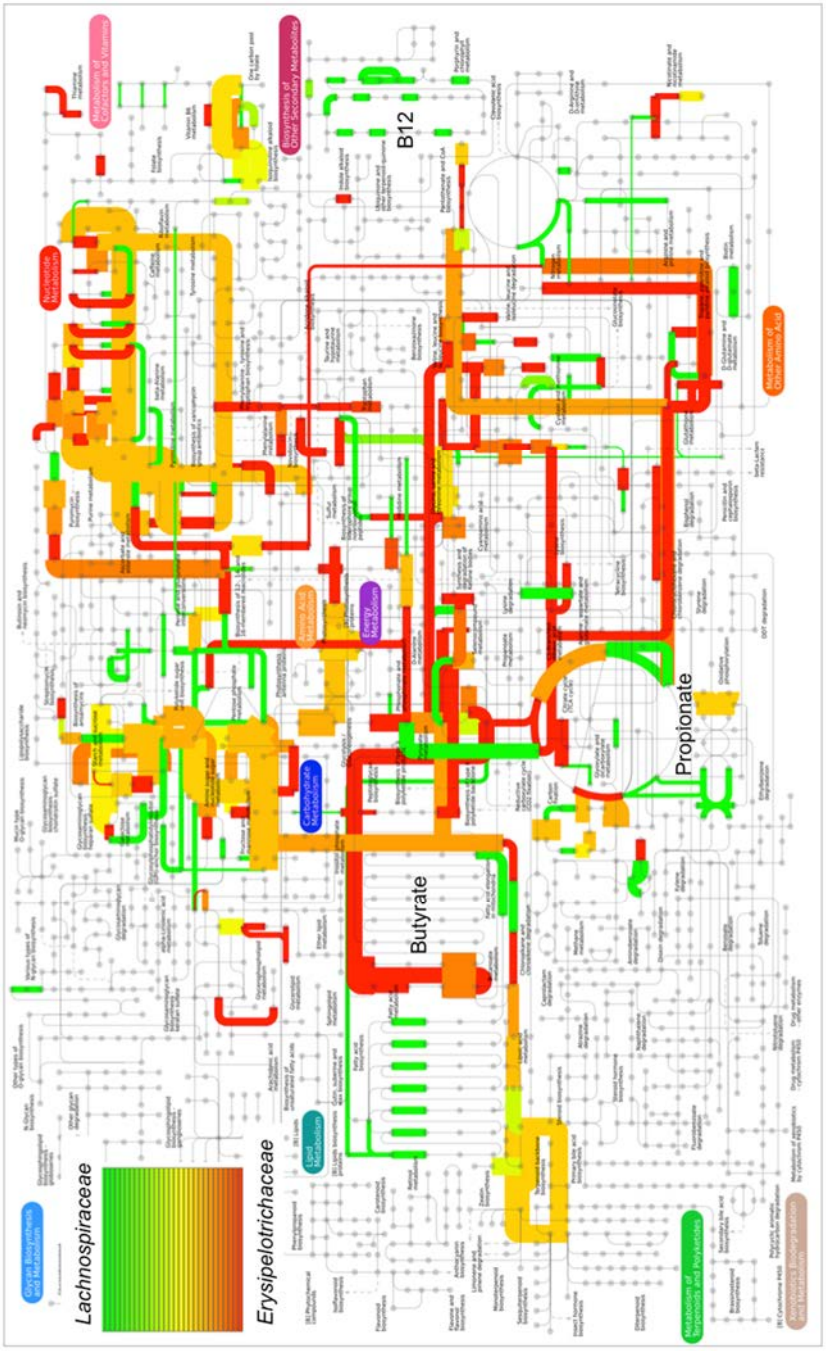


Figure 5.7 Metabolic pathways mapping of Lachnospiraceae and Erysipelotrichaceae expression profiles. Relative contribution of each family (green Lachnospiraceae, red Erysipelotrichaceae) are colour scaled. Line-width indicates the total amount of reads mapped to the corresponding KEGG ortholog (log scaled).

Conclusions

We have developed a validated *de novo* metatranscriptome assembly pipeline suitable for analysis of transcriptome data from unfamiliar complex ecosystems. With this *de novo* metatranscriptome assembly pipeline mRNA reads obtained from RNA-Seq can be assigned to a protein function and a taxonomic rank with high precision. When dealing with familiar ecosystems composed of species that are well represented by complete catalogs of reference sequences at a small evolutionary distance, an assembly strategy will be less efficient due to the required minimal transcript coverage for assembly. However, when dealing with species that are not well represented in the genome databases, the *de novo* assembly pipeline outperforms direct read alignment methods due to a significant increase in information content. Nevertheless taxonomic classification of activity profiles remains a major issue for unfamiliar ecosystems harbouring many novel species at variable evolutionary distances.

Acknowledgements

Functional annotation using interpro was carried out on the Dutch national e-infrastructure with the support of SURF Foundation. This work was (co)financed by the Netherlands Consortium for Systems Biology (NCSB), which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

Material & methods

RNA extraction, mRNA enrichment, cDNA synthesis and illumina sequencing

The cecal intestinal content was collected from four mice on a high fat high protein diet at 10 weeks during the dietary intervention study previously described by Schwarz *et al.,* 2013, snap frozen in liquid nitrogen and stored at -80°C. RNA was extracted from 0.1-0.2 grams of cecal content (Schwarz *et al.* 2012). The content was re-suspended in 500 µL ice-cold TE buffer (Tris-HCL pH 7.6, EDTA pH 8.0). Total RNA was obtained via the Macaloid-based RNA isolation protocol (Zoetendal *et al.* 2006, Leimena *et al.* 2013) with in addition the use of Phase Lock Gel heavy tubes (5 Prime GmbH, Germany) during the phase separation. The RNA purification was done with the RNAeasy mini kit (Qiagen, USA), including an on-column DNAseI (Roche, Germany) treatment [38]. The total RNA was eluted in 30 µL ice-cold TE buffer, and the RNA quantity and quality were assessed using a NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, USA) and Experion RNA Stdsens analysis kit (Biorad Laboratories Inc., USA), respectively. mRNA enrichment was performed by using the mRNA enrichment kit MICROBExpress™ (Ambion, Applied Biosystems, The Netherlands) using the manufacturer's protocol. RNA quantity and quality were assessed as described above to determine the efficiency of the mRNA enrichment. Double stranded cDNA was synthesized from one µg of the enriched mRNA sample with the SuperScript® Double-Stranded cDNA Synthesis kit (Invitrogen, The Netherlands), with addition of SuperScript® III Reverse Transcriptase (Invitrogen, The Netherlands) and random priming using random hexamers (Invitrogen, The Netherlands) (Yoder-Himes *et al.* 2009, Leimena *et al.* 2012, Leimena *et al.* 2013). To remove the RNA an RNAse A (Roche, Germany) treatment was preformed, followed by phenol-chloroform extraction of the cDNA and ethanol precipitation. The product was checked on gel and 3 to 8 µg of cDNA was sent for sequencing (GATC Biotech, Germany). Single read Illumina libraries were prepared from the double-stranded cDNA according to the ChiP protocol with insert sizes between 200-300bp. Sequencing was done with an Illumina Hiseq2000, and each sequence library was barcoded and sequenced at 5pM concentration using the single-end protocol. In total the amount of reads was between 700k and 177 M per sample. The data set supporting the results of this article is available in the NCBI small reads archive (sra) repository, under accession number SRX611064.

PCR and sequencing of a representative selection of assembled contigs

Amplicons targeting a representative selection of assembled contigs were generated by PCR. Primer sets were designed with the NCBI's primer blast (Ye *et al.* 2012) using an optimal melting temperature between 59-61 °C with the potential to amplify fragments of around 800 bp (Table S 5.2). Amplicons were generated in two runs. In the first PCR run amplicons from two primer combinations were generated in multiplex reactions, and checked for expected amplicon size. PCRs were performed in a total volume of 25 µl with the FastStart Taq DNA polymerase (Roche), a

denaturation of 95 °C for 30 seconds an annealing temperature of 60 °C for 40 seconds, and elongation for 30 seconds at 72 °C, and run for 30 cycles, where the size of the PCR products was confirmed by gel electrophoresis. In the second PCR only the two outer primers were used resulting in the largest obtainable amplicon per contig. The Phusion Hot Start II High-Fidelity DNA polymerase (Thermo Scientific) was used for amplification in a total volume of 50 µl during 35 cycles consisting of denaturation at 98 °C for 10 s, annealing at 60 °C for 20 s and 72 °C for 50 s for elongation. The size of the PCR products was confirmed by gel electrophoresis and sent for Sanger sequencing from both the forward and reverse primer.

Metatranscriptome analysis

Data filtering

The data was filtered for ribosomal RNA sequences, adapter sequences and low-quality reads using dedicated tools. SortMeRNA (version 1.2) (Kopylova *et al.* 2012) was used to rapidly filter out rRNA sequences using the precompiled databases for eukaryotes, bacteria and archaea. Truseq adapter sequences were removed from the reads with cutadapt (Martin 2011). Initial results showed a high bias of adenines in the trimmed sequences and therefore all trimmed sequences were discarded. The remaining reads where quality (phred >30) and poly A tail edge trimmed using PRINSEQ (lite-version)(Schmieder and Edwards 2011). Reads smaller than 50 nucleotides were discarded.

Assembly, annotation and classification

Assemblies were performed using the assemblers' default setting. When required kmer size was set to 31. ORF calling was performed using prodigal 2.60 with the meta procedure (Hyatt *et al.* 2010). Functional annotation was performed using Interproscan5 with standard settings for all potential output and the KEGG automated annotation server using the SBH method against the default reference set (Quevillon *et al.* 2005, Moriya *et al.* 2007). COG annotation was performed by rpsblast (v2.2015) against the NCBI COG database (2-2-2011) with a minimum E-value of 0.0001. Reads were mapped using bowtie2 (Langmead and Salzberg 2012) and expression levels for each predicted protein were extracted using samtools and BEDtools (Li *et al.* 2009, Quinlan and Hall 2010, Langmead and Salzberg 2012). The first step in the taxonomic classification of the predicted ORFs was identifying all the essential single copy genes and using these to determine the taxonomic search space. Proteins were aligned against the NR database using BLASTp followed by MEGAN classification (Altschul *et al.* 1997, Huson *et al.* 2011). A list of GI protein identifiers belonging to the ESCG identified bacterial orders was retrieved via an eUtils query (Nadkarni and Parikh 2012). The second step was to classify the remaining proteins by aligning them against a GI restricted database followed by MEGAN classification. SRV scores were calculated by dividing the bitscore of the best alignment by the bitscore of a self-alignment. Tetra nucleotide occrence

regression coefficients of the mock community members were calculated using jspecies (Richards *et al.* 2013).

Supplemental files

Supplemental Figures

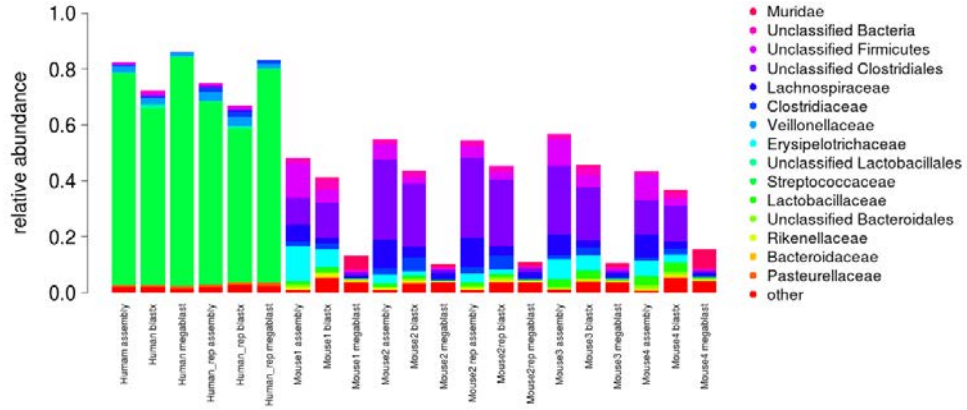


Figure S 5.1 Read assignment of human small intestine- and mouse caecum-derived metatranscriptome samples using alignment and assembly procedures
Description: Taxonomy profiles for all mouse and human samples using BlastX, megaBlast and assembly strategies.

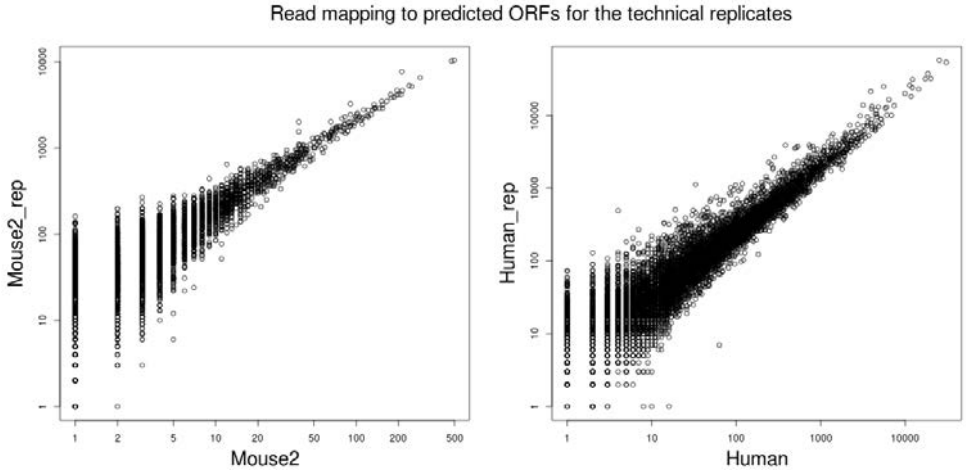
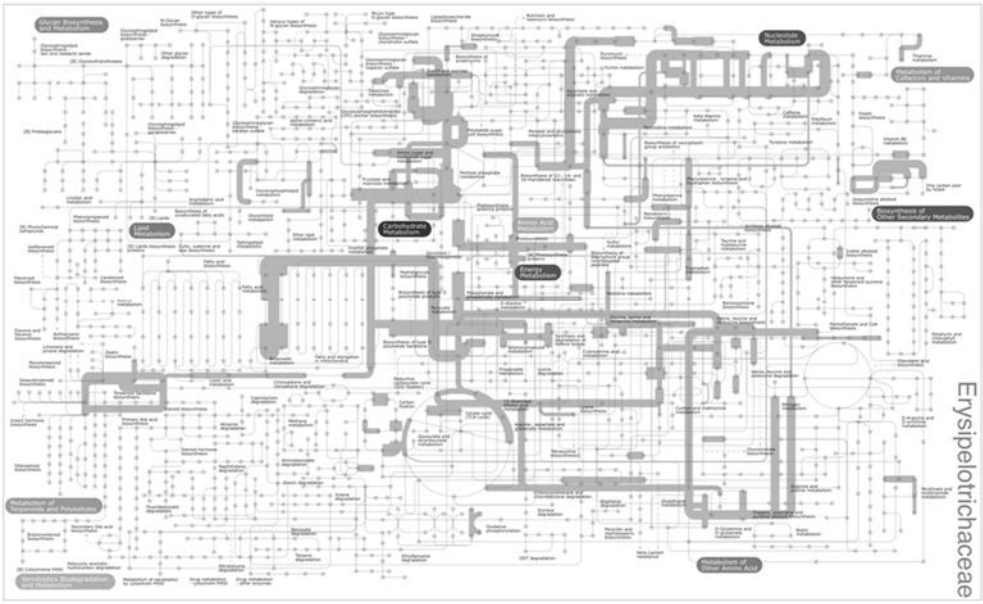
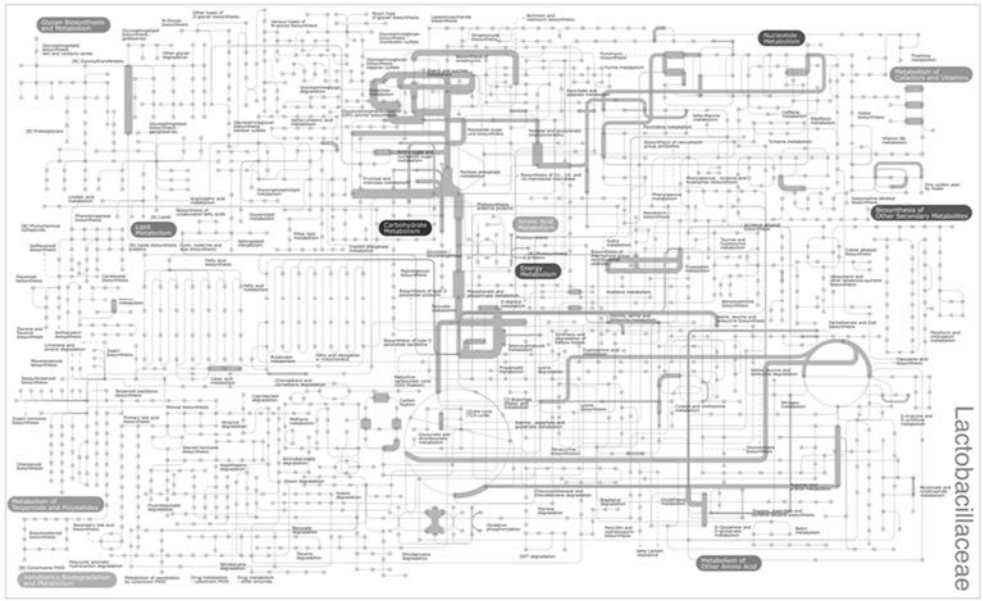


Figure S 5.2 Comparison of technical replicates


Erysipelotrichaceae


Lactobacillaceae

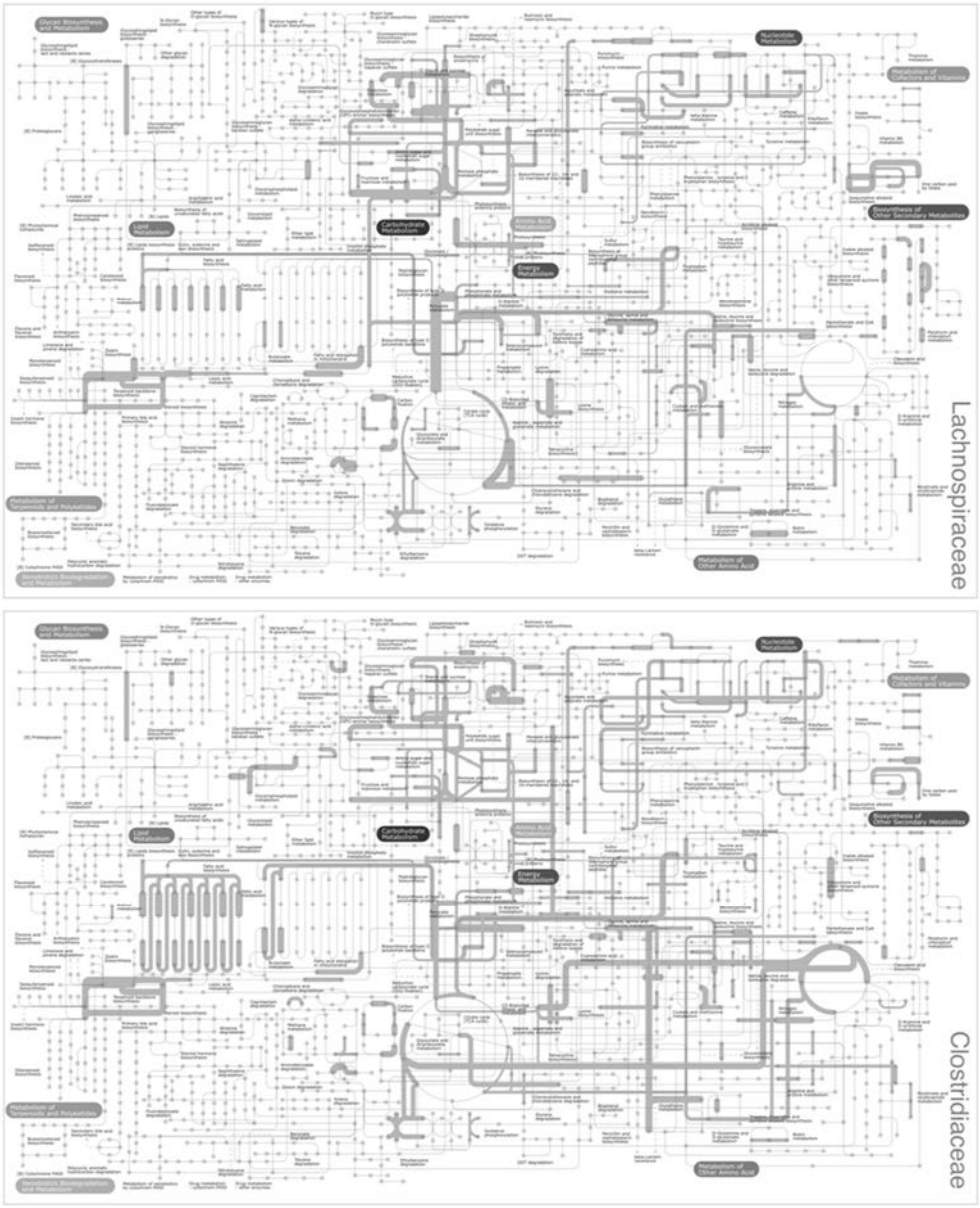


Figure S 5.3 iPATH mapping of the 4 most abundant bacterial families found in the metatranscriptome.

Supplemental Tables

	Sample name	sra code	total	Bacterial 16S	Bacterial 23S	Bacterial/Archeal	Archeal-16S	Archeal-23S	Eukarvotic-18S	Eukarvotic-28S	Eukarvotic-5.8S	adapter	quality	mRNA
Mock	*Streptococcus agalactiae*	SRR922307	166882	172465	772929	74490	301	497	58823	77622	6521	68860	498192	644903
	Clostridium beijerinckii	SRR988002	339168	11556	97858	42351	4	3	8347	30570	47	8835	199140	317258
	Pediococcus claussenii	SRR647762	957888	112830	561321	2088	1416	515	189	2194	73	2843	145686	137118
	Streptococcus pneumoniae	SRR1009263	114795	50313	605041	193	6	2	29	182	3	3176	935852	988478
	Enterococcus faecium	SRR922448	667772	1562	2164	177	3	3	319	1036	3	0	0	667246
	Lactobacillus casei	SRR616266	117911	801115	690348	4962	3151	65	274	3396	76	108117	0	397726
	Streptococcus thermophilus	SRR390316	671339	38016	56412	13	18	2	146	187	0	0	267392	635121
	Clostridium difficile	ERR406251	250880	711404	137409	256	913	161	203	2723	288	481	139438	278845
	Mock community		205475	116769	845129	124530	5812	1248	68330	117910	7011	116536	528914	102506
ASF	NOD501CecQN	SRR452973	146309	558467	382665	9789	524	0	3818	20935	633	394793	15533	75940
	NOD501CecQY	SRR452974	224156	109623	362296	2972	56	6	4046	16263	410	844244	125439	776213
	NOD501ColQN	SRR452975	133421	522974	689690	14378	40	8	9503	36019	3267	20311	5640	32381

	NOD502CecQN	SRR452976	162729	618126	438733	133991	107209	94310	142986	219629	124521
	NOD502CecQY	SRR452977	134777	57414	76856	26941	92521	21338	12296	53489	27096
	NOD502ColQN	SRR452978	152667	154747	103052	317139	201256	237550	50081	681617	475488
	NOD503CecQN	SRR452979	1672	1399	1443	966	1401	812	3223	603	953
	NOD503CecMN	SRR452980	73006	46428	38656	20425	48783	18723	69974	19281	34008
	NOD504CecQN	SRR452981	13127	11844	11230	3106	9608	4198	35275	5819	6742
	NOD504CecQY	SRR452982	6	3	6	9	11	17	15	11	0
	NOD504ColQN	SRR452983	176	129	86	246	175	322	131	48	835
	NOD504CecMN	SRR452984	31253	14473	4546	4772	31471	5166	10565	2380	17007
	ASF total		3898	4061	59780	2407	7793	148772	2768	584	1381	92	118316	442501	16782	456041	649340	535622
Mouse cecum	Mouse1	SRR1425537	128971	959811	123400	559874	171181	620969	314460	380375	925720	769221	347665	543456	428481	392190	673475	638412
	Mouse2	SRR1425604	207461	226412	262251	118089	343574	520950	266087	425134	242585	709267	334360	700322	396030	126448	818204	256319
	Mouse2_rep	SRR1425614	495251	84499	649268	29457	281015	776429	149762	126514	57680	545020	1228	222	2715	104	740	317539
	Mouse3	SRR1425615	674148	875596	174013	493467	317539	210998	134335	240637	99864	317539	577958	336624	558189	256319	638412	177126
	Mouse4	SRR1425616	864	431	4999	275	1381	163	349	8268	422	584	3898	4061	59780	2407	7793	148772

Human small intestine	HSIC	SRR870697	895108	110663	564090	2389	651	235	8167	7752	267	16	489315	169475
	HSIC_rep	SRR886287	297092	275570	196552	166896	7392	959	37310	64072	12142	930066	692478	388499

Table S 5.1 Results of filtering of RNA-Seq datasets for rRNA reads, adapter sequences and low-quality reads

Filtering: Ribosomal RNA, adapter and quality filtering results.

Mock_TNF_hybrids: Number of shared assemblies and tetranucleotide frequency ocrrence regression coefficient between mock members.

Coverage_Assembly: Distribution of assembled proteins based on transcriptome coverage.

Mock_assembly ouput: Precision and recall for functional assignments

NOD_mouse_taxonomy_output: Precision and recall for taxonomy assignments

Target primer pair	Primer Sequence Forward	Primer Sequence Reverse	location Forward	Location Reverse	band size	cross-primer
contig-100_0 P4	TGTGAACTCGG GAACTTCGG	**AGATTTCCGGC ACACTGGAG**	22443	23106	663	1114
contig-100_0 P14	**AGATTGCGTCA CGGGACATT**	ACCGGCTTTGCT CCACAATA	21992	22558	566	115
contig-100_2 P2	**TCCAGGCTGGT CGGTATGTA**	CCTGCAACACT GTTCTGTGC	21183	22021	838	
contig-100_2 P3	CACAGAACAGT GTTGCAGGC	**TATGCACTCCCG GGTTTTCC**	22022	22466	444	**1283**
contig-100_5 P2	**CTGAGGAGTA TGGCGCAGAG**	CTCATGCATAAC GTGTGCGG	8888	9859	971	273
contig-100_5 P3	GTGAGATCATG CCGGAGGAG	**GCATTTGTATG CGCCGGATT**	9586	10132	546	1244
contig-100_122 P1	**TTCCGGACTCC CATCTCCAT**	CTATGCGGGAA AATGCACCG	4702	5170	468	69
contig-100_122 P4	TCCCTGTGGGC ATCCAGATA	**ACTGCCCTGTG AGTTGATCG**	5101	5954	853	**1252**
contig-100_3 P2	TGTCATCCCCC GGATACCTT	**TGTGACACGTC GGACAAGAG**	3129	3534	405	**1180**
contig-100_3 P5	**CGAGAATCCTA CGGCCTTCC**	AAGGTATCCGG GGGATGACA	2354	3148	794	

contig-100_1 P22	**TCCAGGCTGGT CGGTATGTA**	CCTGCAACACT GTTCTGTGC	848	1378	530	112
contig-100_1 P23	AAGGTACCGCA CGGCAATTA	**ATTGTTCAGTA CCAGCGCCA**	1266	1809	543	961
contig-100_3111 P1	**TCGTAGCGACC GGTGAAAAA**	TTCAGTGCGGG AGCATCATT	191	1163	972	225
contig-100_3111 P2&P20	GGCAAGTAGG TCAAGCCCAT	**CTGCGATCGGA CAGCAACTA**	938	1787	849	**1596**
contig-100_4904 P1	**GCATTGTTCTG CGCTGTAGG**	TCCTTCGGAATG TTACCGCC	28	842	814	**271**
contig-100_4904 P9	CCAAACAGCAA GGGTACGGA	**GAGTCAACCGG GTGGAAAGA**	571	962	**391**	934
contig-100_217 P8	**ATTCTGGCAGA GCAACCGAA**	TGTCTGGCGAG CTGTTCATT	2265	2793	528	193
contig-100_217 P9	CTCCGGGTGTT TTCCGATGA	**AGATCCTTTGT GCGGGAGTG**	2600	3470	870	1205
contig-100_2390 P1	**GTTATCTGGGG CAACGGTCA**	ACAATGTTGTTC CGGCAAGC	117	940	823	310
contig-100_2390 P24	TCCCATGCCTT CTTCACCTG	**AAGTCACGCCG GGCTTTTTA**	630	1385	755	**1268**
contig-100_5984 P2	**TGCCCGATCGT TAAGGGTTC**	CCACCCTCGTCT TTCGTCAA	133	590	457	131
contig-100_5984 P9	CATCGGTGTTC TGCTTCGTG	**GCAAACGTAAG ACCCTGCTC**	459	764	305	631
contig-100_3802 P2	**TCTGCACGTTA TCTCCAGGC**	AGAGCGGAGCA GACATCAAC	122	676	554	163
contig-100_3802 P10	TGTTGCAACTG TTCCACGTC	**ATGCCACAGAC AAGAGAGCA**	513	887	*374*	*765*

Table S 5.2 PCR primers and Sanger sequencing output of amplicons. Table consists of primer pair names, the corresponding location and the expected fragment lengths. In bold are the fragments not found in the first PCR, underlined are fragments found in the second PCR that were send for sequencing. Primers in bold were used for amplification in the second PCR.

Chapter 6

Linking the fate of dietary fibres in the murine caecum to microbial transcriptome patterns

Floor Hugenholtz*, Katja Lange*, Mark Davids, Peter J. Schaap, Michael Müller, Guido JEJ Hooiveld, Michiel Kleerebezem and Hauke Smidt

*These authors contributed equally

Abstract

The microbiota of the gastrointestinal tract plays a key role in the degradation of food components that escape digestion by host enzymes. Complex metabolic networks of interacting microbes in the gastrointestinal tract of humans and other mammals yield a wide range of metabolites of which the short chain fatty acids (SCFA), in particular butyrate, acetate, and propionate, are the most abundant products of carbohydrate fermentation. Here we studied the quantitative interactions between diet, microbiota and host and modelled the multivariate data using Systems Biology approaches.

The experiments targeted the caecum of conventionally raised mice that were fed different fibre-containing diets. Microbiota composition was assessed using phylogenetic microarray technology, and was complemented with metatranscriptome, metabolome and host mucosal tissue transcriptome data. Relative abundance of butyrate producing bacteria correlated with host genes involved in energy metabolism and affecting the immune system. Moreover the metatranscriptome revealed distinct activities of bacterial families in the fermentation of fibres into SCFA. The *Bacteroidaceae*, *Porphyromonadaceae*, *Verrucomicrobiaceae, Bifidobacteriaceae*, *Lachnospiraceae*, *Clostridiaceae, Eubacteriaceae*, several Bacilli families, *Ruminococcaceae* and *Erysipelotrichaceae* all took part in fibre utilization, expressing genes encoding glycosidases and/or sugar transport systems. All families expressed in different ratios genes that code for enzymes involved in the production of SCFA. Overall different dietary fibres induce distinct changes in the caecal microbiota, their functional activities and SCFA production with profound effects on host metabolic and immune function in the caecum.

Introduction

A diet rich in fibre has a beneficial health effect by promoting gastrointestinal homeostasis, and decreasing risk for obesity, cancer and metabolic disorders. The enzyme repertoire of humans and mammals is commonly not able to catabolize non-starch polysaccharides, like those derived from plant cell walls, resistant starches, and oligosaccharides, whereas a broad range of intestinal bacteria can ferment these dietary fibres. The intestinal microbiota is diverse, dense, metabolically active, and largely saccharolytic (Qin *et al.* 2010, Tasse *et al.* 2010, Flint *et al.* 2012b). Most members of the large intestinal microbiota are not depending on the availability of simple sugars, but are rather able to derive carbon and energy from the breakdown of a range of complex carbohydrates, sometimes involving single members of the ecosystem, but often requiring a concerted community effort.

The bulk of the fibre fermentation takes place in the large intestine, where the fibres are the main energy source for the microbiota. Dietary fibre is fermented into a range of metabolites, of which the short chain fatty acids (SCFA) acetate, propionate and butyrate are the most abundant (Cummings *et al.* 2001, Flint *et al.* 2008, Bird *et al.* 2010, Flint *et al.* 2012a). The SCFA production pattern is dependent on diet composition, transit time in the small and large intestine, intestinal pH, and the microbial species and their relative abundance within the microbiota (Mortensen and Clausen 1996, Ouwehand *et al.* 2005, Macfarlane *et al.* 2006). SCFA are taken up by the mucosa, and have been reported to affect numerous processes related to the immune system (Al-Lahham *et al.* 2011) and energy metabolism (Donohoe *et al.* 2011). Additionally, the intestinal microbiota exerts a major influence on host physiology, including the tuning of the host's immune and metabolic state (Turnbaugh and Gordon 2009). SCFA are readily absorbed by the colonic epithelium where butyrate serves as the main energy source, although in smaller amounts acetate and propionate can also be metabolised (Cook and Sellin 1998). After absorption and metabolic conversion, the remainder of the SCFA enter the portal blood and are processed by the liver, the central metabolic organ in the body (Bloemen *et al.* 2009). Intestine-derived acetate is incorporated in fatty acid synthesis in the liver and in part transferred to the peripheral tissue, whereas propionate induces lipogenesis and gluconeogenesis in liver (Al-Lahham, 2010).

The amount and ratio of SCFA can be affected by specific types of fibre (Flint *et al.* 2008, Van den Abbeele *et al.* 2011); **Chapter 4**). The relative abundance of specific groups of bacteria is known to increase when certain types of fibres are consumed. For example, when resistant starch is consumed, the relative abundance of species of *Clostridium* cluster IV that includes butyrate producers is increased, while consumption of fructooligosaccharides (FOS) and inulin tend to expand bifidobacterial populations (Barboza *et al.* 2009, Riviere *et al.* 2014). The explanation for these type of community responses are rather complex. For example a dietary intervention with resistant starch usually results in increased butyrate concentrations in the intestine, indicative of complex food webs consisting of primary degraders

such as those belonging to the *Actinobacteria* (including the *Bifidobacteriaceae*) and Bacteroidetes that ferment starch and produce lactate, succinate, acetate and propionate, which are subsequently taken up by secondary fermenters to produce mainly acetate, propionate and butyrate, explaining the emergence of elevated butyrate levels upon resistant starch feeding via a cross-feeding (syntrophic) interaction within the microbial consortium (Duncan, 2004; Flint, 2008).
In our previous study mice were fed five different fibres or a control diet containing corn starch that is readily degraded and largely absorbed within the small intestine (**Chapter 4**). Colonic samples obtained from the non-starch fibre-fed mice showed an overall increase in luminal SCFA concentrations, which coincided with an increased relative abundance of the *Bacteroides distasonis* group and *Clostridium* cluster XIVa species in the colonic microbiota. These bacteria most likely act as primary or secondary fibre degraders, although data from cultured representatives of the detected species are not available. To further specify the microbial involvement in the degradation of these fibres, and to study the effects on the host mucosal tissue, this study focuses on the main fibre fermentation location in the mouse intestine, i.e. the caecum (**Chapter 4**). The large caecal volume is compatible with multiple analyses in individual mice, including the determination of SCFA concentrations, microbiota composition and its activity using metatranscriptome analysis. Furthermore, these multivariate microbial and metabolic datasets were correlated to host mucosal responses in the caecum, determined by tissue transcriptome analyses, aiming to construct models of diet-microbe-host interactions.

Results

The effects of different dietary fibres - Resistant Starch (RS), Arabinoxylan (AX), Fructooligosaccharides (FOS), Inulin (IN) and Guar Gum (GG) - were studied to evaluate whether the responses of the microbiota and the host to different dietary fibres are congruent or diverse. Male C57BL/6J mice of 9 weeks (young adults) were given the control diet, a standard semi-synthetic low fat diet containing corn starch as digestible carbohydrate, for two weeks, followed by a 10 days dietary intervention with the fibre diets. Mice were allocated according to their body weight to one of six diet groups (n=10 per diet group): control (CON), IN, FOS, AX, GG or RS type 3. The diets enriched in fibre were identical to the control diet, except that 10% (w/w) of corn starch was replaced by each fibre (20% for RS, since it was 50% resistant) (**Chapter 4**, Table S 4.1). The effects of these dietary interventions were measured in caecal samples by determination of the steady state levels of luminal SCFA, the microbiota composition by 16S rRNA gene profiling, and the microbiota activity by metatranscriptomics, respectively. In addition, the cognate host responses in the caecal mucosa were determined using murine microarrays and this was correlated to the 16S microbial composition dataset and to the microbial metatranscriptome data (Figure 6.1).

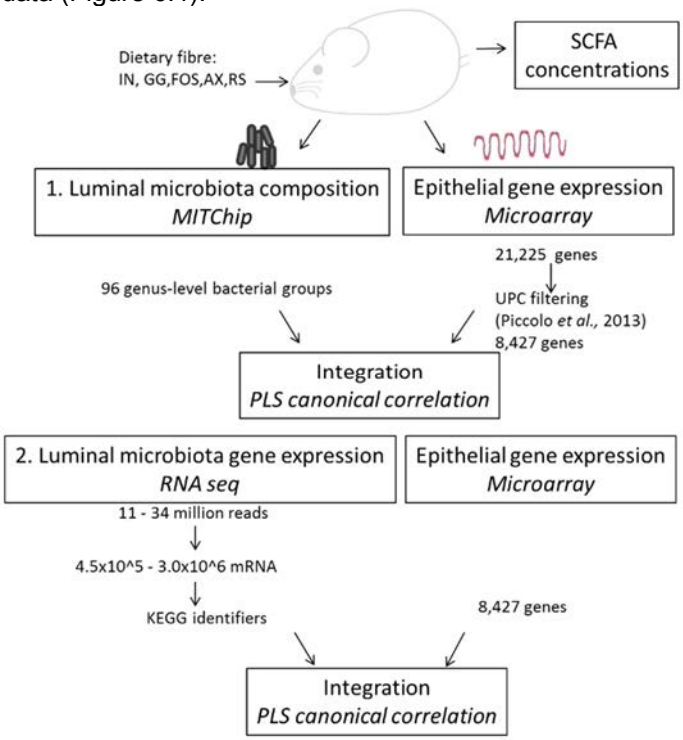


Figure 6.1 Flow-chart of data analysis.

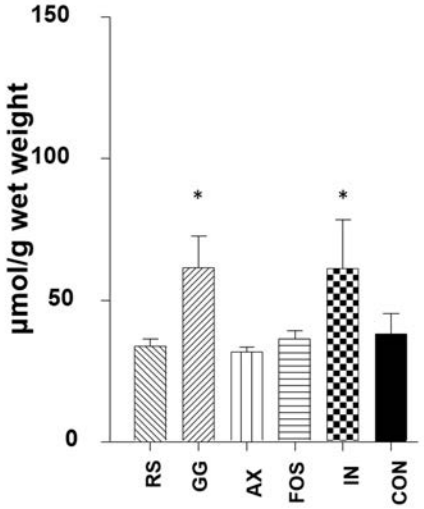


*Figure 6.2 Ceacal luminal SCFA concentrations in µmol/g content measured with gas chromatography. Abbreviations are for control (CON), Resistant Starch (RS), Arabinoxylan (AX), Fructooligosaccharides (FOS), Inulin (IN) and Guar Gum (GG). * indicates significance (Student T.test P<0.05) between the dietary group and control (CON)*

Dietary fibres differentially modulate luminal SCFA levels

In all mice the main metabolic products of dietary fibre fermentation in the caecal lumen were analysed by gas chromatographic quantification of acetate, propionate, butyrate, valerate, and the branched-chain SCFA iso-butyrate and iso-valerate. Total SCFA concentrations significantly increased in caecal lumen of mice in the IN and GG diet groups as compared to the CON diet ($p<0.05$), and the highest SCFA concentrations were observed for mice fed the IN diet (Figure 6.2). In contrast, the RS, AX and FOS diets no changes the overall SCFA concentrations were observed.

Dietary fibres modulate the microbiota composition

The caecal content of five mice per dietary treatment was used to analyse the microbiota after 10 days of dietary treatment. The density of the caecal microbiota was analysed by 16S rRNA gene-targeted quantitative PCR (qPCR), whereas the MITChip platform was employed for compositional profiling.

Although all fibre diets, and especially the IN-diet, tended to increase the microbiota density in the caecum as compared to the CON diet, none of these effects were significant (Figure S 6.1). Microbial diversity was also not affected by the diets used in the present study (data not shown). Analogous to what has previously been reported for the colon in these mice, caecal microbiota composition of mice fed the GG, AX, IN, or FOS diet, but not those receiving an RS-diet, could be discriminated from that of mice receiving the control diet (Figure S 6.2). In order to correlate changes in microbiota composition and metabolism to the different diets, the SCFA concentrations and the hybridization signals of in total 96 genus-level phylogenetic groups targeted by the MITChip were subjected to redundancy analysis (RDA). The RDA included the concentrations of acetate, propionate and butyrate and the diets as explanatory variables, which were concluded to explain 58.6% of the total

variation in microbial composition, of which 73.5% was captured within the first two canonical axes of the RDA analysis (Figure 6.3). The RS and CON diet groups clustered separately from the IN, AX, FOS and GG groups along the first canonical axis, whereas the FOS and IN diet-groups appeared to be largely overlapping. Genus-like groups that correlated in their relative abundance with the RS and the control diet included *Collinsella*, *Propionibacterium*, *Olsenella* et rel., *Alistipes* et rel., *Bacteroides splanchnicus* et rel., *Porphyromonas asaccharolytica* et rel., *Fibrobacter succinogenes* et rel., *Lactobacillus salivarius* et rel., *Staphylococcus aureus* et rel., *Turicibacter* et rel., *Clostridium perfringens* et rel., *Faecalibacterium prausnitzii* et rel., *Ruminobacter amylophilus* et rel., *Clostridium difficile* et rel., *Ruminococcus obeum* et rel., *Fusobacterium*, *Bilophila* et rel., *Desulfovibrio* et rel., *Sphingomonas* et rel., *Labrys methylaminiphilus* et rel. and Unclassified TM7. In the opposite direction relative abundance of the groups *Clostridium herbivorans* et rel., *Clostridium sphenoides* et rel., an unclassified *Clostridium* cluster XIVa group and *Bacteroidetes distasonis* et rel. was correlated with IN, FOS, GG and AX diets. Moreover, luminal acetate and propionate concentration also correlated with the IN, FOS, GG and AX diets and the four microbial groups that associated positively with these diets. An increased luminal butyrate concentration was observed with the GG and AX diets, but did not strongly correlate with a specific microbial group.

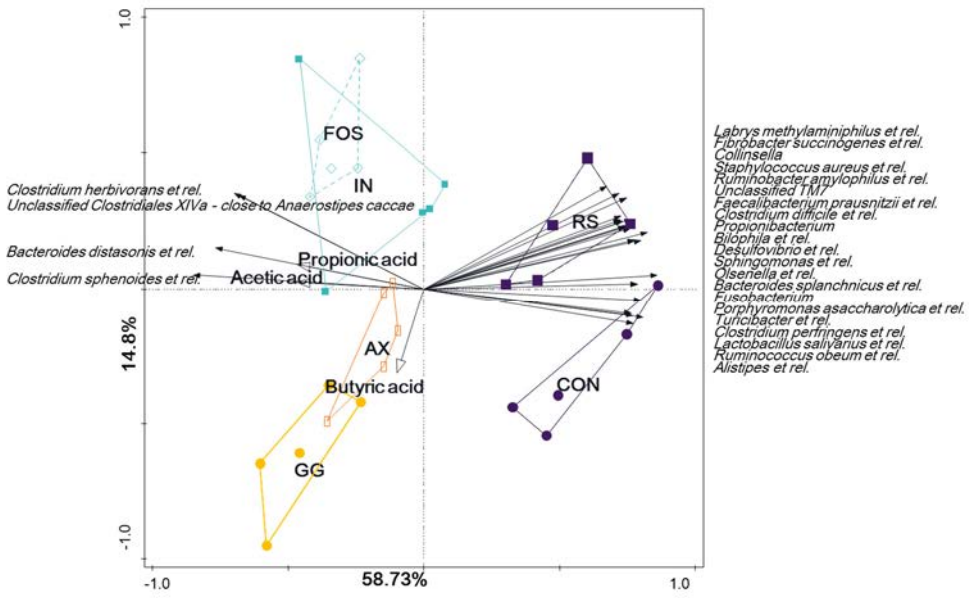


Figure 6.3 Redundancy analysis, RDA, where the explanatory variables are acetate, propionate, butyrate and the diets: control (CON), Resistant Starch (RS), Arabinoxylan (AX), Fructooligosaccharides (FOS), Inulin (IN) and Guar Gum (GG). These variables explain 58.6% of total variation. In this plot 73.5% of the explained variation is shown.

Effect of dietary fibre on caecum mucosal gene expression

We used Partial Least Square Discriminant Analysis (PLS DA) to analyse gene expression changes in caecal mucosa in mice fed different diets. The expression of 8427 filtered protein encoding genes was used as input for a sparse PLS DA. The analysis revealed that samples could mainly be distinguished between mice fed on CON and RS diets and mice fed IN, FOS and GG diets (Figure 6.4). AX fed mice were clustering between these two main clusters, indicating intermediate effects on gene expression as compared to all other diet groups.

The data revealed that the different fibre diets modulate gene expression differentially. This observation is similar to effects of dietary fibre on gene expression in colonic mucosa (**Chapter 4**), although contrary to the colon transcriptomes, the caecal tissue transcriptomes were clearly discriminated by diet, except for IN and FOS. This observation implies that fibre diets can elicit more specific effects on caecal mucosa as compared to colonic mucosa, which appears to be congruent with the notion that the caecum is the main fermentative organ in the mouse intestinal tract.

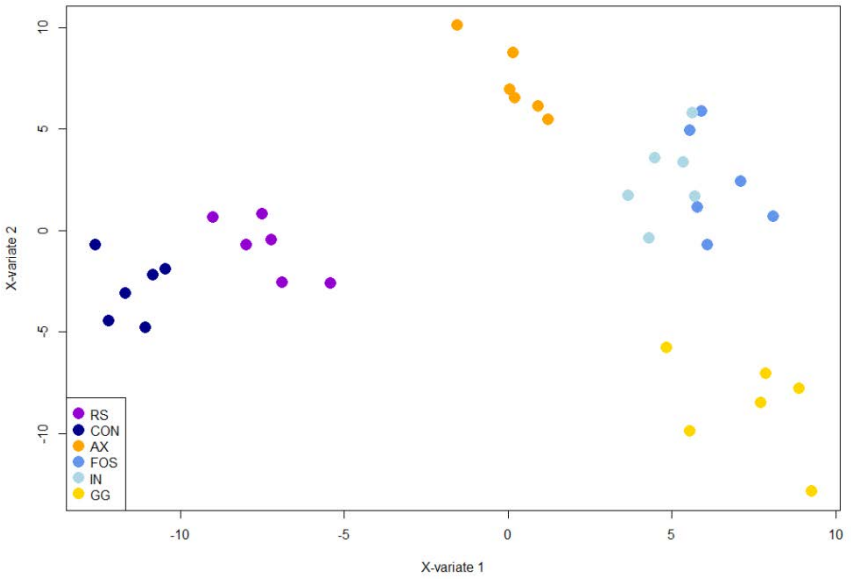


Figure 6.4 Partial Least Square Discriminant Analysis (PLS-DA) score plot on gene expression profiles in caecal epithelial cells of mice fed different diets: control (CON), Resistant Starch (RS), Arabinoxylan (AX), Fructooligosaccharides (FOS), Inulin (IN) and Guar Gum (GG). In the plot the samples (individual mice) were plotted based on the two main variates.

Correlation between microbiota composition and caecal mucosa gene expression profiles

To reveal associations between microbiota composition and mucosal gene expression profiles in the different diet groups, we integrated these two microarray datasets using a PLS-based canonical correlation approach (Figure 6.1). The results for the first three components were represented in a Clustered Image Map (CIM) with correlation coefficients depicted by different colours (Figure 6.5).

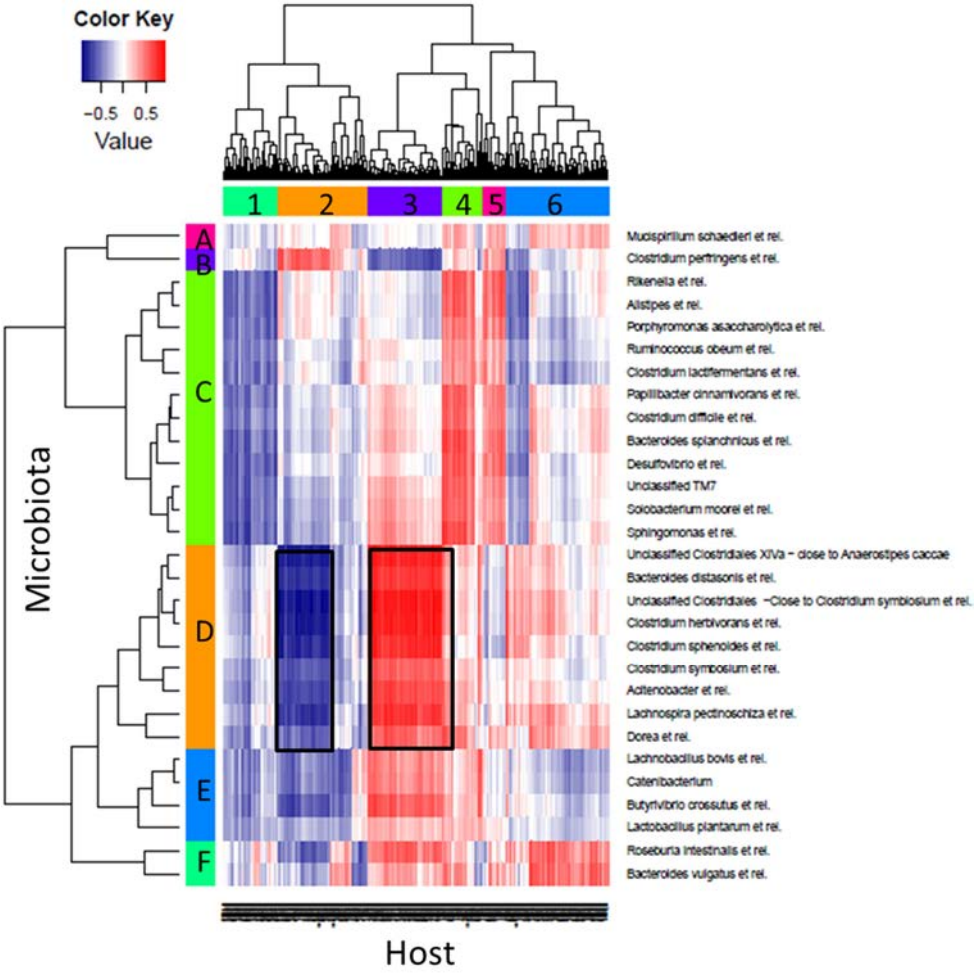


Figure 6.5 Correlation of epithelial cell gene expression with luminal microbiota composition. Sparse PLS canonical correlation analysis was performed to integrate gene expression values with relative abundance data of bacteria for individual mice. The heatmap represents the correlation structure of both dataset; red: positively correlated, blue: negatively correlated. More intense colours indicate stronger correlation. Correlation values were subjected to unsupervised hierarchal clustering based on Euclidean distance for both genes and microbial groups. Six main gene clusters (1-6) and six main bacterial clusters (A–F) were identified.

In total 599 mucosal genes and 29 bacterial groups (abundance > 1%) were retained for the first three components, and clustering of the correlation coefficients revealed six main clusters of host genes that correlated to six main clusters of bacteria (Gene lists per cluster are available upon request).
The largest microbiota cluster in this correlation analysis (Figure 6.5, cluster C) contained a range of bacterial groups, of which some were previously shown to be associated with the control and RS diets, like *Alistipes* et rel., *Bacteroides splanchnicus* et rel., *Porphyromonas asaccharolytica* et rel., *Clostridium difficile* et rel., *Ruminococcus obeum* et rel., *Desulfovibrio* et rel., *Sphingomonas* et rel., and Unclassified TM7. This microbiota cluster correlated with the repression of epithelial gene sets associated with phosphorylation, other signalling processes and protein modification (cluster 1), although this effect appeared not completely specific for this microbial cluster. Likewise, this microbial cluster also correlated with enhance expression levels of genes associated with protein catabolism and RNA related processes and apoptosis (cluster 4), although this host response was again not exclusively correlated to this group of bacteria. Stronger correlations were identified between the second largest bacterial cluster (cluster D) that contained microbial groups including *Parabacteroides distasonis* et rel., *Acitenobacter* et rel. and seven genus groups within the *Clostridium* cluster XIVa group. Remarkably, this microbiota cluster encompassed all the microbial groups that were previously found to be associated with the FOS, IN, AX, and GG diets, e.g. *Clostridium herbivorans* et rel., Unclassified *Clostridiales* XIVa (*close to Anaerostipes caccae*), *Clostridium sphenoides* et rel., and *Bacteroides distasonis* et rel.. Cultured representatives of these bacterial groups are known for their capacity to degrade diet and host derived oligo- and poly-saccharides. Higher relative abundances of these bacterial groups were correlated with an elevated expression of genes associated with NF-κB related processes, as well as proteolysis and energy-generating processes, including oxidative phosphorylation and glycolysis. Inversely, the relative abundance of these microbial groups correlated negatively with the expression of mucosal functions related to immune-responses (T cell activation), DNA/RNA and glycosaminoglycan biosynthesis, and cell-cell communication. A single microbial group (*Clostridium perfringens* et rel.) appeared to correlate with exactly opposing mucosal transcriptome responses. The remaining microbial cluster A, E and F, encompassed functionally as well as phylogenetically broad microbial groups that correlated with relatively mild changes in caecal mucosa gene expression patterns.
These analyses illustrate that bacterial groups that were associated with the fibre containing diets (except the RS diet) and the elevated caecal SCFA concentrations, also were associated with alterations in gene expression profiles in the caecal mucosa of the mice in variable ways, affecting processes that include energy metabolism and immune response, which are the essential pillars of host-microbe homeostasis in the intestinal tract.

Effect of IN and GG on gene-functions expressed by the microbiota

The IN and GG diet groups showed the highest SCFA concentrations and most distinct microbial profiles compared to the CON diet. Therefore, the caecal content of 4 mice of the IN, GG and CON groups were used for metatranscriptome analysis, using a procedure encompassing extraction of total microbial RNA, rRNA depletion, double-strand cDNA synthesis and single-end shotgun Illumina sequencing. A total of 11 to 34 million metatranscriptome reads were generated per sample. These reads were filtered for non mRNA and low quality reads, leading to the extraction of 4.5 x 10^5 to 3.0 x 10^6 mRNA reads that were assembled and functionally and phylogenetically assigned to unravel the overall as well as group-specific microbiota activity profiles (see M&M for details; **Chapter 5**). Using this approach approximately 46-69% of the overall mRNA reads could be assigned to specific transcripts (Table 6.1), of which 70-85% was functionally annotated.

	Total reads after quality filtering	mRNA	Assembled mRNA reads	Bacterial protein coding in assembled contigs
GG_1	1.3E+07	8.2E+05	61.3%	80.6%
CON_1	1.4E+07	9.4E+05	55.1%	83.6%
IN_1	1.7E+07	1.2E+06	54.6%	71.6%
GG_2	2.0E+07	1.8E+06	55.3%	84.9%
IN_2	1.6E+07	2.0E+06	58.6%	83.6%
GG_3	2.5E+07	1.7E+06	68.7%	75.8%
IN_3	1.4E+07	7.0E+05	58.4%	70.4%
CON_2	2.5E+07	1.5E+06	57.0%	79.4%
GG_4	2.0E+07	1.3E+06	45.7%	70.6%
CON_3	1.1E+07	4.5E+05	58.4%	79.8%
CON_4	2.8E+07	1.0E+06	54.0%	75.1%
IN_4	3.4E+07	3.0E+06	56.4%	83.0%

Table 6.1 . Reads of the Iilumina sequences and the result of data processing per sample.

Using the sample specific functional assignments of the detected transcripts, without taking their phylogenetic origin into account, RDA analysis revealed that these function-patterns clustered according to diet, clearly separating CON, and GG and IN diets (Figure 6.6). To further unravel the functional impacts of the different fibre diets (GG and IN versus control), the functions (i.e., KEGGs) associated with the first axis of the RDA (explaining ~ 27.6 % of the overall variation within this analysis) were evaluated. Most of the differentially expressed KEGGs were within the KEGG category of “Metabolism” (>70%), predominated by carbohydrate, amino-acid, and central energy metabolism associated functions, as well as a scattering of other metabolic functions (Table S 6.1). In addition, genes belonging to the KEGG category “Environmental Information Processing” appeared to be substantially represented in the microbial activity patterns that discriminated the IN and GG diets

from the CON diet. This category includes functions involved in response and adaptation of bacteria to their environment and encompasses both signal transduction pathways like two-component systems, as well as "membrane transport" associated functions like ABC-transporters and Phosphotransferase systems (PTS), that are important for the utilization of alternative carbon sources (e.g., fibres from the diet), and are required to drive the adaptations of the overall microbiota metabolism that was detected. Intriguingly, the CON diet microbiota appeared to express genes with KEGG identifiers of the "cell motility" category at a higher level as compared to the microbiota of mice fed GG or IN diets. The induction of bacterial motility within the microbiota, specifically in the caeca of the CON-diet animals, may imply that the microbiota experiences nutrient starvation under these conditions, which appears in agreement with the lower abundance of dietary nutrients (e.g. fibres) in the caecal lumen of these animals, and is known to induce motility *in vitro* (Aizawa and Kubori 1998, Koirala *et al.* 2014).

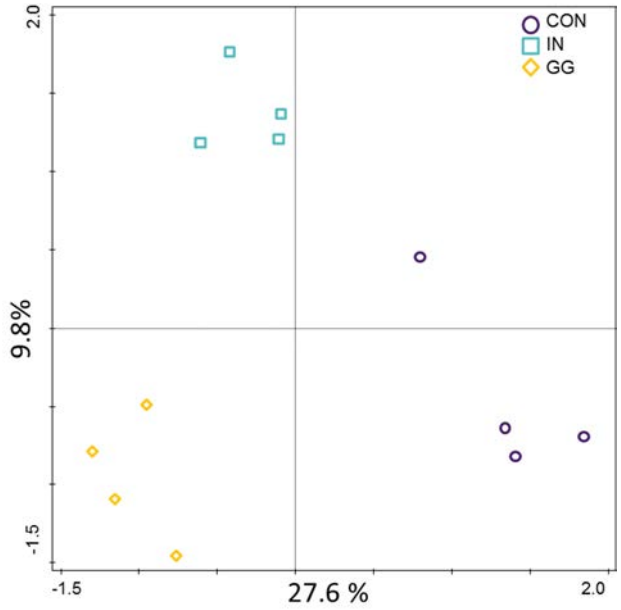


Figure 6.6 Redundancy analysis (RDA) of metatranscriptome data, where the explanatory variables are the diets control (CON), Inulin (IN) and Guar Gum (GG). These variables explain 37.4% of total variation. In this plot all the explained variation is shown.

Correlation between microbiota expressed functions and caecal mucosa gene expression patterns

To pinpoint microbial and host functions that are associated (Figure 6.1), the metatranscriptome KEGG functions were correlated with the caecal mucosa gene expression patterns using sPLS canonical correlations, leading to the detection of two strongly correlating co-clusters of bacterial and host transcripts (Figure 6.7). The strongest correlations (Figure 6, marked areas of the correlation analysis),

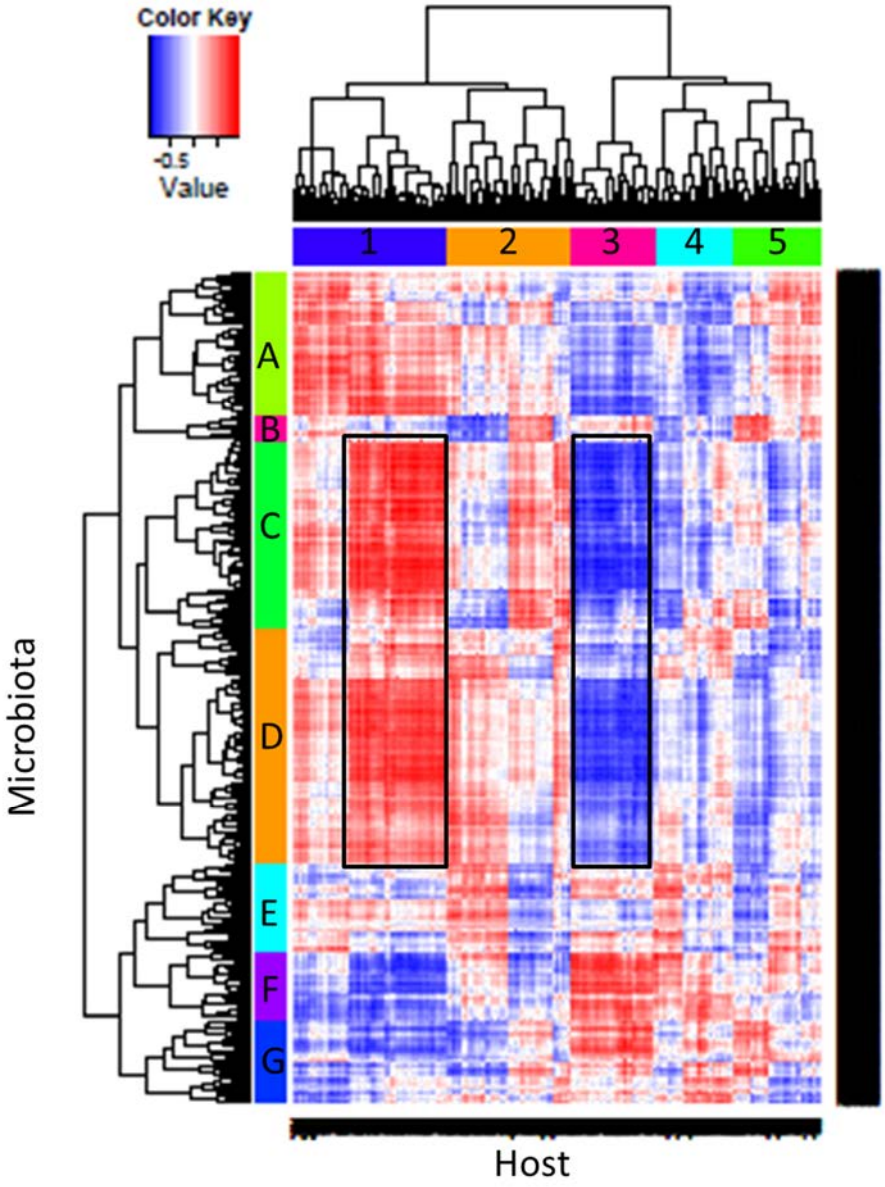


Figure 6.7 Correlation between microbiota expressed functions and caecal gene mucosa expression patterns. Sparse PLS canonical correlation analysis was performed to integrate epithelial cell gene expression with microbial metatranscriptome KEGG functions for individual mice. The heatmap represents the correlation structure of both datasets; red: positively correlated, blue: negatively correlated. More intense colours indicate stronger correlation. Correlation values were subjected to unsupervised hierarchal clustering based on Euclidean distance for both genes and microbial groups. Five main gene clusters (1-5), and seven main bacterial clusters (A-G) were identified.

encompassed host-functions associated with glutathione, glutamate metabolism and glycerolipid metabolic processes, and PPAR signalling, (negatively correlated; gene enrichment lists per cluster are available upon request), as well as immune-related processes such as B cell and T cell receptor and toll- like receptor signalling (positively correlated, gene enrichment lists per cluster are available upon request). The microbial functions correlated with these host processes, were scattered across a variety of different microbial pathways and processes, and only displayed a minor enrichment of the KEGG category "Amino-acid metabolism". Intriguingly, the microbiota-associated functions that displayed an opposite correlation with these host functions (Figure 8; the pink (B), purple (F) and dark blue (G) clusters), predominantly belonged to the KEGG categories "Environmental Information Processing", "Cellular Processes", "Genetic Information Processing" and "Nucleotide metabolism" (Table S 6.2).

Effect of IN and GG on active microbial community

The microbiota functional profiles determined in the different diet groups could be due to a shift in the relative contribution to the overall activity patterns by specific microbial groups rather than a function adaptation of the overall microbial community per se. The differentially expressed KEGG annotated functions could be assigned to five bacterial families, i.e., *Verrucomicrobiaceae, Bacteroidaceae, Lactobacillaceae,* and *Erysipelotrichaceae* were detected at a significantly higher level in the GG diet compared to the control diet (Error! Reference source not found.). Conversely, the *Lachnospiraceae* assigned functions were in lower abundance in the GG diet compared to control. The same bacterial families displayed a similar trend in the IN samples compared to control samples, albeit not reaching significance.

To identify which bacterial families are active in the degradation of the fibres, all metatranscriptome datasets were analysed in detail for the taxonomic origin of expressed genes involved in glycoside hydrolysis (Figure 6.8a) and saccharide transport (Figure 6.8bc). Bifidobacteriaceae, Lachnospiraceae, Clostridiaceae and the Erysipelotrichaceae particularly expressed genes associated with glycoside hydrolysis and saccharide transport, possibly indicating active degradation of fibres and sugar transport into the cells. The *Bifidobacteriaceae* appeared to have increased expression of both glycosidases and sugar transporters in the IN and GG compared to the control. This family is known for its increase in abundance during interventions with fructooligosaccharides, like IN and FOS (Gibson and Wang 1994, Cummings *et al.* 2001, Barboza *et al.* 2009, Kleerebezem and Vaughan 2009). The increased activity of *Bifidobacteriaceae* during GG intervention has not been reported to date, but according to the metatranscriptome data generated in this study appeared to be even somewhat more elevated in GG diet as compared to the IN diet. This finding implies that the *Bifidobacteriaceae* are strongly involved in GG fibre catabolism. The other bacterial families detected displayed quite distinct responses to the different diets. The *Lachnospiraceae* increased the glycosidase expression in IN and GG diets compared to the control diet, while their expression of sugar

transport functions was decreased. Conversely, the *Erysipelotrichaceae* elevated expression of sugar transport functions in the IN and GG diets, while their glycosidase expression appeared not to respond to the diet changes. Finally, the *Clostridiaceae* decreased the glycosidase expression levels in the fibre diets while the carbohydrate transport remained expressed at the same level in all diets.

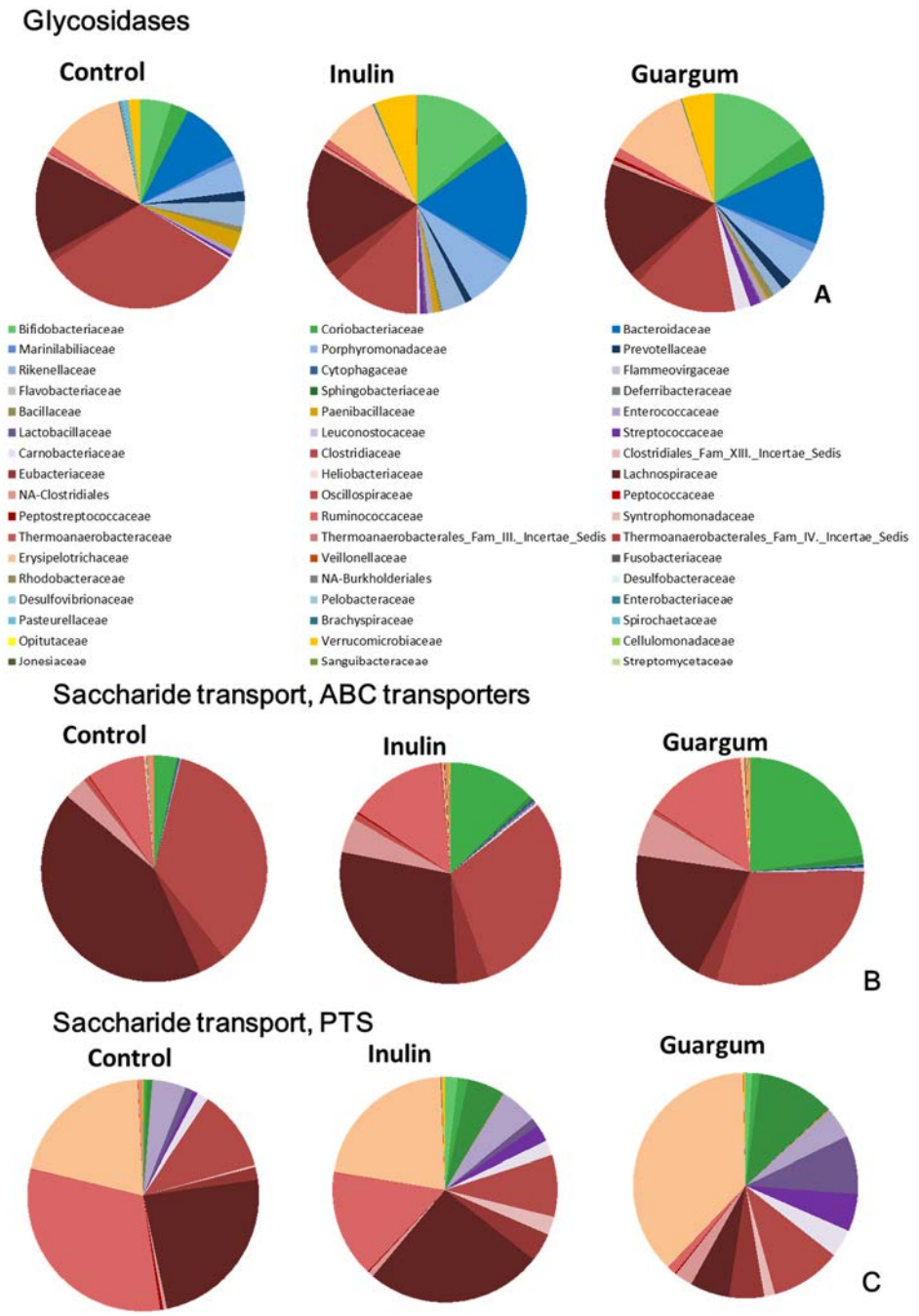


Figure 6.8 Relative abundance of gene expression of a) Glycosidases, b) ABC-transporters and c) Phosphotransferase systems (PTS) at family level.

Notably, although 25 % of all the detected glycosidase transcripts was assigned to the Bacteroidetes, this phylum hardly expressed saccharide transporter functions. The Bacteroidetes expressed a relatively large fraction of the overall (and diverse) hydrolase encoding genes, but appeared less active in importing the monomeric sugars produced by enzymatic polysaccharide hydrolysis. However, isolates within the *Bacteroides* and *Prevotella* genera encode operons that cluster genes encoding fibre binding, several carbohydrate degrading enzymes and transport functions (Dodd, 2011; Nelson 2010), and the encoded polysaccharide degradation machinery of *Bacteroides thetaiotaomicron* was proposed to be closely associated with the transport machinery anchored in the outer membrane. This assembly could ensure direct transfer from the hydrolytic enzymes to the transport system to minimize loss of monosaccharides to other bacteria (Dodd 2011), which may not require elevated expression levels of other transport functions.

The Bacilli displayed a relatively lower expression of genes encoding glycosidases and higher expression of sugar transporter-encoding genes in the IN and GG groups. In particular the Phosphotransferase systems (PTS) expression by the *Bacilli* was strongly elevated in the IN and GG groups, suggesting that these populations depend on other microorganisms to generate the mono- and disaccharides through the initial hydrolysis of dietary polysaccharides. Analogously, also the *Ruminococcaceae* actively expressed sugar transport functions, but did not contribute to the expression of glycosidase encoding genes. This is somewhat unexpected, since members of the *Ruminococcaceae* are known as fibre degraders that possess glycosidases, like *Ruminococcus bromii*, *Faecalibacterium prausnitzii* and *Ruminococcus flavevacies* (Biddle 2013; Rajilic 2014). Notably, the *Ruminococcaceae* shifted the expression of sugar transport systems from PTS systems to ABC transporters in the GG diet (and to a lesser extent in the IN diet) as compared to the CON diet. The substrate predictions of the PTS that are strongly expressed in the control diet group were related to saccharides that are derived from mucus degradation, which may be suppressed by the supply of dietary carbohydrates in the GG and IN diets (Figure S 6.4). PTS transport enables efficient import of mono- and/or disaccharides, which is coupled to substrate phosphorylation (Biddle 2013) and is commonly the preferred mode of transport in bacteria when they reside in carbohydrate limited environments. This is likely the situation for the *Ruminococcaceae* in the CON group caeca, where mucus-derived mono- and di-saccharides as well as other carbon sources were most likely depleted and/or not accessible for the members of this family. ABC transport allows import of mono- up to oligosaccharides, which could be energetically favourable for the *Ruminococcaceae* in the enriched environment associated with the fibre diets. This transport-flexibility is absent from many other bacterial families, like the *Lachnospiraceae*, of which the overall expression of carbon metabolism associated functions was suppressed in the GG and IN diet compared to the CON diet, implying that this bacterial family is unable to compete with other species for the dietary fibres. With the intention to reconstruct the activity profiles involved in SCFA production,

KEGGs assigned to key enzymes in SCFA production pathways were selected (Figure S 6.5). The expression of genes encoding these enzymes and their predicted taxonomic assignment differed between the CON, IN and GG groups in all of the pathways evaluated (Figure 6.9). In the GG diet the *Erysipelotrichaceae* seemed to more abundantly express various SCFA pathways compared to the control diet. The *Erysipelotrichaceae* were highly active in the conversion of pyruvate to lactate and/or vice-versa. In addition the *Erysipelotrichaceae* did express the acetyl-coA to butyryl-coA via crotonyl-coA pathway, where the conversion of crotonyl-coA to butyryl-coA allows anaerobes to conserve energy (Figure S 6.6; (Seedorf *et al.* 2008) and should lead to butyrate production. However the butyrate kinase and the butyryl-CoA:acetate CoA-transferase enzymes of *Erysipelotrichaceae* were not identified in our dataset, which may be due to the erroneous annotation of acetate and butyrate kinases and SCFA transferases since these enzymes display a high-level of similarity (Louis and Flint 2007, Vital *et al.* 2013, Vital *et al.* 2014). Moreover, some of the phosphate butyryl-transferase and butyrate kinase were predicted to be expressed by members of the *Bacillaceae, Bacteroidaceae* and *Porphyromonadaceae* families, implying that these families are involved in the production of butyrate, which is in clear contrast with previous studies that indicated that *Bacillaceae* and *Bacteroidaceae* family members are not producing butyrate (Rajilic 2014). Again this may indicate that the inaccurate annotations of gene sequences related to kinases and transferases involved in SCFA pathways confuse these metabolic interpretations of the metatranscriptome data.

In the IN group there was a high activity in ethanol consumption by members of the *Desulfovibrionaceae*, in line with previous reports that isolates within this family use ethanol as a carbon source (Scanlan and Marchesi 2008). Transcripts associated with the production of propionate were predominantly assigned to members of the *Clostridiaceae*, *Bacteroidaceae* and *Porphyromonadaceae*, where the latter two groups probably produce propionate from oxaloacetate via succinate. Although increased luminal concentrations of acetate, propionate and butyrate were detected in the GG and IN diets, only increased relative expression levels of genes associated with propionate were detected in the IN and GG metatranscriptomes compared to those of the CON diet, whereas transcripts associated with acetate and butyrate production appeared to be present at a relatively lower level in GG and IN compared to control diet.

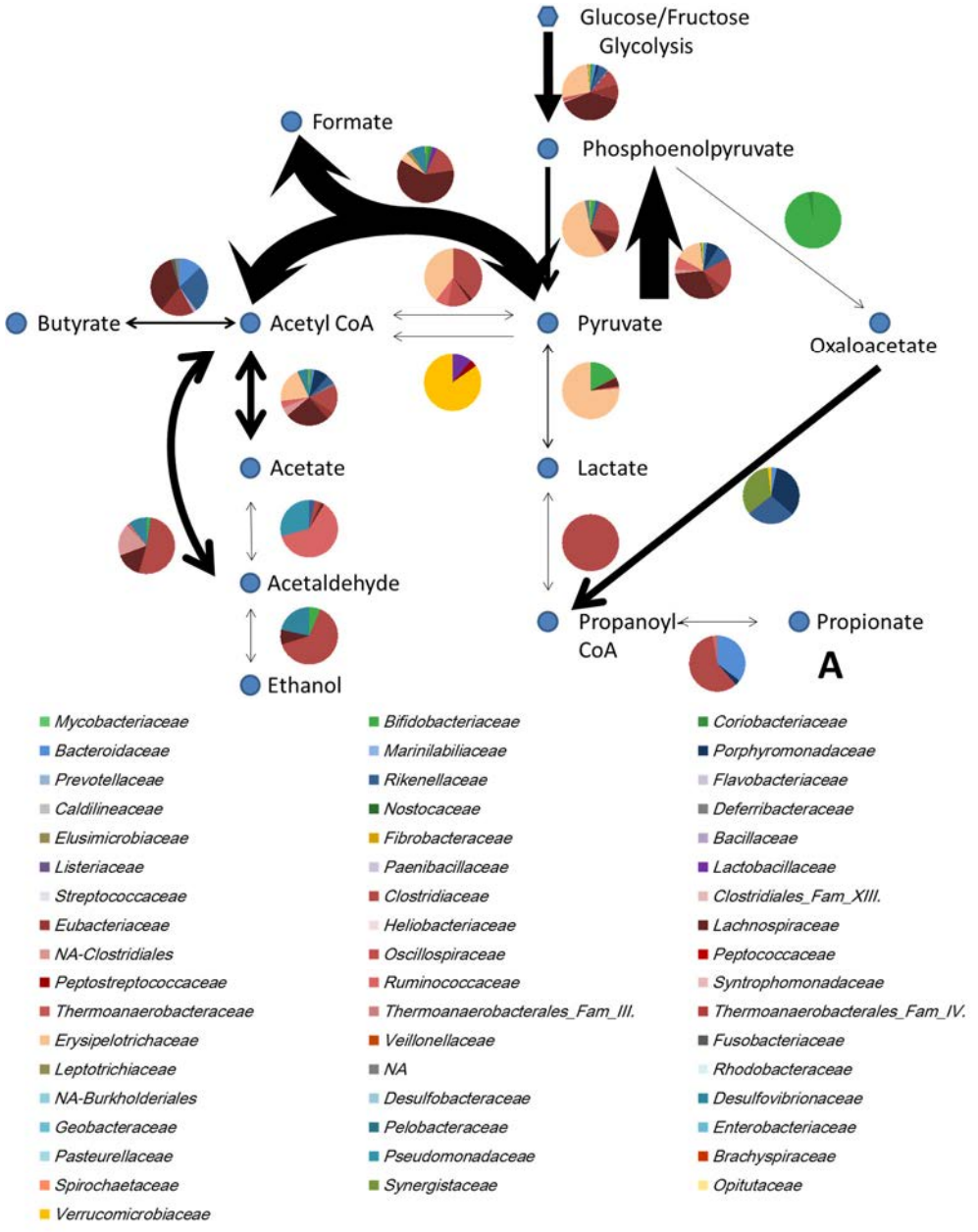
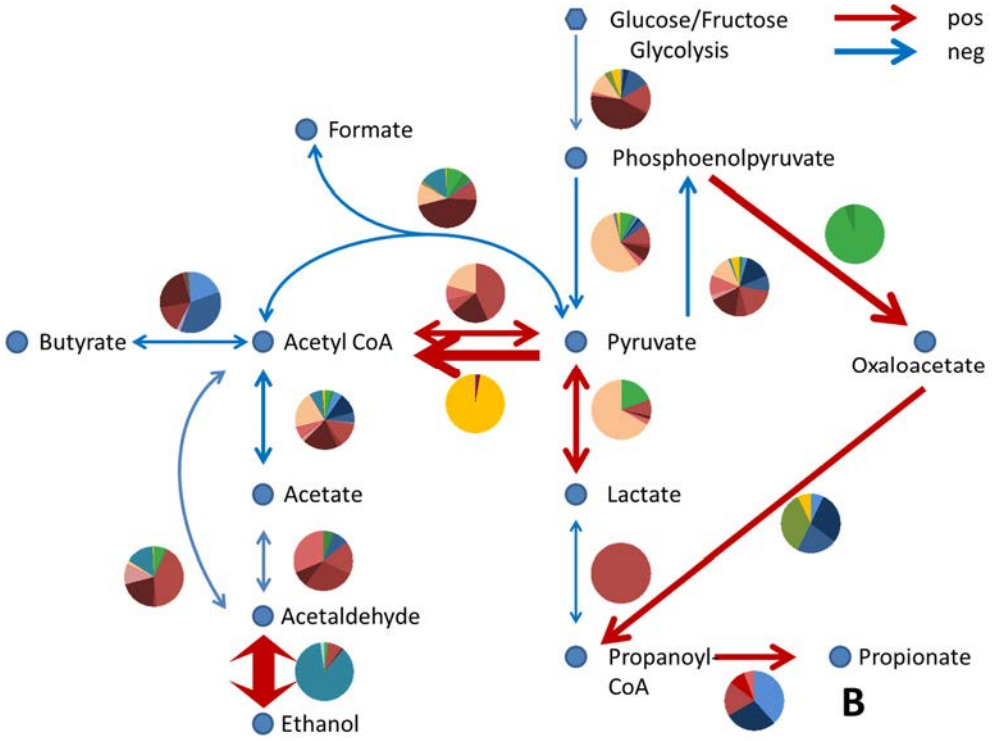
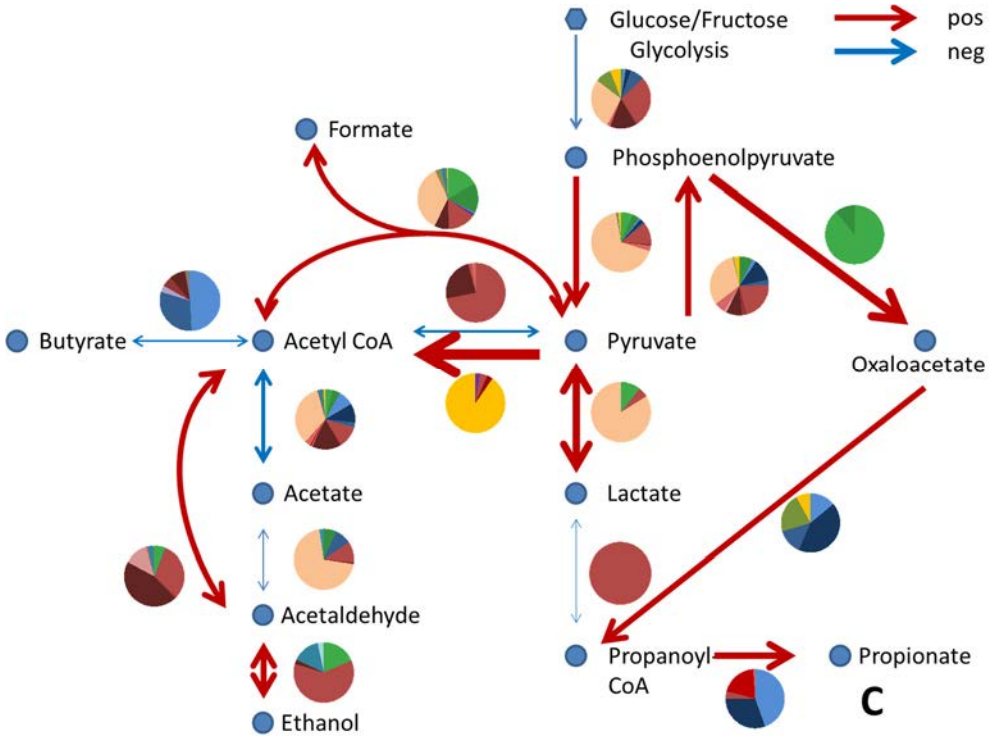


Figure 6.9 Activity of SCFA metabolism pathways. A) Control animals. Arrow thickness indicates activity of this pathway relative to the total activity. B) and C) are the Inulin and Guargum dietary treatment, respectively. The colour and thickness of the arrows indicate a fold change compared to the same path in the control diet.


Glucose/Fructose
Glycolysis
pos
neg
Formate
Phosphoenolpyruvate
Butyrate
Acetyl CoA
Pyruvate
Oxaloacetate
Acetate
Lactate
Acetaldehyde
Propanoyl CoA
Propionate
Ethanol
B


Glucose/Fructose
Glycolysis
pos
neg
Formate
Phosphoenolpyruvate
Butyrate
Acetyl CoA
Pyruvate
Oxaloacetate
Acetate
Lactate
Acetaldehyde
Propanoyl CoA
Propionate
Ethanol
C

Discussion

Dietary fibres have been associated with health benefits. After their degradation by the intestinal tract microbiota, SCFA are generated and taken up into the epithelium of the host. Here we show that the dietary fibres Inulin (IN) and Guar Gum (GG) yield increased total SCFA concentrations in the caecum and in parallel increase the expression of metabolic processes involved in central energy metabolism in the caecal mucosa of the mice. Moreover microbial analysis revealed shifts in composition and activity patterns when the mice were given the different carbohydrate sources.

Mice fed the fructooligosaccharides (FOS), arabinoxylan (AX) or resistant starch (RS) diets did not have increased total SCFA concentrations in their caeca, although FOS and AX could still stimulate enhanced expression of the central energy metabolism pathways in the mucosa of these mice. Similar observations were reported previously for the colonic mucosa of these mice, although in the colon lumen both the FOS and AX diets increased SCFA levels (Lange & Hugenholtz, 2014). The experimental design of this study allowed sampling of luminal content at a single time point and thus generated a snapshot view of the intestinal system, which excludes interpretations of the rate of SCFA production and or consumption, which may be drastically affected by the different diets. Enhanced SCFA flux into the host mucosa could explain the similarity in the local mucosal responses measured in the IN, GG, FOS and AX diets. Such a snapshot determination of SCFA concentrations in the lumen is clearly a measurement of limited value for the determination of the microbiota fibres-fermentation output (den Besten, 2014). Microbiota composition and activity shifts may provide a better proxy for the estimation of in situ fibre-fermentation rates in the intestinal lumen, and these parameters were found strongly affected by the addition of fibres to the diet.

The fibre digestion by the microbiota could be further specified by analysis of the expression of glycosidase and sugar transport functions. Here we could distinguish three categories of bacteria, (i) bacteria that express glycosidases, but hardly sugar transporters; (ii) bacteria that express both glycosidases and sugar transporters; (iii) bacteria that hardly express glycosidases, but do express sugar transporters (Figure 6.10). The first group is mainly represented by the *Bacteroidaceae*, the *Porphyromonadaceae* and the *Verrucomicrobiaceae*, of which the latter displayed elevated expression in the IN and GG diet. The *Verrucomicrobiaceae* member detected predominantly belonged to *Akkermansia muciniphila* (data not shown). The increase of activity in the IN and GG groups of this typical mucin digesting microbe (Derrien *et al.* 2008) may indicate that these diets lead to increased mucus production, although this was not apparent from the mucosal transcriptome analyses, but may not be primarily regulated at the level of gene transcription (Hedemann *et al.* 2009). The second group corresponds to members of the *Bifidobacteriaceae, Lachnospiraceae, Erysipelotrichaceae* and *Clostridiaceae*.

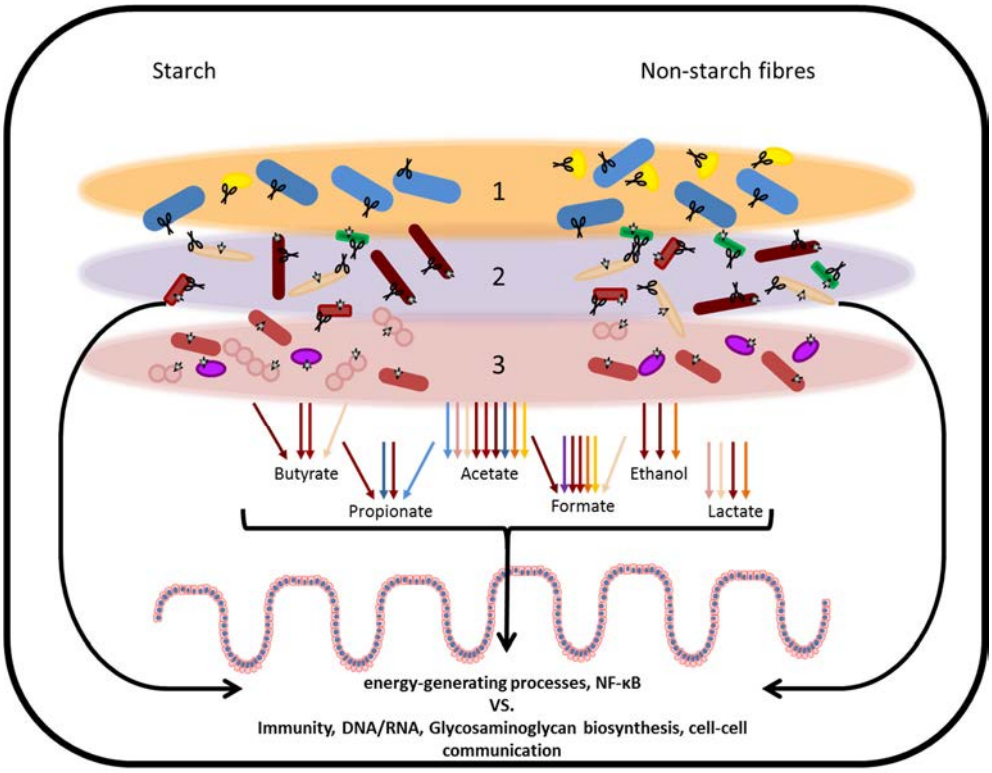


Figure 6.10 Overview of identified processes during fibre fermentation.

These four families all expressed sugar transporter encoding genes parallel to the glycosidases, which is in clear contrast to the first group and may be due to a lack of mechanistic coupling of polysaccharide hydrolysis and saccharide transport as has been proposed for some members of the first group of bacteria (e.g. *Bacteroides thetaiotaomicron*; Dodd, 2011). Notably, the *Erysipelotrichaceae* have only recently been recognized as a separate bacterial family, and many members still need to be

characterized and re-assigned to this family ((Rajilic-Stojanovic and de Vos 2014), NCBI taxonomy, September 2014). Our results show that this family plays an important role in the murine microbiota and contributes strongly to its overall metabolite conversions (Figure 6.10). The third group are bacteria that profit from glycosidase activity of other bacteria and import the released sugars. These are the *Eubacteriaceae*, several Bacilli families and the *Ruminococcaceae*. All bacterial families in three different categories ferment the carbohydrates they ingest to produce SCFA in different composition and ratio, although their individual contribution to the overall SCFA production by the microbiota may differ as a function of the dietary treatment. However, deciphering of the specific activity in particular SCFA pathways of individual bacterial families was hampered by the apparent inaccuracy of SCFA pathway mapping of genes, where many functions appear to be wrongly assigned, due to the high degree of similarity of the enzymes involved. To overcome this, advanced annotation and domain recognition tools need to be developed to accurately dissect these different enzyme families, which is a prerequisite to enable SCFA pathway reconstruction for environmental samples on basis of metatransciptome or similar metagenomic information.

In summary the SCFA can be used as an energy source for the epithelial cells of the host. The correlation of MITChip and host gene-expression revealed fibre degrading and possibly butyrate producing bacteria activating energy metabolism in the host and repress transcriptional regulation and immune system processes. Moreover, based on the KEGG functions derived from the metatranscriptome data we observed a correlation with similar host genes to KEGG functions related to bacterial growth. This could indicate that active and fibre utilizing bacteria influence the host mucosa directly by enhancing its energy metabolism and affecting the immune system. Next to the known fibre responding families - *Bacteroidaceae*, *Porphyromonadaceae*, *Verrucomicrobiaceae*, *Bifidobacteriaceae*, *Lachnospiraceae*, *Clostridiaceae, Eubacteriaceae*, several *Bacilli* families and the *Ruminococcaceae* – we identified a new family, the *Erysipelotrichaceae*, as a prominent and active member of the murine gut microbiota.

Material and methods

Short-chain fatty acid analysis in caecal luminal content

Short chain fatty acids were measured in mouse intestinal samples at section. Luminal content of the caecum (ten mice per group) was collected in H_3PO_4 and isocaproic acid (as an internal standard) containing buffer solution. Samples were stored at -20ºC until further processing. The day of analysis, samples were thawed, centrifuged at 14.000 rpm (5 min), and supernatant was collected and stored at 5ºC. The samples were then subjected to gas chromatography (Fisons HRGC Mega 2, CE Instruments, Milan, Italy) at 190°C using a glass column fitted with Chromosorb 101 with a carrier gas (N2 saturated with methanoic acid).

Microbial Composition

Total DNA was extracted from 0.01-0.1 grams of caecal content samples using the repeated bead beating plus column (RBB+C) method of (Yu and Morrison 2004). The composition of microbial communities in the intestinal samples was analysed with the Mouse Intestinal Tract Chip (MITChip). This phylogenetic microarray was designed using criteria of the Human Intestinal Tract Chip (HITChip) (Rajilic-Stojanovic *et al.* 2009). The MITChip consists of 3,580 different oligonucleotides specific for the mouse intestinal microbiota (Rajilic-Stojanovic *et al.* 2009, Geurts *et al.* 2011a, Reikvam *et al.* 2012). The oligonucleotides on the array target the V1 and V6 regions of bacterial 16S rRNA genes. The 16S rRNA genes were amplified from twenty nanogram of DNA extracted from intestinal samples, with the primers T7prom-Bact-27-F and Uni-1492-R (Table 6.2). PCR products were then transcribed, and RNA was labelled with Cy3 and Cy5 dyes and fragmented as described previously (Rajilic-Stojanovic *et al.* 2009, Geurts *et al.* 2011a, Reikvam *et al.* 2012). Finally the samples were hybridized on the arrays at 62.5ºC for 16h in a rotation oven (Agilent Technologies, Amstelveen, The Netherlands). After washing and scanning of the slides, data was extracted with the Agilent Feature Extraction software, version 9.1. The data was normalized and analysed using a set of R-based scripts in combination with a custom-designed relational database, which operates under the MySQL database management system. To determine correlation of the Robust Probabilistic Averaging (RPA) signal intensities of 2667 specific probes for the 96 genus-level bacterial groups detected on the MITChip with a specific diet or SCFA, redundancy analysis (RDA) in Canoco 5.0 was used, and visualized in a triplot (Lahti *et al.* 2011, ter Braak and P 2012). The Monte Carlo Permutation test was used to assess the significance of the variation in the dataset in relation to the diet and SCFA.

The Unpaired Wilcoxon signed-rank test was used to determine bacterial groups significantly different between the control and resistant starch, IN, FOS, arabinoxylan and the guar gum diet. The RPA signal intensities of the 96 genus-level groups were tested.

Primer name	Sequence	Application
T7prom-Bact-27-F	5'-TGA ATT GTA ATA CGA CTC ACT ATA GGG GTT TGA TCC TGG CTC AG–3'	MITChip
Uni-1492-R	5'-CGG CTA CCT TGT TAC GAC-3'	MITChip
PROK1492R	5' -GGW TAC CTT GTT ACG ACT T-3'	QPCR
BACT1369F	5'-CGG TGA ATA CGT TCY CGG-3'	QPCR

Table 6.2 List of primers (Suzuki et al. 2000, Rajilic-Stojanovic et al. 2009).

Quantification of bacterial community

Quantification of the bacterial 16S rRNA gene was done by a qPCR assay using the primers developed by (Suzuki *et al.* 2000). The qPCRs were performed in 384-well plates (BioRad) sealed with a film (Microseal B film, Bio-Rad) using a MyIQ cycler with MyIQ software (version 1.0.410, Bio-Rad). The reactions were carried out in a total volume of 12.5 µl consisting of 1x IQ SYBR green Supermix (Bio-Rad), 200 nM of the forward and reverse primer and 2 µl of template DNA, and the cycling program and melting curve analysis as previously described (van den Bogert *et al.* 2011). The standard curve consisting of an 8-fold dilution series was a 16S rRNA gene PCR product of *Escherichia coli* top10. The copy number was calculated per caecum weight.

RNA extraction, mRNA enrichment, cDNA synthesis and illumina sequencing.

Four intestinal caecum content samples from each dietary treatment were used to analyse the metatranscriptome activity profiles. The RNA was extracted from 0.1-0.2 grams of caecal content. The content was re-suspended in 500 µL ice-cold TE buffer (Tris-HCL pH 7.6, EDTA pH 8.0). Total RNA was obtained via the Macaloid-based RNA isolation protocol (Zoetendal *et al.* 2006, Leimena *et al.* 2013) with in addition the use of Phase Lock Gel heavy tubes (5 Prime GmbH, Germany) during the phase separation. The RNA purification was done with the RNAeasy mini kit (Qiagen, USA), including an on-column DNAseI (Roche, Germany) treatment (Zoetendal *et al.* 2006). The total RNA was eluted in 30 µL ice-cold TE buffer and the RNA quantity and quality were assessed using a NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, USA) and Experion RNA Stdsens analysis kit (Biorad Laboratories Inc., USA), respectively. mRNA enrichment was performed by using the mRNA enrichment kit (MICROBExpressTM, Ambion, Applied Biosystem, The Netherlands) using the manufacturer's protocol. As with the total RNA the quantity and quality were assessed, the latter was done to check on the efficiency of the mRNA enrichment. One µg of the enriched mRNA sample was used to transcribe the RNA in cDNA. Double stranded cDNA was synthesized with the SuperScript® Double-Stranded cDNA Synthesis kit (Invitrogen, the Netherlands), with addition of SuperScript® III Reverse Transcriptase (Invitrogen, the Netherlands) and random priming using random hexamers (Invitrogen, the Netherlands) (Yoder-Himes *et al.* 2009, Leimena *et al.* 2012, Leimena *et al.* 2013). To remove the RNA a RNAse A

(Roche, Germany) treatment was preformed, followed by phenol-chloroform extraction of the cDNA and ethanol precipitation. The product was checked on gel and 3 to 8 µg of cDNA was send to the sequencing provider for sequencing (GATC Biotech, Germany). Single read Illumina Libraries were prepared from the double-stranded cDNA according to the ChiP-seq protocol (Schmidt *et al.* 2009) with insert sizes between 200-300bp, suing barcoded tags for library constructions to enable parallel sequencing (GATC Biotech, Germany). Sequencing was performed using Illumina Hiseq2000 and using 5pM concentration of the library and the single-end protocol (Leimena *et al.* 2013).

Sequence data processing

In total, sequencing yielded between 11 and 34 million reads per sample. The data is available in the NCBI small reads archive (sra) repository, under accession numbers SRR1569190-SRR1569206.The reads were processed via a previously described protocol (**Chapter 5**). Briefly the data was filtered for ribosomal RNA sequences, adapter sequences and poor quality reads using SortMeRNA (version 1.2) (Kopylova *et al.* 2012), cutadapt (Martin 2011), PRINSEQ (lite-version) (Schmieder and Edwards 2011), respectively. Reads smaller than 50 nucleotides were discarded. The resulting mRNA fractions were merged and de novo assembled into larger contigs, creating one reference set for all samples. A total of 70710 contigs were assembled with a total length of 59 Mb (n50=1074). Encoded in these contigs a total of 104110 potential open reading frames were predicted. To determine the taxonomic origin of the contigs, the predicted protein sequences were aligned with NCBI's non-redundant database, and the taxonomical family classification of the best hit was retrieved. Functional annotation was done by assigning KEGG orthology identifiers using the KEGG's KAAS server. Expression levels of the predicted proteins were determined by aligning the mRNA reads with assembled contigs and counting the total amount of nucleotides aligned with the corresponding ORFs.

Transcriptome analysis and Functional implications

All steps for Microarray hybridization and analysis, including RNA isolation and purification, were performed as described before (**Chapter 4**). Functional implications were analysed using Enrich (Chen *et al.* 2013b).

Multivariate statistical analysis

We used Partial Least Square analysis (PLS) from mixOmics library in R to first, analyse effects of dietary fibre on gene expression (PLS Discriminant analysis) and second, to integrate microbiota composition with mucosal gene expression (PLS canonical correlation). For latter analysis, twenty-three animals were chosen for analysis of microbiota composition and host gene expression of which for 15 mice both datasets were available and 8 mice were added that were cage partner to increase the power. For transcriptome-transcriptome analysis 11 mice were included in the analysis of which for 7 mice both datasets were available and 4 mice were added that were cage partner. To normalise the data, both datasets were log

transformed before analysis. The PLS was done in a canonical mode. A pair-wise similarity matrix for the first three components of both datasets was computed. The matrix was visualized in a clustered image map representing colour-coded variable associations (correlation coefficients) between the two datasets (from blue negative to red positive). The columns and rows were reordered according to hierarchical clustering (Euclidean distance). Analyses were performed in R using the library mixOmics (Lê Cao *et al.* 2009).

Supplemental information

Supplemental Figures

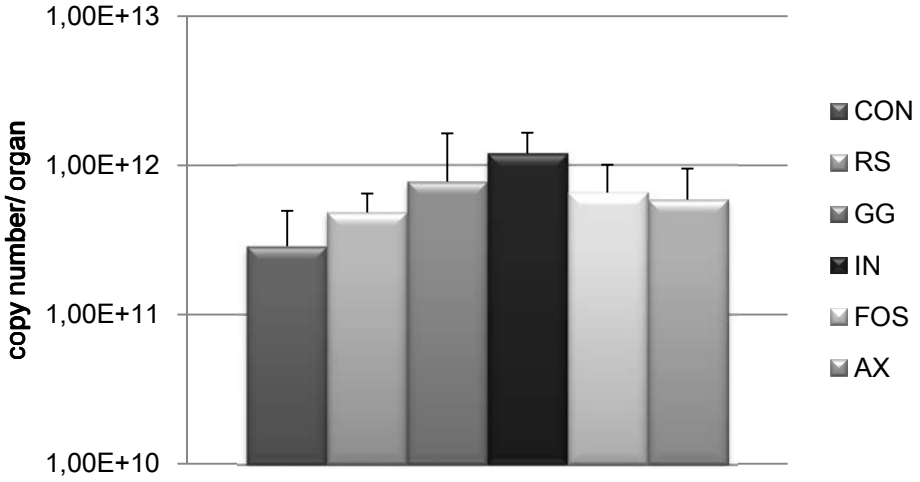


Figure S 6.1 Microbial abundance as measured by 16S rRNA gene-targeted quantitative PCR (qPCR).

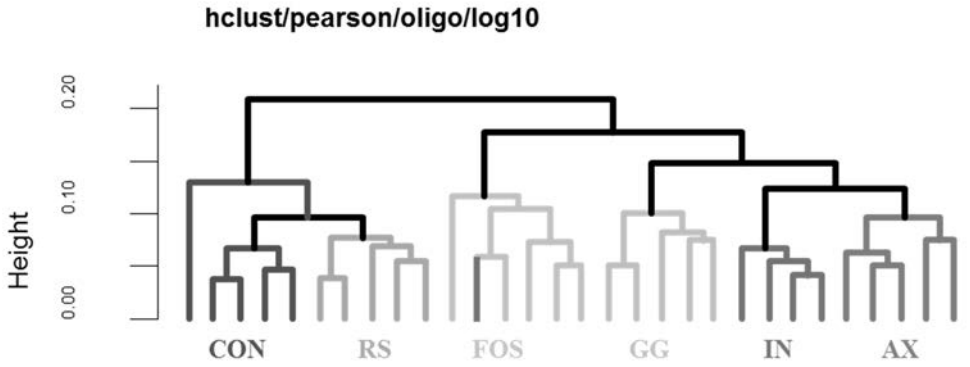


Figure S 6.2 Pearson distance clustering of the samples on log10 transformed probe level data of the MITChip. Abbreviations are for control (CON), Resistant Starch (RS), Arabinoxylan (AX), Fructooligosaccharides (FOS), Inulin (IN) and Guar Gum (GG).

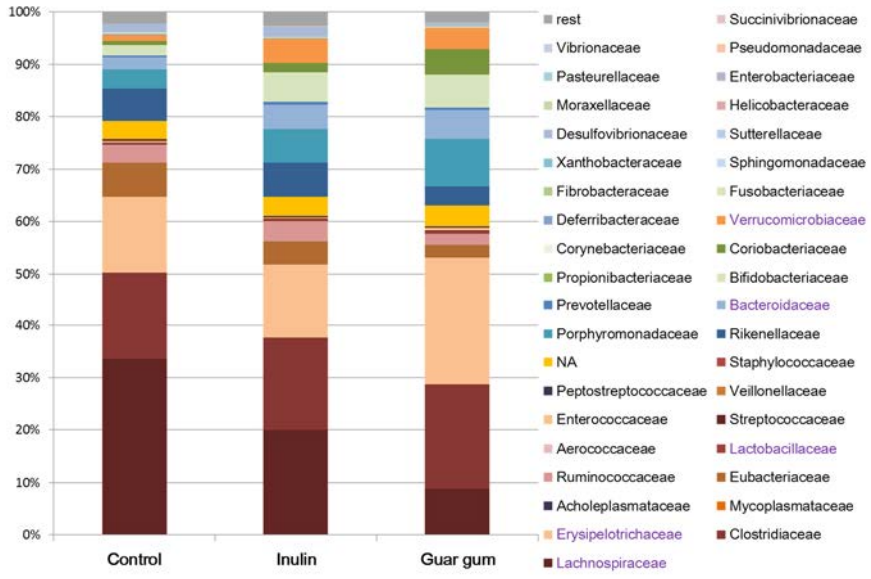


Figure S 6.3 Relative abundance of metatranscriptome (activity) at family level. Family names depicted in purple are significantly different in activity between control and Guargum groups.

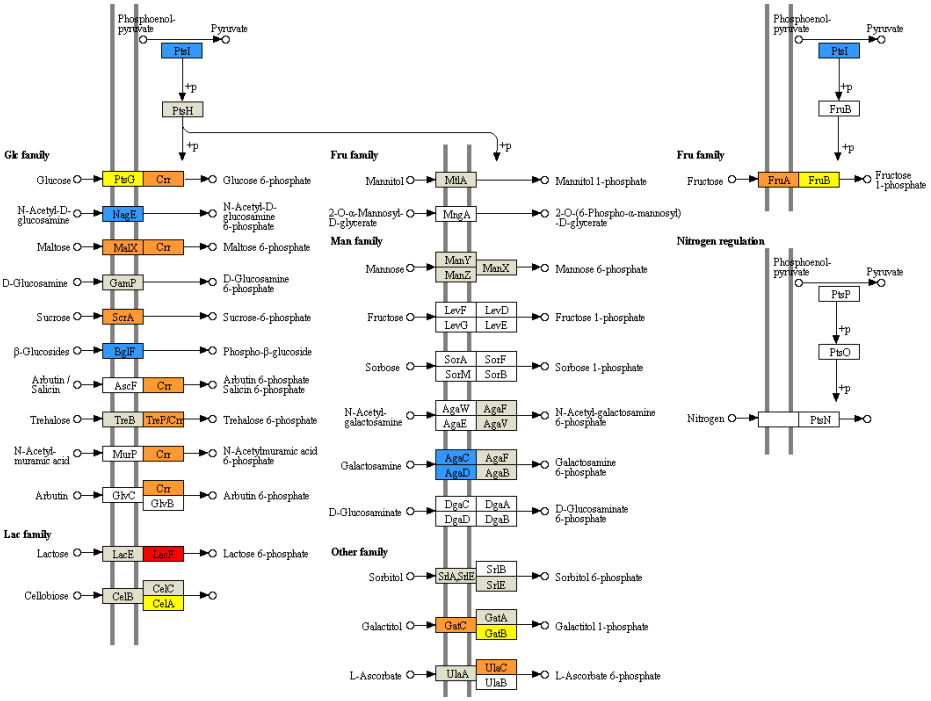


Figure S 6.4 Expressed Phosphotransferase systems (PTS) in the metatransciptome. Uncoloured units were not found in the data. Grey, equally expressed in all groups; orange, higher expression in IN and GG; blue, lower expression in IN and GG; yellow, higher expression in GG compared to CON; red, higher expression in IN compared to CON.

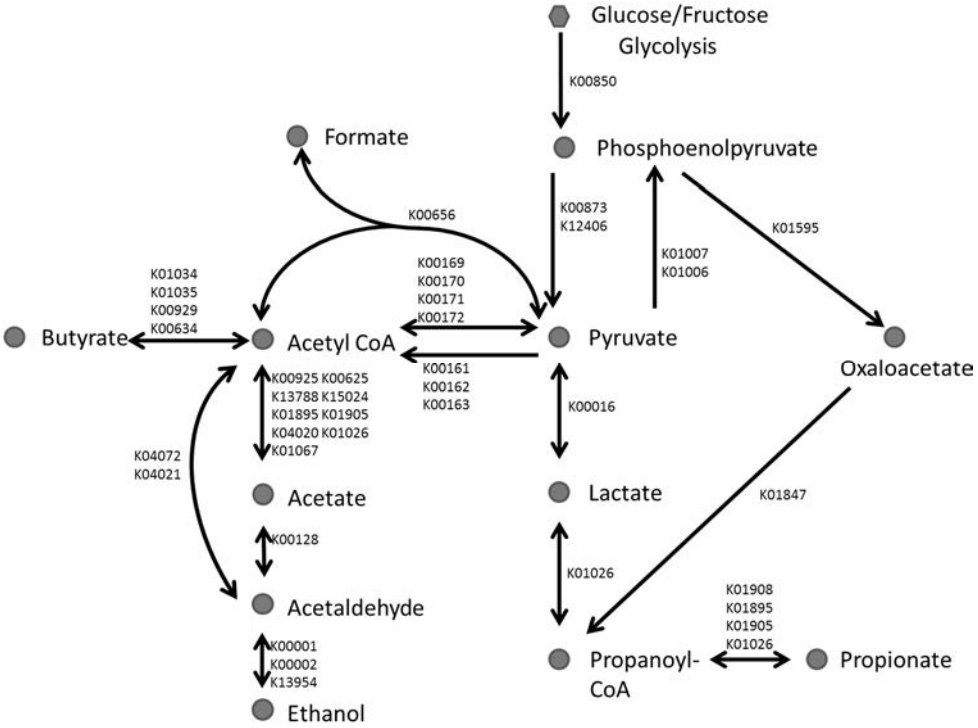


Figure S 6.5 KEGG numbers of the SCFA metabolism pathways used to create Figure 6.9.

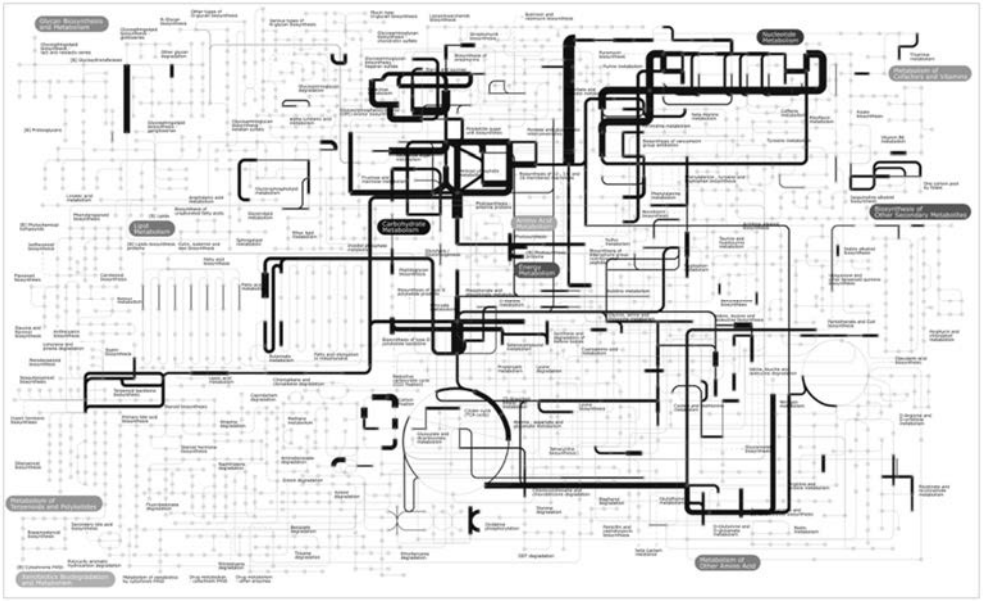


Figure S 6.6 iPATH visualization of the expression levels in the metabolic processes assigned to Erysipelotrichaceae.

Supplemental Tables

	Nr of Keggs		Total Kegg
	More expressed in IN and GG	*Less expressed in IN and GG*	
Metabolism	**95**	**87**	**70.5%**
Carbohydrate metabolism	**33.1%**	**38.6%**	**20.9%**
Glycolysis / Gluconeogenesis	5	4	
Citrate cycle (TCA cycle)	4	-	
Pentose phosphate pathway	1	3	
Pentose and glucuronate interconversions	3	2	
Fructose and mannose metabolism	2	6	
Galactose metabolism	7	4	
Ascorbate and aldarate metabolism	1	1	
Starch and sucrose metabolism	2	-	
Amino sugar and nucleotide sugar metabolism	6	9	
Pyruvate metabolism	2	1	
Glyoxylate and dicarboxylate metabolism	2	1	
Propanoate metabolism	1	1	
Butanoate metabolism	4	2	
C5-Branched dibasic acid metabolism	1		
Energy metabolism	**8.9%**	**6.7%**	**10.3%**
Oxidative phosphorylation	3	-	
Carbon fixation in photosynthetic organisms	2	1	
Carbon fixation pathways in prokaryotes	-	1	
Methane metabolism	-	3	
Nitrogen metabolism	2	-	
Sulfur metabolism	4	3	
Lipid metabolism	**6.5%**	**4.2%**	**3.8%**
Fatty acid biosynthesis	-	1	
Fatty acid degradation	-	1	
Synthesis and degradation of ketone bodies	-	1	
Sphingolipid metabolism	2	1	
Linoleic acid metabolism	1	-	
Biosynthesis of unsaturated fatty acids	-	1	
Primary bile acid biosynthesis	1	-	
Glycerolipid metabolism	3	-	
Glycerophospholipid metabolism	1	-	
Nucleotide metabolism	**6.5%**	**5.0%**	**7.4%**
Purine metabolism	7	3	
Pyrimidine metabolism	1	3	
Amino acid metabolism	**10.5%**	**7.6%**	**11.9%**
Alanine, aspartate and glutamate metabolism	2	-	
Glycine, serine and threonine metabolism	1	1	

Cysteine and methionine metabolism	2	2	
Valine, leucine and isoleucine biosynthesis	1	-	
Valine, leucine and isoleucine degradation	-	1	
Lysine biosynthesis	1	-	
Lysine degradation	-	1	
Arginine and proline metabolism	4	1	
Phenylalanine, tyrosine and tryptophan biosynthesis	2	-	
Phenylalanine metabolism	-	2	
Tryptophan metabolism	-	1	
Metabolism of other amino acids	**1.6%**	**1.7%**	**2.1%**
Selenocompound metabolism	1	1	
D-Alanine metabolism	1	-	
beta-Alanine metabolism	-	1	
Glycan biosynthesis and metabolism	**2.4%**	**5.0%**	**3.2%**
N-Glycan biosynthesis	-	1	
Various types of N-glycan biosynthesis	-	2	
Glycosaminoglycan degradation	-	1	
Glycosphingolipid biosynthesis - globo series	1	-	
Peptidoglycan biosynthesis	1	-	
Other glycan degradation	1	2	
Metabolism of cofactors and vitamins	**3.2%**	**5.9%**	**6.2%**
Nicotinate and nicotinamide metabolism	1	-	
Porphyrin and chlorophyll metabolism	3	1	
Pantothenate and CoA biosynthesis	-	4	
Biotin metabolism	-	1	
Ubiquinone and other terpenoid-quinone biosynthesis	-	1	
Metabolism of terpenoids and polyketides	**1.6%**	**0.8%**	**1.9%**
Terpenoid backbone biosynthesis	1	1	
Carotenoid biosynthesis	1	-	
Biosynthesis of other secondary metabolites	**0.8%**	**0.8%**	**1.2%**
Streptomycin biosynthesis	1	-	
Caffeine metabolism	-	1	
Xenobiotics biodegradation and metabolism	**1.6%**	**6.7%**	**1.6%**
Chloroalkane and chloroalkene degradation	1	-	
Bisphenol degradation	1	-	
Benzoate degradation	-	3	
Xylene degradation	-	1	
Atrazine degradation	-	1	
Dioxin degradation	-	1	
Drug metabolism - other enzymes	-	2	
Genetic Information Processing	**4**	**5**	**13.2%**
Transcription	**0%**	**0.8%**	**0.3%**
RNA polymerase	-	1	

Translation	**0.8%**	**2.5%**	**6.0%**
Aminoacyl-tRNA biosynthesis	1	1	
Ribosome	-	2	
Folding, sorting and degradation	**0.8%**	**0.8%**	**2.4%**
RNA degradation	1	-	
Protein processing in endoplasmic reticulum	-	1	
Replication and repair	**1.6%**	**0%**	**4.5%**
Nucleotide excision repair	1	-	
Homologous recombination	1	-	
Environmental Information Processing	**23**	**21**	**11.8%**
Membrane transport	**11.3%**	**11.8%**	**7.5%**
ABC transporters	8	10	
Phosphotransferase system (PTS)	6	4	
Signal transduction	**7.3%**	**5.9%**	**4.2%**
Two-component system	6	6	
HIF-1 signaling pathway	1	-	
FoxO signaling pathway	1	-	
Phosphatidylinositol signaling system	1	-	
Calcium signaling pathway	-	1	
Cellular Processes	**2**	**6**	**4.5%**
Transport and catabolism	**1.6%**	**0.8%**	**0.9%**
Peroxisome	2	1	
Cell motility	**0%**	**4.2%**	**2.9%**
Bacterial chemotaxis	-	2	
Flagellar assembly	-	3	

Table S 6.1 KEGG orthology distribution of 250 KEGGs explaining the differentiation of the diets in Figure 6.6 best. More expressed in IN and GG are the KEGGs in the direction of the IN and GG groups, less expressed are the KEGGs that correlated to the CON group.

	All Keggs	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Metabolism	**70.5%**	**70.2%**	**81.4%**	**76.0%**	**75.2%**	**61.1%**	**60.5%**	**34.8%**
Carbohydrate metabolism	20.9%	18.4%	24.6%	22.4%	21.4%	20.4%	24.2%	2.2%
Energy metabolism	10.3%	4.3%	18.2%	7.3%	10.3%	18.5%	5.6%	8.7%
Lipid metabolism	3.8%	4.0%	3.0%	2.8%	0.9%	5.7%	8.1%	0.0%
Nucleotide metabolism	7.4%	9.4%	5.5%	7.7%	13.7%	0.0%	4.8%	**19.6%**
Amino acid metabolism	11.9%	**18.4%**	12.7%	12.6%	11.1%	4.5%	8.1%	0.0%
Metabolism of other amino acids	2.1%	2.0%	1.7%	2.4%	0.9%	2.5%	3.2%	2.2%
Glycan biosynthesis and metabolism	3.2%	3.3%	2.5%	5.7%	5.1%	1.9%	0.0%	0.0%
Metabolism of cofactors and vitamins	6.2%	5.7%	9.3%	9.3%	4.3%	2.5%	3.2%	2.2%
Metabolism of terpenoids and polyketides	1.9%	1.7%	1.3%	1.6%	3.4%	2.5%	2.4%	0.0%
Biosynthesis of other secondary metabolites	1.2%	1.7%	0.8%	1.6%	1.7%	1.3%	0.0%	0.0%
Xenobiotics biodegradation and metabolism	1.6%	1.3%	1.7%	2.4%	2.6%	1.3%	0.8%	0.0%
Genetic Information Processing	**13.2%**	**18.4%**	**10.2%**	**7.7%**	**4.3%**	**14.6%**	**6.5%**	**60.9%**
Transcription	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%	6.5%
Translation	6.0%	1.7%	5.9%	2.8%	1.7%	11.5%	4.8%	47.8%
Folding, sorting and degradation	2.4%	2.0%	3.4%	2.4%	2.6%	3.2%	0.8%	6.5%
Replication and repair	4.5%	14.7%	0.8%	2.4%	0.0%	0.0%	0.0%	0.0%
Environmental Information Processing	**11.8%**	**9.7%**	**7.6%**	**14.2%**	**18.8%**	**9.6%**	**19.4%**	**2.2%**
Membrane transport	7.5%	6.0%	5.1%	8.5%	17.1%	5.1%	9.7%	2.2%
Signal transduction	4.2%	3.7%	2.5%	5.7%	1.7%	4.5%	9.7%	0.0%
Cellular Processes	**4.5%**	**1.7%**	**0.8%**	**2.0%**	**1.7%**	**14.6%**	**13.7%**	**2.2%**
Transport and catabolism	0.9%	1.3%	0.0%	0.8%	0.9%	1.9%	0.0%	2.2%
Cell motility	2.9%	0.0%	0.0%	0.4%	0.0%	**11.5%**	**12.9%**	0.0%
Cell growth and death	0.7%	0.3%	0.8%	0.8%	0.9%	1.3%	0.8%	0.0%

Table S 6.2 KEGG category distribution of KEGGs from the different clusters in the microbial expressed functions and mucosal caecal gene expression.

Chapter 7

Fermentation of dietary milk protein by murine gut microbiota.

Floor Hugenholtz*, Mark Davids*, Jessica Schwarz, Michael Müller, Daniel Tomé, Peter J. Schaap, Guido JEJ Hooiveld, Hauke Smidt and Michiel Kleerebezem

*These authors contributed equally

Abstract

The undigested carbohydrates and proteins fractions of the food are converted by the microbiota into a large range of metabolites, of which short chain fatty acids (SCFA) are the most abundant. These microbial metabolites are subsequently available for absorption by the host mucosa and can serve as an energy source. Amino acids derived from dietary proteins can also serve as substrates for SCFA formation, leading to expansion of the fermentation end-product diversity, including branched-SCFA. So far metabolic networks related to protein fermentation were documented in *in vitro* models. Here the long-term effects of high protein-diets on microbial community composition and functionality were analysed. To this end, determinations of the microbiota composition using phylogenetic microarray (MITChip) technology were complemented with metatranscriptome and metabolome analyses to obtain further insights in the protein fermentation pathways and associated groups.

Introduction

Components of our daily food such as fibres and a part of our dietary protein are not efficiently digested by us, leaving that fraction of our diet available for digestion by the intestinal microbiota. These food ingredients proceed toward the large intestine where they are converted by the microbiota into a large range of metabolites, of which short chain fatty acids (SCFA) are the most abundant. These microbial metabolites are subsequently available for absorption by the host mucosa and can serve as an energy source.

Approximately ten grams of protein reach the human colon daily (Chacko and Cummings 1988), which include both host (proteases from the pancreas and mucins) as well as dietary proteins. The gut microbiota has a high proteolytic capacity and ferments the proteins into SCFA, branched chain fatty acids (BCFA), ammonia and phenolic and indolic compounds (Macfarlane *et al.* 1988, Gibson *et al.* 1989, Smith and Macfarlane 1997, Windey *et al.* 2012b). These BCFA are generated by branched-chain amino acid catabolism, i.e., the degradation of valine, leucine and isoleucine (Macfarlane *et al.* 1986), while the phenolic and indolic compounds are degradation products of aromatic amino acids. Bacterial degradation of sulfur containing amino acids, including methionine, cysteine, and taurine, results in the formation of the hydrogen sulfide. At present there is contradicting evidence concerning the potentially damaging effects in the large intestine elicited by high protein intake (Windey *et al.* 2012a, Windey *et al.* 2012b). However, the source of the protein seems to be an important factor for colonic damage, as animal protein has a more pronounced effect than plant derived protein. For example the risk of colorectal cancer and inflammatory bowel disease are correlated with red meat intake, although white meat, like fish, is not associated with these diseases (Norat *et al.* 2010, Windey *et al.* 2012b). In contrast whey protein reduced body weight gain and increased insulin sensitivity in rats (Belobrajdic *et al.* 2004).

Nowadays, the consumption of diets that contain high protein levels is quite common and has been proposed to support body weight reduction (Lacroix *et al.* 2004, Westerterp-Plantenga *et al.* 2009, Westerterp-Plantenga *et al.* 2012). With respect to the effect of high protein dietary intake on the composition of the gut microbiota, it has been shown that long term consumption of high levels of protein and animal fat are associated with the *Bacteroides* enterotype (Wu *et al.* 2011, Zoetendal and de Vos 2014), but it should be noted that so far there is no confirmation that the high protein content in the diets directly leads to this enterotype. In rats a short term (2 weeks) high protein dietary intervention did change the microbial community, where *Clostridium coccoides, Clostridium leptum* and *Faecalibacterium prausnitzii* decreased in abundance due to the intervention (Liu *et al.* 2014).

In a previous study the effect of long-term high protein (HP) diets was studied in a mouse model, and could be shown to result in a lower body weight, reduced adiposity and hepatic lipid accumulation (Schwarz *et al.* 2012). However, to date little is known about the effect of this type of high-protein diet on the gut microbiota

community composition and activity. Here we describe the effects on the gut microbiota community and activity on long-term high-protein dietary interventions in mice, using 16S rRNA gene-targeted profiling and metatranscriptome approaches, to unravel patterns of activity within the microbial ecosystem residing in the caeca of the mice.

Results & Discussion

Long-term effects of high protein diets on composition and activity of the gut microbiota were assessed in a 12-week dietary intervention study in mice. The effects of high protein was studied both in a low- and high-fat dietary background, to evaluate whether the fat content of the diet affects the outcome of the high protein intervention. Male C57BL/6J mice of 10 weeks (young adults) were given the control, a normal protein and low fat diet, for two weeks, followed by the 12 week dietary intervention. Four groups of mice were selected (n=10 per group), for the four diets: the control diet (normal protein low fat, NPLF), a normal protein high fat diet (NPHF), a high protein low fat diet (HPLF) and a high protein high fat diet (HPHF) (Table 7.1). The effects of these dietary interventions were measured by determination of the fermentation output, measured by luminal SCFA levels, the microbiota composition and activity at DNA and mRNA levels, respectively.

	NPLF	NPHF	HPLF	HPHF
	(g/kg) dry matter			
Milk protein	140	160	484	580
Corn starch	361.35	291.3	189.35	80
Sucrose	361.35	291.4	189.35	80
Soybean oil	40	40	40	40
Palm oil	0	120	0	123
Minerals	35	35	35	35
Vitamins	10	10	10	10
a-Cellulose	50	50	50	50
Choline	2.3	2.3	2.3	2.3

Table 7.1 The four diets. NPLF: Normal protein low fat. NPHF Normal protein high fat. HPLF: High protein low fat. HPHF: high protein high fat (Schwarz et al. 2012).

Dietary proteins differentially modulate luminal SCFA levels

The main fermentation metabolites of dietary protein are short chain fatty acids, predominated by acetate, propionate and butyrate. During protein fermentation also branched-chain SCFA are formed from the degradation of branched amino acids, and these metabolites were measured to obtain an indication whether the protein fermentation increased due to the dietary intervention. Gas chromatography was used to measure concentrations of acetate, propionate, butyrate, valerate and the branched-chain SCFA, iso-butyrate and iso-valerate in caecal luminal content of mice receiving the different diets.

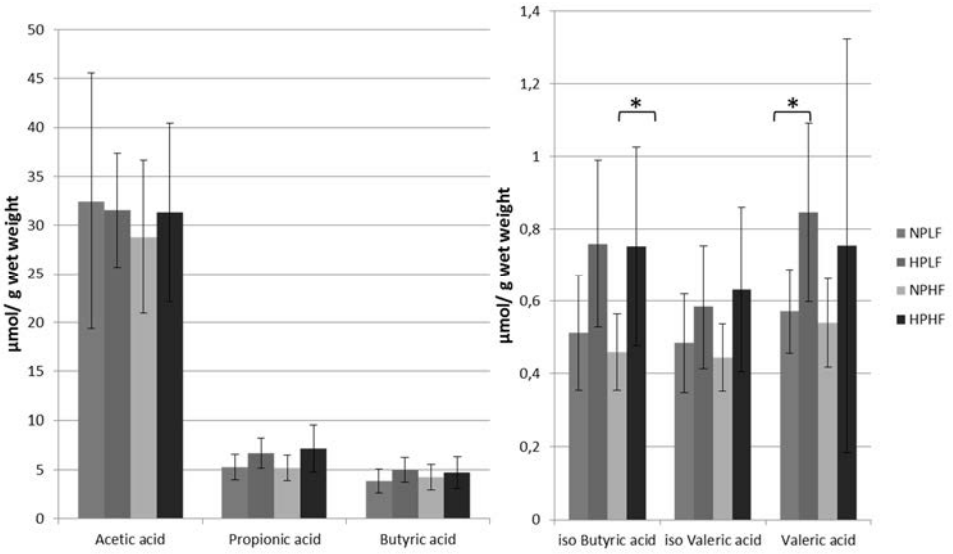


*Figure 7.1 Caecal luminal SCFA concentrations in µmol/g content measured with gas chromatography. * Indicates significance between two groups (Ttest, $p < 0.05$).*

The high protein diets led to an apparent increase of the SCFA and branched-chain SCFA concentrations in the caecum (Figure 7.1). However due to the high variation between individual mice, only a significant increase could be detected in the concentration of iso-butyrate in the HPLF group relative to the HPHF group, while valeric acid was significantly increased in the HPLF group compared to the NPLF group. Total SCFA concentrations tended to be higher (not significant) in caecal luminal samples obtained from mice that were fed the HPLF (data not shown). Evaluation of overall trends in these luminal metabolites indicated that the high protein levels in the diet resulted in a higher concentration of these fermentation products, whereas high fat diets were associated with slightly reduced concentrations. However, in these analyses it should be taken into account that the relative amount of corn-starch in the diets was drastically decreased in diets with increased relative protein and fat concentrations. Notably, this observation implies that despite the reduced corn-starch availability for microbial fermentation in the HPHF diet, the microbiota still generated higher overall SCFA concentrations, which are likely derived from protein fermentation, which has been proposed to account for up to 17% of the overall SCFA production in the caecum (Macfarlane, 1992). Apparently high level protein fermentation by the microbiota supports higher SCFA concentrations in the caecal lumen as compared to fermentation of the alternative nutrients (e.g. corn-starch).

Dietary proteins modulate the microbiota composition

Intestinal content of four mice per dietary treatment was used to analyse the microbiota composition after 12 weeks in the ileum, caecum and colon, using the

MITChip platform, a 16S rRNA targeted phylogenetic microarray designed for the comprehensive and deep profiling of mouse intestinal microbiota composition (Rajilic-Stojanovic *et al.* 2009, IJssennagger *et al.* 2012, El Aidy *et al.* 2013a, Everard *et al.* 2013). MITChip analysis revealed clearly distinct microbiota composition profiles in the ileum as compared to those obtained for caecal and colonic content. Notably, the ileal microbiota appeared to be unaffected by the dietary intervention, whereas the caecum and colon microbiota profiles clustered closely together and sub-clustered according to the diet (Figure S 7.1). The caecum is considered as the intestinal region where most prominent microbiota fermentation takes place. Moreover, this region of the intestine allowed the extraction of an amount of intestinal content that was sufficient for RNA extractions that are compatible with metatranscriptome analysis. Therefore, our analyses focused on the metatranscriptome analysis of the microbiota residing in this intestinal region as described below.

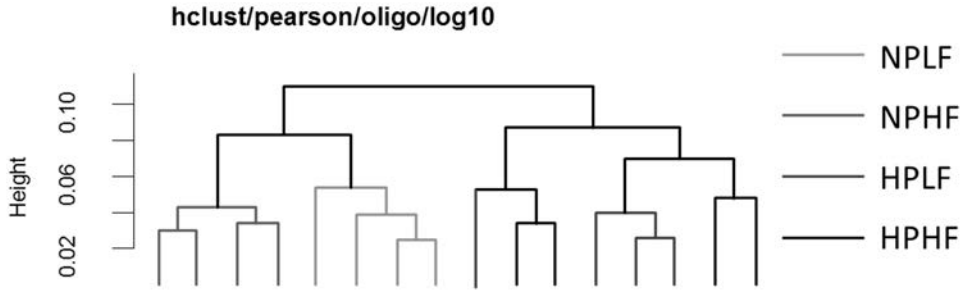


Figure 7.2 Cluster of the caecal samples on log10 transformed probe level data of the MITChip. The clustering was made using Pearson distance.

MITChip analyses revealed distinct microbiota composition profiles in the caecum from animals that were fed the normal protein (NPLF and NPHF) or high protein content (HPLF and HPHF) diets (Figure 7.2). Notably, within the NP diets the fat level resulted into distinct clustering of the microbiota from mice on the low (NPLF) and high (NPHF) fat content diets, whereas the microbiota profiles failed to discriminate the HP diets on basis of their fat-content. This finding illustrates that within the NP diet context the other main nutritional component (i.e., fat content) has a prominent influence on the microbiota, while this effect is lost or overruled by the high protein content in the HP diet context.

Microbiota and fermentation-metabolite data integration

In order to relate changes in caecal microbiota composition to the different diets, the weight of the mice, as well as to SCFA as the main metabolites of microbial fermentation, hybridization signals of in total 96 genus-level phylogenetic groups were subjected to redundancy analysis (RDA). The RDA included the concentrations of acetate, propionate, butyrate, iso-butyrate, valerate, iso-valerate, the weight of the mice and the diets as explanatory variables. Overall, these explanatory variables

accounted for 85.8% of the total variation, 57.5% of which was covered by the first two canonical axes (Figure 7.3). All the diets clustered separately on these canonical axes. The LFNP, HFNP and HFHP diets as explanatory variables had a significant (Monte Carlo Permutation test, $p<0.05$) impact on the total variation of the data. The genus like microbial groups in the analyses that correlated with the LFNP diet belonged to the *Bifidobacterium*, *Lactobacillus delbrueckii* et rel., *Lactobacillus plantarum* et rel., *Lactobacillus acidophilus* et rel., *Lactobacillus gasseri* et rel., and *Ruminococcus obeum* et rel. Notably, these groups also correlated with the acetate concentration that was measured in the caecum content. The higher abundances of these typical saccharolytic bacterial groups implies that in the mice on the LFNP diet, the higher relative amount of the carbohydrates in this diet, i.e., corn-starch and sucrose, are incompletely (digested and) absorbed in the small intestine, and thus available for microbial fermentation in the caecum. *Ruminococcus obeum* can use a wide range of sugars (Liu *et al.* 2008) and the lactobacilli are known for their rapid sugar import and metabolism (Turroni *et al.* 2014). The HFNP diet as well as the body-weight of the mice strongly correlated with higher abundances of several groups within *Clostridium* cluster XIVa. Inversely, *Akkermansia muciniphila* was anti-correlated with the HFNP diet and body-weight, which is notable since the anti-correlation of this microbial group with body-weight has also been reported (Everard *et al.* 2013, Cani and Everard 2014). Samples taken from animals fed the HP diets grouped closely together and correlated with elevated levels of branched SCFA as well as with a higher abundance of *Parabacteroides distasonis* et rel.

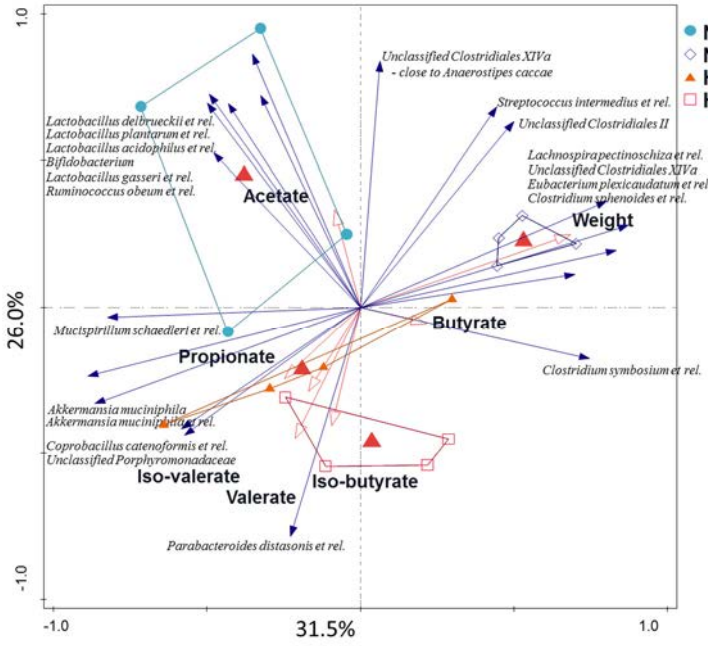


Figure 7.3 Redundancy analysis. Explanatory variables are the weight of the mice, acetate, propionate, butyrate, valerate, iso-butyrate and iso-valerate. Explanatory variables account for 85.8 % of the total variation.

Effect of high protein and high fat on overall microbial metatranscriptome patterns

The activity profiles of the microbiota obtained from the caecum of mice were determined by metatranscriptome analysis in each of the diet-groups at the end of the dietary intervention period. To this end, the caecal contents of four mice of the LFHP, HFHP, HFNP and three mice of the LFNP group were used for RNA extraction, mRNA enrichment, cDNA synthesis and illumina metatranscriptome sequencing. The sequencing efforts generated 13.4-34.9 million reads per sample, with a single outlier that generated 177.0 million reads. These reads were filtered for non-mRNA and low quality reads, resulting in between 3.4 - 19.2 x 10^5 (with the outlier at 56 x 10^5) of mRNA derived sequence reads. To determine the function and taxonomy of these reads the mRNA fractions were merged and *de novo* assembled into larger contigs using the pipeline described previously (**Chapter 5**), creating one reference set for all samples. A total of approximately 38 thousand contigs could be assembled with an overall length of 29.8 x 10^6 bases (n50=945). These contigs encoded a total of close to 5.5 x 10^4 predicted open reading frames. The taxonomic origin of the contigs was analysed by alignment of the predicted protein sequences with NCBI's non-redundant database, retrieving the taxonomic family classification of the protein sequence with the highest similarity. Functional annotation was performed by assignment of KEGG orthology identifiers to all predicted protein sequences, using the KEGG KAAS server. Expression levels of individual genes were determined by aligning the mRNA reads with the assembled protein-encoding contigs and enumerating the total amount of nucleotides aligned with the corresponding ORFs. Between 54% and 74% of the mRNA reads could be assigned to the predicted protein-encoding transcripts (Table 7.2). The expression levels of each of the protein encoding regions in each of the samples were normalized by scaling each gene by the total number of nucleotides mapped to ORFs of that same sample. Clustering of the normalized expression dataset revealed a clustering of the different samples that resembled the clustering of the MITChip derived microbiota composition profiles (Figure S 7.2). This indicates that analogous to microbial composition, the microbiota activity profiling enabled the detection of the impact of the dietary fat content in the NP diets, whereas this effect of the fat content appeared to be lost or overruled by the high protein content in the HP diets.

Sample name	Total reads	mRNA	Assembled mRNA reads	Bacterial protein coding in assembled contigs
HPLF_3	3.49E+07	7.14E+05	73.7%	83.1%
NPLF_2	3.29E+07	1.58E+06	66.2%	83.1%
NPLF_3	3.11E+07	1.92E+06	64.0%	88.6%
HPLF_4	1.93E+07	3.83E+05	64.7%	80.2%
HPLF_1	1.81E+07	4.02E+05	76.4%	78.2%
NPHF_1	1.40E+07	4.48E+05	62.1%	83.7%
NPLF_1	1.61E+07	7.28E+05	62.6%	80.1%
HPHF_1	2.49E+07	6.38E+05	73.2%	73.0%
HPHF_2	1.77E+08	5.58E+06	73.3%	82.0%
HPHF_3	1.34E+07	3.37E+05	75.8%	75.5%
NPHF_2	1.83E+07	4.03E+05	68.0%	74.9%
NPHF_3	2.35E+07	5.60E+05	72.2%	59.7%
HPLF_2	3.29E+07	6.78E+05	75.0%	70.6%
NPHF_4	2.40E+07	5.99E+05	64.7%	65.2%
HPHF_4	1.77E+07	5.78E+05	80.5%	53.1%

Table 7.2 Reads of the illumina sequences and the result of data processing.

Effect of high protein and high fat diets on microbiota function profiles

To focus only on functions that are differentially expressed within the microbiota as a function of the different diets, we employed an in house R script to detect KEGG modules that are differentially expressed. Remarkably, the different fat levels in the diets (HF versus LF) could not be correlated to KEGG modules that were differentially expressed. In contrast, the samples derived from NP and HP diet fed mice enabled the detection of KEGG modules that displayed significant differential expression levels. The KEGG modules enriched in the NP diet derived samples were all associated with sugar metabolism, whereas the modules enriched in the HP diet derived samples were consistently associated with protein metabolism (Table 7.3). These findings are in good agreement with the clustering analyses as well as the predicted dietary impacts on the nutrients available for fermentation in the caecum of the mice that were fed on the different diets.

	Module	nr of KEGGs in modules	nr of KEGGs found	Module explanation
Enriched in NP	M00377	10	4	Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) [PATH:map01200 map00720]
	M00422	5	3	Acetyl-CoA pathway, CO2 => acetyl-CoA [PATH:map00680]
	M00196	4	3	Multiple sugar transport system [PATH:map02010]
Enriched in HP	M00018	10	4	Threonine biosynthesis, aspartate => homoserine => threonine [PATH:map01230 map00260]
	M00299	4	3	Spermidine/putrescine transport system [PATH:map02010]
	M00236	3	3	Putative polar amino acid transport system [BR:ko02000]

Table 7.3 Enriched modules within the NP or HP dataset.

Effect of high protein and high fat on active microbial community

To further support an eventual role of microbial groups in protein fermentation, the metatranscriptome datasets were screened for genes involved in proteolysis (Figure 7.4a), amino acid metabolism (Figure 7.4b), and amino acid transport transport (Figure 7.4c) based on the kegg orthology annotations. The microbial families *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae* were all dominantly to the overall activity associated with the degradation of proteins, supporting their high relative contribution in the total mRNA activity profile in the samples from HP diets (Figure S 7.3). Especially the *Erysipelotrichaceae* displayed a significant increased contribution to the overall protein degradation, which was most significantly detected in samples obtained from mice that were fed the LFHP diet, whereas the trend was also observed in the samples obtained from HFHP-fed mice, albeit not reaching significance. On the contrary *Lachnospiraceae* tends to decrease its contribution to the overall protein degradation, where the expression of peptidases in the HPLF diet is significantly lower than in the NPLF diet.

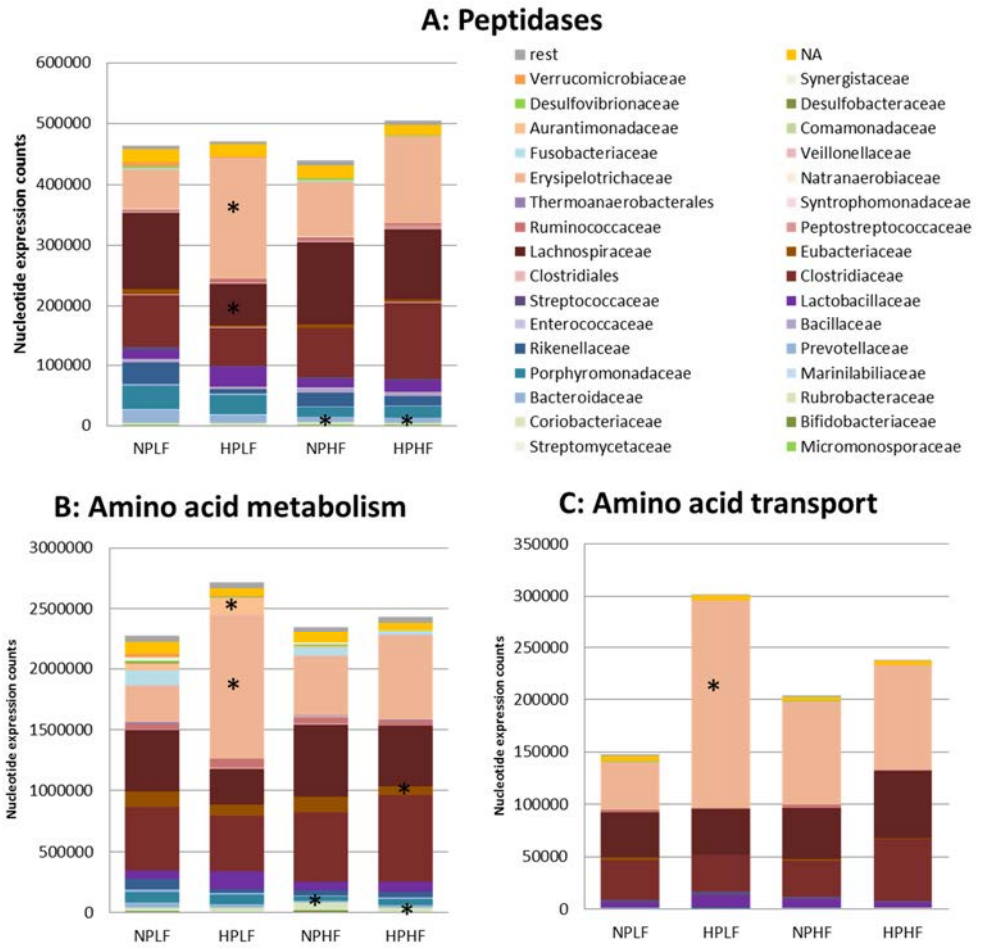


*Figure 7.4 Relative abundance of families expressing peptidases (A), amino acid metabolism related proteins (B) and amino acid transporters (C). * Indicates significance between NPLF and another group (Ttest, $p < 0.05$).*

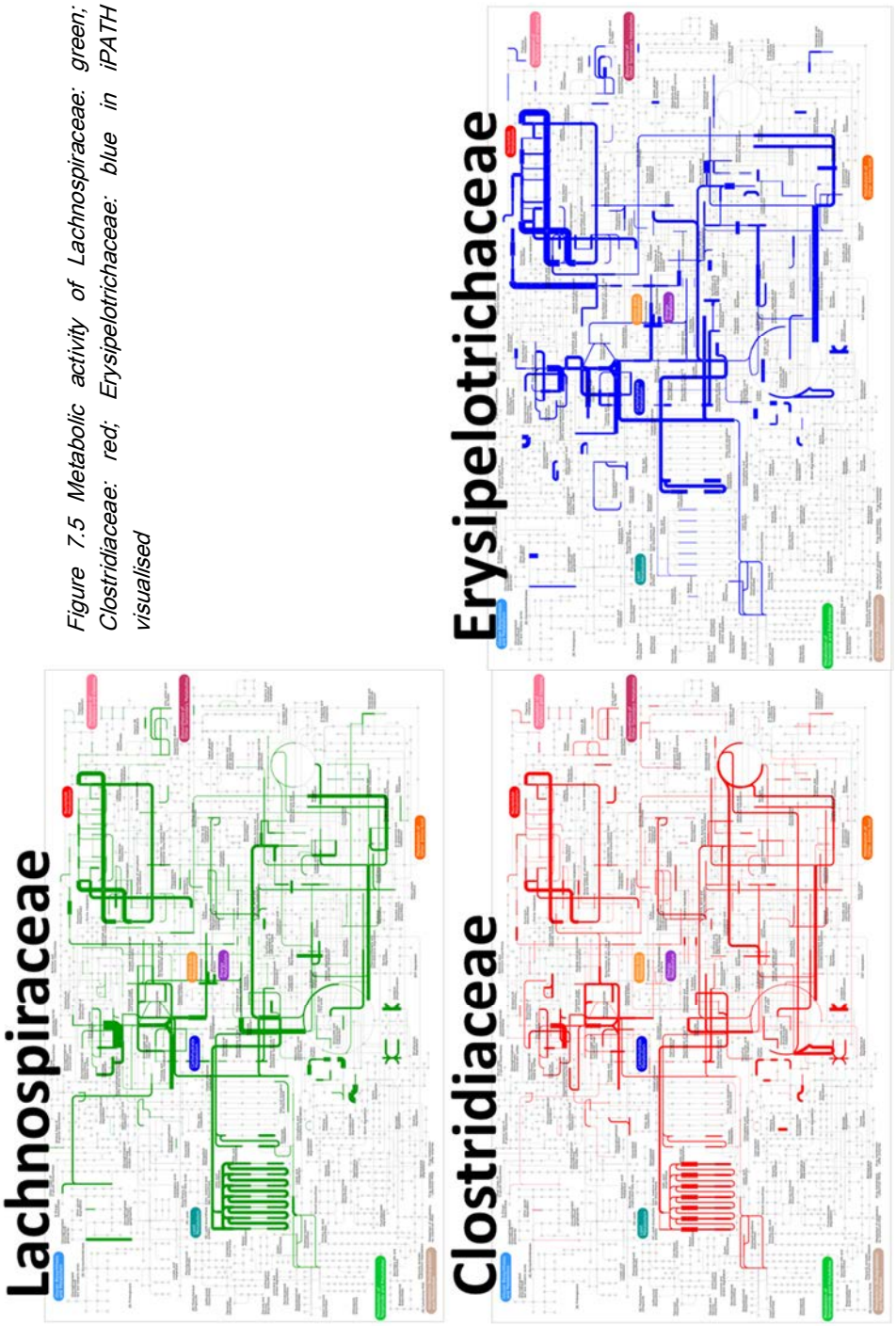


Figure 7.5 Metabolic activity of Lachnospiraceae: green; Clostridiaceae: red; Erysipelotrichaceae: blue in iPATH visualised

In contrast to the observed increased relative abundance of protein degradation related transcripts, one of the glycolysis-associated genes encoding the 6-phosphofructokinase, was less expressed in the HPLF diet, which was mainly due to the decreased expression from *Lachnospiraceae* (Figure S 7.4). The *Lachnospiraceae* seemed less able to use the protein sources and are more dependent on the sugar fermentation in the gut. Confirming earlier observations by Liu *et al.*, 2014, that the *Clostridium coccoides* group, belonging to the *Lachnospiraceae,* decreases due to high protein in the diet (Liu *et al.* 2014). To more closely inspect the specific activity patterns of these three microbial families their specific-activity was plotted on the metabolic map available in the iPATH software suite (Figure 7.5). Remarkably, the specific expression patterns did not seem to differ in the different dietary regimes for the three predominantly active microbial families (data not shown). Nevertheless, each of the microbial families displayed quite distinct expression patterns. As an example, *Lachnospiraceae*, strongly expressed genes encoding enzymes involved in the conversion of phosphoenolpyruvate to oxaloacetate, and lipid biosynthesis activity, whereas the *Erysipelotrichaceae* hardly displayed these activities. In turn, the *Erysipelotrichaceae* appeared to be much more focused on the conversion of malate, fumarate and succinate. Finally, the predominant *Clostridiaceae* representatives were concluded to express both these activities, as well as a broader spectrum of pathways related to amino acid metabolism. Notably, each of the iPATH mapped pathways for the activity patterns of the families *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae* appeared to display gaps, suggesting incomplete or incorrect annotations of genes in the metatranscriptome data. This may either be due to incorrect functional annotation due to taxonomical misclassification of the predicted proteins or low expression levels. Inaccurately annotated genomes in the NCBI database or substantial dissimilarity between the reference genomes and the protein transcription data obtained here that are derived from bacterial species residing in the murine intestinal tract. However the depth of our metatranscriptome analysis could also be insufficient to detect these complete pathways.

Overall, in animals fed the HPLF diet, the *Erysipelotrichaceae* family appeared to have an advantage over the other families, as the gene expression data suggest a focus on pathways for the degradation of a selected set of amino acids. In the HPHF, possibly due to the higher fat content, the advantage of the *Erysipelotrichaceae* appeared to be reduced relative to the LF diets and appeared to be at least partly taken over by members of the *Clostridiaceae*, which appeared to express a wider range of amino acid catabolic pathways.

The observed expression profiles of genes encoding enzymes involved in SCFA production, suggest that *Erysipelotrichaceae* produced predominantly acetate as the main end product of protein catabolism, whereas *Clostridiaceae* and *Lachnospiraceae* were predicted to produce both acetate and butyrate (Figure S 7.5). Notably, also the *Erysipelotrichaceae* appeared to express genes that code for enzymes involved acetyl-coA to butyryl-coA conversion, using the crotonyl-coA

pathway, which allows anaerobic microbes to conserve energy (Seedorf *et al.* 2008). This may imply that also this family contributes to butyrate production. However, expression of genes encoding the butyrate kinase and the butyryl-CoA:acetate CoA-transferase enzymes involved in this pathway could not be detected in the *Erysipelotrichaceae* associated transcriptional activity patterns, which could be due to erroneous annotations of acetate kinase and other SCFA transferase functions in this group, since the sequences of these enzymes are known to be highly similar (Louis and Flint 2007, Vital *et al.* 2013, Vital *et al.* 2014). Genes that encode the enzymes required for propionate production appeared barely expressed, and were exclusively assigned to the bacterial families *Porphyromonadaceae* and *Sphingomonadaceae*. The very low expression detected for the propionate production pathway may indicate that for a more complete reconstruction of the microbiome activity profiles, metatranscriptome datasets with a higher depth of analysis would be required. Analogously, we failed to detect the expression of genes encoding enzymes involved in branched SCFA production, which may also require a higher depth of metatranscriptome analysis considering that concentrations of these metabolites were two – three orders of magnitude lower than those observed for acetate, propionate and butyrate (Figure S 7.5), but some of enzymes involved in this process might not have classified in the KEGG system. Nevertheless, the enriched modules associated with amino acid metabolism in the metatranscriptome data establishes the role of microbiota in the fermentation of the increased dietary protein levels, while the unaltered pattern of specific activities assigned to *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae* under the different dietary regimes indicates that their specific contribution to the *in situ* protein catabolic activity remains the same despite the substantial differences in protein content of the respective diets.

Conclusion

Here we show that extended feeding of high protein level diets for a period of 12 days exerted a prominent effect on the composition of caecal microbiota and its protein fermentation capacity, supporting elevated SCFA production in the caecal lumen as compared to fermentation of the alternative nutrients (i.e., corn-starch or fat). In addition, the data also revealed a prominent influence on the microbiota composition of the fat content in diets that contain normal protein levels. The microbial community members most abundant in the diets belonged to the families of the *Lachnospiraceae*, *Erysipelotrichaceae* and *Clostridiaceae*, and were all predicted to produce mainly acetate and butyrate based on observed metatranscriptome profiles. The relative activity of especially the *Erysipelotrichaceae* appeared to be increased in mice consuming the high protein diets, although the *Clostridiaceae* appeared to express a wider range of different amino acid metabolism associated pathways. A more complete reconstruction of the microbiome activity profiles is necessary to provide a comprehensive understanding of the role of *Erysipelotrichaceae* and *Clostridiaceae* in protein fermentation *in situ* in the caecal fermentation of dietary proteins. Such improved understanding would be strongly facilitated by *in vitro* pure and defined mixed culture studies using representatives of these microbial families to better characterize and more completely identify their genetic repertoire involved in these pathways to enhance the accuracy of the metatranscriptome mapping to the pathways involved. Moreover, improved understanding of the precise role of these microbial groups in protein catabolism in the intestine is likely to require metatranscriptome datasets with a higher depth of analysis.

Methods

Mice and diets

Male C57BL/6J mice (age 8 weeks) were purchased from Charles River (L'Arbresle, France) and were housed in the animal facility of the Wageningen University. The mice were divided into four groups of 20 animals and housed in pairs in light and temperature-controlled animal housing facilities (12/12 (light/dark), 20°C). The mice had free access to food and tap water. During the first two weeks of the study all mice received the same diet, containing (in %w/totalw) casein (14), corn starch (36.1), sucrose (36.1), soy oil (4), mineral mixture (3.5), vitamin mixture (1), cellulose (5) and choline (0.23). This control diet (NPLF) was given to one group of mice during the whole experiment. In the other groups the amount of protein, fat and carbohydrates was changed (Table 7.1), the responses of the mice to the dietary interventions was reported previously (Schwarz *et al.* 2012). The mice were sacrificed after 12 weeks of dietary intervention, the intestinal content was collected from the ileum, caecum and colon and snap frozen in liquid nitrogen and stored at -80°C.

Short-chain fatty acid analysis in caecal luminal content

Short chain fatty acids were measured in mouse intestinal samples at section. Luminal content of the caecum (ten mice per group) was collected in H_3PO_4 and isocaproic acid (as an internal standard) containing buffer solution. Samples were stored at -20ºC until further processing. The day of analysis, samples were thawed, centrifuged at 14.000 rpm (5 min), and supernatant was collected and stored at 5ºC. The samples were then subjected to gas chromatography (Fisons HRGC Mega 2, CE Instruments, Milan, Italy) at 190°C using a glass column fitted with Chromosorb 101 with a carrier gas (N2 saturated with methanoic acid).

Microbial community

Metagenomic DNA was extracted from the ileum, caecum and colon samples (4 mice per diet and time point) using the repeated bead beating plus column (RBB+C) method (Yu and Morrison 2004). The microbial population in the intestinal samples were analysed with the Mouse Intestinal Tract Chip (MITChip). This phylogenetic microarray consists of 3,580 different oligonucleotides specific for the mouse intestinal microbiota ((Van den Abbeele *et al.* 2010)Derrien, et al., in preparation). The array targets the V1 to V6 region of 16S rRNA genes of bacteria. The 16S rRNA genes were amplified from 20 ng of intestinal metagenomic DNA with the primers *T7prom*-Bact-27-F and Uni-1492-R (Table 7.4). The PCR products obtained were transcribed, labelled with Cy3 and Cy5 dyes and fragmented. Finally, the samples were hybridized on the arrays at 62.5ºC for 16 hours in a rotation oven (Agilent Technologies, Amstelveen, The Netherlands). After the slides were washed and scanned, data was extracted with the Agilent Feature Extraction software (version 9.1). The data was normalized and analysed using a set of R-based scripts combined

with a custom-designed relational database, which operates under the MySQL database management system.

Primer name	Sequence	Application
T7prom-Bact-27-F	5'-TGA ATT GTA ATA CGA CTC ACT ATA GGG GTT TGA TCC TGG CTC AG–3'	MITChip
Uni-1492-R	5'-CGG CTA CCT TGT TAC GAC-3'	MITChip
PROK1492R	5' -GGW TAC CTT GTT ACG ACT T-3'	QPCR
BACT1369F	5'-CGG TGA ATA CGT TCY CGG-3'	QPCR

Table 7.4. List of primers (Suzuki et al. 2000, Rajilic-Stojanovic et al. 2009, van den Bogert et al. 2011).

RNA extraction, mRNA enrichment, cDNA synthesis and illumina sequencing.
Four intestinal caecum content samples from each dietary treatment were used to analyze the metatranscriptome activity profiles. The RNA was extracted from 0.1-0.2 grams of ceacal content. The content was suspended in 500 µL ice-cold TE buffer (Tris-HCL pH 7.6, EDTA pH 8.0). Total RNA was obtained via the Macaloid-based RNA isolation protocol (Zoetendal *et al.* 2006, Leimena *et al.* 2013) with in addition the use of Phase Lock Gel heavy tubes (5 Prime GmbH, Germany) during the phase separation. The RNA purification was performed using the RNAeasy mini kit (Qiagen, USA), including an on-column DNAseI (Roche, Germany) treatment (Zoetendal *et al.* 2006). The total RNA was eluted in 30 µL ice-cold TE buffer and the RNA quantity and quality were assessed using a NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, USA) and Experion RNA Stdsens analysis kit (Biorad Laboratories Inc., USA), respectively. The mRNA enrichment was performed using the mRNA enrichment kit (MICROBExpressTM, Ambion, Applied Biosystem, The Netherlands) according to the manufacturer's protocol. Following the enrichment, the quantity and quality of the RNA were assessed again (see above) to confirm the efficacy of the mRNA enrichment procedure. One µg of the enriched mRNA sample was used to reverse-transcribe the RNA to cDNA, and subsequently generate double stranded cDNA using the SuperScript® Double-Stranded cDNA Synthesis kit (Invitrogen, the Netherlands, 11917–010), and employing the SuperScript® III Reverse Transcriptase (Invitrogen, the Netherlands 18080–044) and random priming using random hexamers (Invitrogen, 48190–011) (Yoder-Himes *et al.* 2009, Leimena *et al.* 2012, Leimena *et al.* 2013). To remove RNA from the double stranded cDNA preparations, RNAse A (Roche, Germany) treatment was performed, followed by phenol-chloroform

extraction and subsequent cDNA purification and concentration by ethanol precipitation. The product was checked on 1% agarose gel and 3 to 8 µg of cDNA was sent for sequencing (GATC Biotech, Germany). Single read Illumina Libraries were prepared from the double-stranded cDNA according to the ChiP-seq protocol (Schmidt *et al.* 2009) with insert sizes between 200-300bp, suing barcoded tags for library constructions to enable parallel sequencing (GATC Biotech, Germany). Sequencing was performed using Illumina Hiseq2000 and using 5pM concentration of the library and the single-end protocol (Leimena *et al.* 2013).

Data filtering

Sequencing generated between 13.4 and 177 million reads per sample. The data set supporting the results of this article is available in the NCBI small reads archive (sra) repository, under accession number SRP043409. The data was filtered for ribosomal RNA sequences, adapter sequences and poor quality reads using the following tools. SortMeRNA (version 1.2) (Kopylova *et al.* 2012) was used to rapidly filter out rRNA sequences using the precompiled databases for eukaryotes, bacteria and archaea. Truseq adapter sequences were removed from the reads with cutadapt (Martin 2011). Initial results showed a high bias of adenines in the trimmed sequences and therefore all trimmed sequences were discarded. The remaining reads where quality (phred >30) and poly A tail edge trimmed using PRINSEQ (lite-version) (Schmieder and Edwards 2011). Reads smaller than 50 nucleotides were discarded.

Assembly, annotation and classification

Assembly was done using IDBA UD with default settings and ORF calling was performed using prodigal 2.60 with the meta procedure (Hyatt *et al.* 2010). Functional annotation was performed using KEGG automated annotation server using the SBH method against the default reference set (Quevillon *et al.* 2005, Moriya *et al.* 2007). Expression levels were determined by aligning the reads with the assembled contigs using MEGABLAST and counting the total nucleotide coverage of each ORF (Morgulis *et al.* 2008). Only alignments with a bitscore over a hundred were kept. Taxonomy of the predicted ORFs was determined by aligning the protein sequences with NR database using blastp (Altschul *et al.* 1997) and retrieving the lineage of the tophit.

Supplemental Information

Supplemental Figures

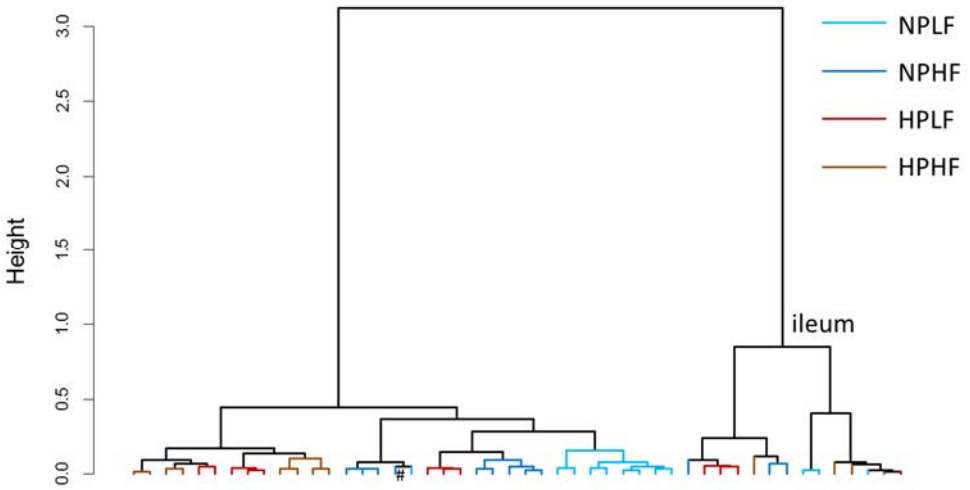


Figure S 7.1 Cluster of all the samples on log10 transformed probe level data of the MITChip. The clustering was done using Pearson distance.

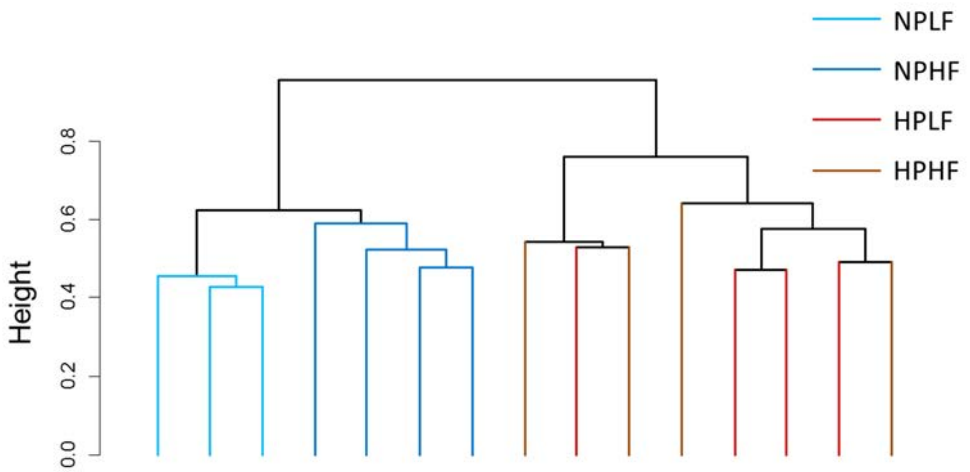


Figure S 7.2 Cluster on normalized (annotated) bacterial metatranscriptome data. Clustering via hclust script of R, using Pearson distance.

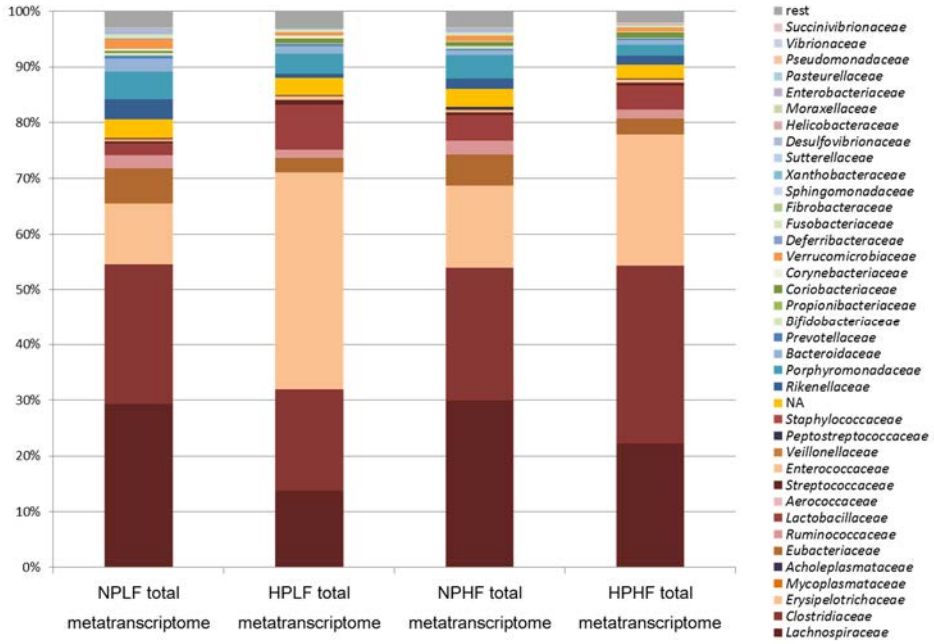


Figure S 7.3 Relative abundance of total metatranscriptome (activity) on family level.

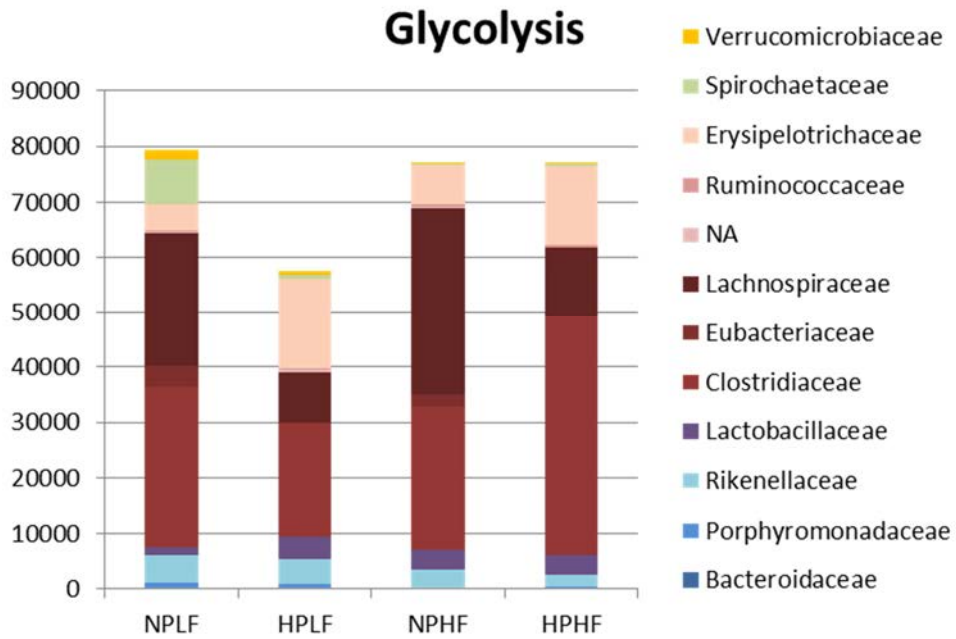


Figure S 7.4 Expression of 6-phosphofructokinase in the glycolysis pathway.

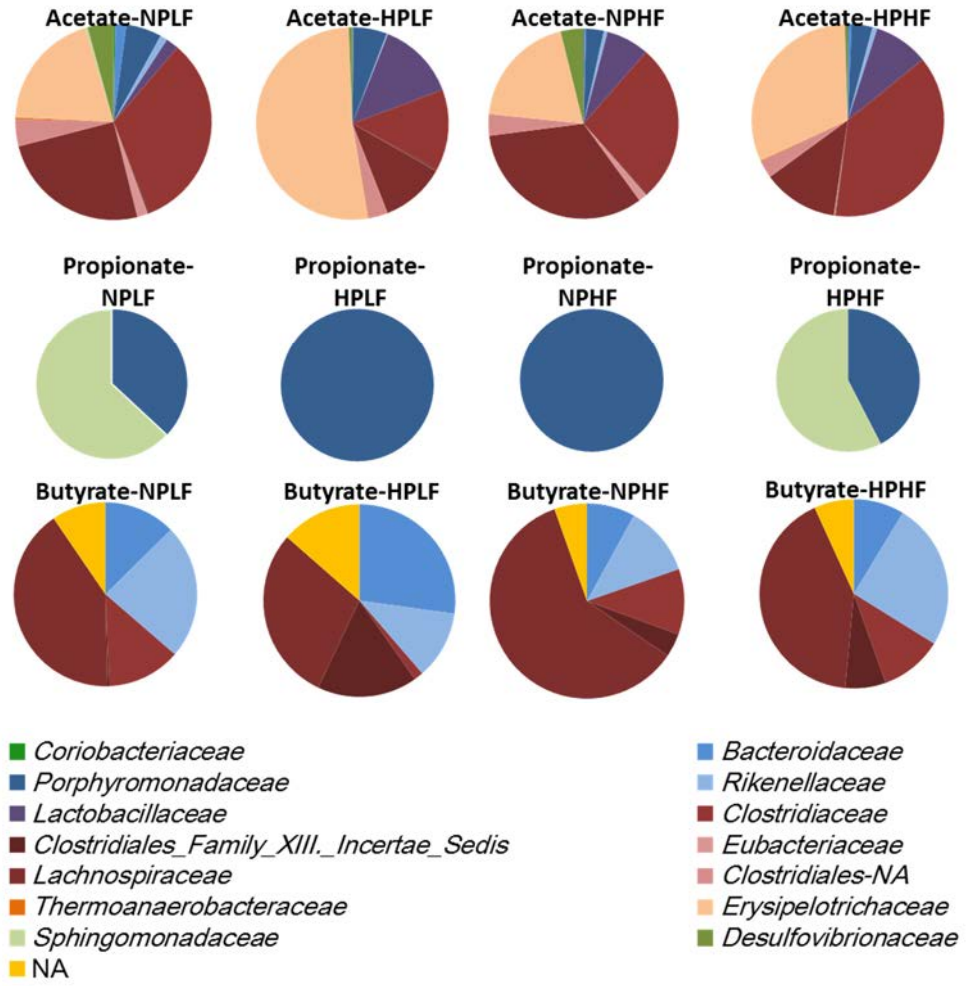


Figure S 7.5 Relative abundances of families that have activity towards SCFA production. Acetate production via acetate kinase, acetyl-CoA synthase and phosphate acetyltransferase (K00925, K00625, K01895, K13788; Propionate production via propionate CoA-transferase and acetyl-CoA synthase (K01895, K01026); Butyrate production via butyrate kinase, acetoacetate CoA-transferase and phosphate butyryltransferase (K01034, K01035, K01896, K00929, K00634).

Chapter 8

General Discussion

Mouse models

In the early 20th century Miss Abbie E.C. Lathrop started breeding mice in Granby, Massachusetts, to sell them as pets (Morse III 1978). However, within a few years her business had grown into an establishment selling mice to scientific laboratories in the area. She also started the inbreeding of mice to generate specific mouse lines, without cross-breeding with wild mice. These initial inbred mouse lines are the ancestors of many presently used laboratory mouse lines, of which the C57BL/6 is probably the most commonly used. Since approximately 1910 these mouse lines were inbred, with 1,5 to 2 generations per year, implying that the currently employed mice are inbred over 150 generations (Beck *et al.* 2000). Other frequently used laboratory mouse lines, like BALB/c, have a similar origin and were bought by scientists and mouse or small pet dealers since the beginning of the last century. The advantages of using inbred mouse lines in scientific research were discovered around that same time, and include their relatively small size and easy maintenance, as well as their applicability in animal models for the study of human diseases, which at that time focused on cancer and the evaluation of drugs to fight this disease (Morse III 1978), but have since then expanded to a huge variety of mouse models for a broad range of human diseases (Lyon and Searle 1989). Although the usefulness of mouse models for human diseases and the validity of extrapolating results obtained in mice to the human situation is a subject of constant debate, inbred mice are still widely used as model organisms for human (Ley *et al.* 2005, Turnbaugh *et al.* 2006, Martin *et al.* 2008b, Marco *et al.* 2009, Turnbaugh *et al.* 2009, Marco *et al.* 2010, Schwarz *et al.* 2012, Everard *et al.* 2013, Ridaura *et al.* 2013).

Murine gut microbiota

Studies related to the role of diet and environmental factors on mammalian health often use mice as a model, which also enables the evaluation of the influence of the host genotype on health and disease through the availability of genetic tools to modify the mouse genome (e.g., knock-out mouse lines etc.) (Cani *et al.* 2008, Zhang *et al.* 2009, Vijay-Kumar *et al.* 2010). Mouse models are also employed to study the microbial diversity in the GI tract and its responses to diet and pharmaceutical treatments, again using the mice as a model for the human situation. Although the phylogenetic makeup of the bacterial communities in human and mouse appear to be similar at phylum level, there are differences in the bacterial composition when evaluating the microbiota at a more refined phylogenetic level (Ley *et al.* 2006, Rawls *et al.* 2006). Humans and mice have 80% of their gut microbial genera in common, although their relative abundances differ significantly (Krych *et al.* 2013). However, significant differences are also seen at genus level, e.g., the murine GI tract harbours members of the phylum *Deferribacteres*, predominantly represented by the species *Muscispirillum schaedleri*, which appear to be absent in the human intestinal microbiota {Krych, 2013 #843}. Nevertheless, since the majority of the genera are present in both ecosystems, external factors like

diet and antibiotic treatments, could elicit similar responses in these ecosystems. This notion is supported by the finding presented in **Chapter 4** that higher abundances of *Ruminococcus bromii* and several species of *Clostridium* cluster IV were detected in the mice that were fed resistant starch, which is congruent with the changes elicited by this dietary component in human and pig intervention studies (Ze *et al.* 2012, Haenen *et al.* 2013b). Apparently, these bacterial groups are indeed responding similarly to an external factor, i.e., resistant starch, in different ecosystems. Conversely, interventions with fructooligosaccharides (FOS) and inulin usually result in an increase of bifidobacteria in the human faecal microbiota (Gibson and Wang 1994, Langlands *et al.* 2004, Ouwehand *et al.* 2005, Barboza *et al.* 2009, Riviere *et al.* 2014), whereas this modulation was not detected in the colon or faeces of mice, but was detected in the mouse caecum. Therefore, although the bifidobacterial members of the microbiota appear to respond similarly to FOS and inulin dietary interventions in both humans and mice, the localization of this effect is apparently different in these two host organisms. FOS and inulin are oligo- and polysaccharide compounds that are fermented in the proximal large intestine (Macfarlane *et al.* 1992), hence in the caecum or proximal colon, which implies that in humans the increase of bifidobacteria in faeces is sustained throughout the large intestine, whereas in mice the increase is diminishing along the length of the colon. Notably, the bifidobacteria in mice likely belong to different species compared to those in human, which could contribute to the different apparent persistence of these bacteria along the length of the colon.

Metatranscriptome-based analysis of murine gut microbiota activity profiles

Ley *et al*. showed that the human and murine microbiota is quite different at the species level. Owing to large scale projects like MetaHIT and the Human Microbiome Project we have now a basic understanding of the number of species present in the GI tract, their functional capacities and possible habitats (Turnbaugh *et al.* 2007, Qin *et al.* 2010, Consortium 2012, Rajilic-Stojanovic and de Vos 2014). 588 Bacterial strains that are considered as representatives for the human intestine microbiota have been cultured in vitro and their genome has been sequenced, and are currently further studied to understand their physiology (reviewed in (Rajilic-Stojanovic and de Vos 2014). However much fewer representative isolates have been cultured for the microbiota of mice, with only 115 mice-related microbial genomes compared to the 11748 human-associated bacteria with (partially) sequenced genomes in the GOLD database (September 2014), of which 588 are gastrointestinal microbiota (Rajilic-Stojanovic and de Vos 2014). Increasing the number of murine isolates and bacterial genome sequences will give a better picture on the make-up of the murine microbiota. Alternatively metagenomic sequencing of the murine GI tract microbiota could result in metagenomic species, following a recent approach developed within the MetaHIT consortium that resulted in 238 metagenomic species (Nielsen *et al.*

2014). With this information we could start comparing the human and mouse microbiota at a functional level, and start understanding particular similarities and differences of both microbiota's and their responses to specific interventions or treatments. Moreover, the lack of sequenced mouse microbiota species results in poor alignment of metatranscriptome datasets to the current genome databases, and thus limits taxonomic and functional assignments. In **Chapter 5** we described an assembly tool to handle murine metatranscriptome data, to overcome the absence of a database that contains genomes originating from the mouse intestine, or at least genomes of species of sufficient similarity to allow significant functional and taxonomic mapping. With this tool we were able to analyse the functional responses of the murine intestine microbiota to dietary interventions, such as in the studies described in **Chapter 6** and **7**. Metatranscriptome analysis was used to elucidate the role of different bacterial groups in the degradation of dietary fibres and proteins, and their participation in the production of SCFA. Butyrate, as one of the main SCFA, increasingly received attention during the last decade due to its proposed health benefits (Barcenilla *et al.* 2000, Bird *et al.* 2010, Louis *et al.* 2010, Scott *et al.* 2013, Vital *et al.* 2014). In **Chapter 6** and **7** we looked at the butyrate producing enzymes in the community, butyryl-CoA: acetyl-CoA transferase and butyrate kinase. These two enzymes share regions of high similarity to genes with other substrate specificities. For example, butyrate kinases are highly similar to acetate kinases (Vital, 2013). In Figure 9.1 the pathways towards acetate, lactate and butyrate from three different bacterial families discussed in **Chapter 6** are shown using iPATH. When studying gene expression patterns in response to different polysaccharides in the diet, *Lachnospiraceae* were characterized by the activity of the pathway that converts acetate towards butyrate. In addition, in the *Bacteroidaceae* transcription of a gene annotated as butyrate kinase was detected, whereas in these bacteria the transcripts associated with the acetyl-CoA and butyryl-CoA pathway were not found. However, in the *Erysipelotrichaceae* transcription of genes that encode proteins catalysing the conversion of acetyl-CoA to butyryl-CoA were detected, but not of the genes encoding the enzyme involved in the last step of the butyrate production pathway. Vital et al., 2014, already reported that the genome-based prediction of butyrate-production capacity in members of the *Bacteroidetes* phylum still requires experimental validation (Vital *et al.* 2014). In our metatranscriptome data of the *Bacteroidaceae* it seems that not the entire pathway required for the conversion of acetate to butyrate is expressed.

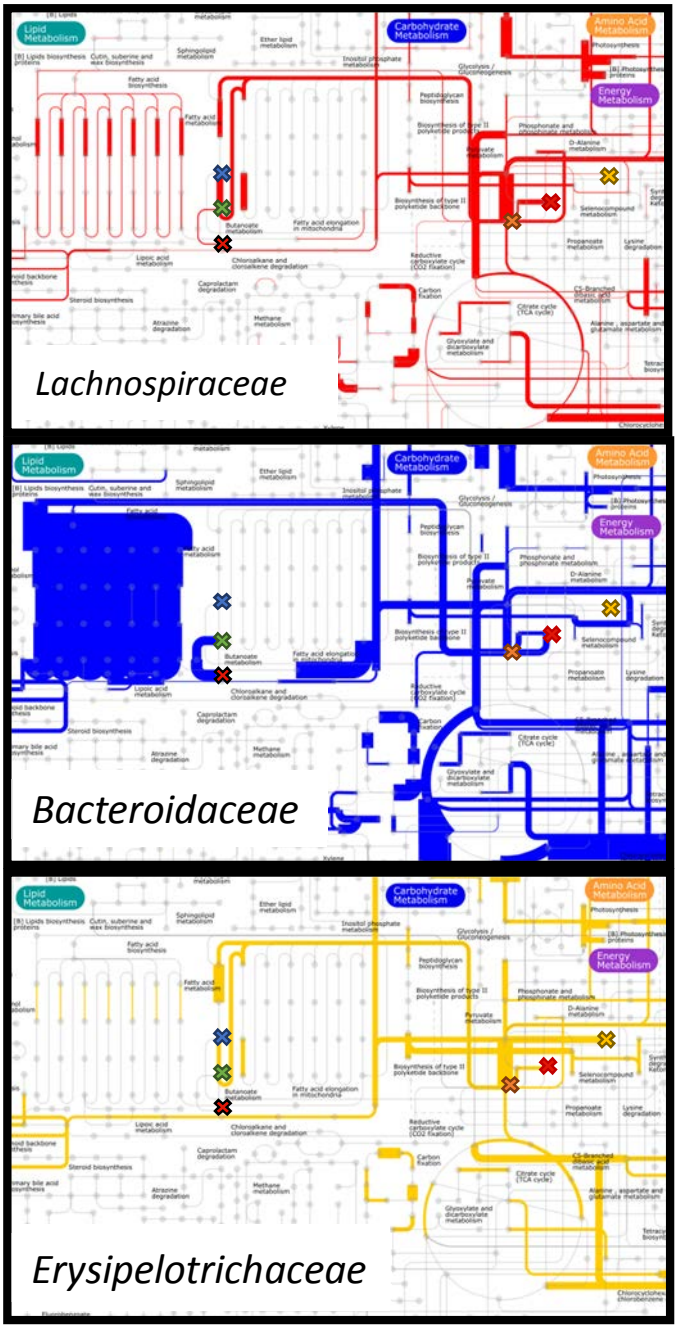


*Figure 9.1 Metatranscriptome profiles of Lachnospiraceae, Bacteroidaceae and Erysipelotrichaceae from **Chapter 6** using the visualization tool of iPATH.*

- *= Acetate*
- *= Acetyl-CoA*
- *= Crotonyl-CoA*
- *= Butyryl-CoA*
- *= Butyrate*
- *= Lactate*

This may suggest that members of the *Bacteroidaceae* are either producing butyrate via the lysine or glutarate pathway (Vital *et al.* 2014), although the gene encoding the enzyme that catalyses the conversion of crotonoyl-CoA to butyryl-CoA, and is required for the production of butyryl-CoA, is not detected in the metatranscriptome data. Another possibility would be that the butyrate kinase is wrongly annotated as a butyrate kinase in the genomes that we used as a template for similarity searches. The latter implies that the enzyme could be an acetate kinase rather than a butyrate kinase, suggesting that the members of the *Bacteroidaceae* described in Chapter 6 and 7 actually do not produce butyrate, but rather acetate, which would be in line with the fact that so far no observations of butyrate production by *Bacteroidaceae* were made (Louis *et al.* 2004, Louis and Flint 2009, Rajilic-Stojanovic and de Vos 2014, Vital *et al.* 2014). Consequently, the prediction of the participation of specific bacterial families in the production of SCFA using metatranscriptome analysis, at this stage requires careful manual curation, especially regarding the substrate and product specificities of key enzymes encoded by the expressed gene pool.

SCFA measurements

In the mouse experiments described in **Chapter 4**, **6** and **7** the luminal SCFA levels were measured as indicators for microbial fermentation. However, as indicated in **Chapter 6,** the snap-shot determinations of SCFA concentrations in the luminal content are not a valuable measurement for fibre fermentation output, since it is the result of production by the microbiota and absorption by the host, which prohibits the interpretation of these measurements in terms of microbial production rate or capacity. Rather, the shift in microbiota composition, gene expression and especially the multivariate integration of host mucosal gene expression and the microbiota composition provided a better proxy for the fermentation activity of the microbiota. Nevertheless, the increased luminal colonic SCFA concentrations elicited by the fibre diet that were determined in **Chapter 4** were reflected in microbiota composition changes, as well as gene expression changes and were especially apparent from the multivariate data integration analyses. In a different study that was executed within this project, the production and absorption fluxes of SCFA were calculated from experiments where different guar gum levels were provided in the diet, combined with infusions of ^{13}C labelled acetate, propionate and butyrate in the caecum (den Besten *et al.* 2014). These studies revealed a dose dependent increase of SCFA production and uptake fluxes, and illustrated that host metabolic markers, which included body weight and insulin and glucose levels in blood plasma after fasting, correlated with SCFA absorption fluxes. The same metabolic markers did not correlate with the measured SCFA concentration in the caecum, confirming our observation of **Chapter 6** that SCFA concentrations measured in the caecal content are not a valuable measurement for the estimation of the microbial fermentation output.

SCFA supplementation in the diet can prevent and reverse high-fat diet induced obesity (den Besten 2014). These studies employed sodium salts of acetate, propionate and butyrate as supplements in the diet, which also elicited slight alterations in microbial composition of the caecal microbiota (Figure 9.2). The dietary SCFA are already absorbed by the host mucosa in the small intestine and may have local effects on these mucosal tissues that could play a role in the observed physiological effects with respect to obesity. It should be noted, however, that the variation in the microbial composition explained by the SCFA in the diets (18%) was much smaller than the variation that could be explained by the fibre diets in the **Chapters 4** and **6** (43.7% and 58,6%, respectively). Dietary fibres will generate SCFA predominantly in the large intestine. This consideration may largely explain the very moderate differences in microbiota composition elicited by the dietary SCFA and could in fact be predominantly induced by the altered physiology of the host organism as apparent from their reduced bodyweight (den Besten 2014).

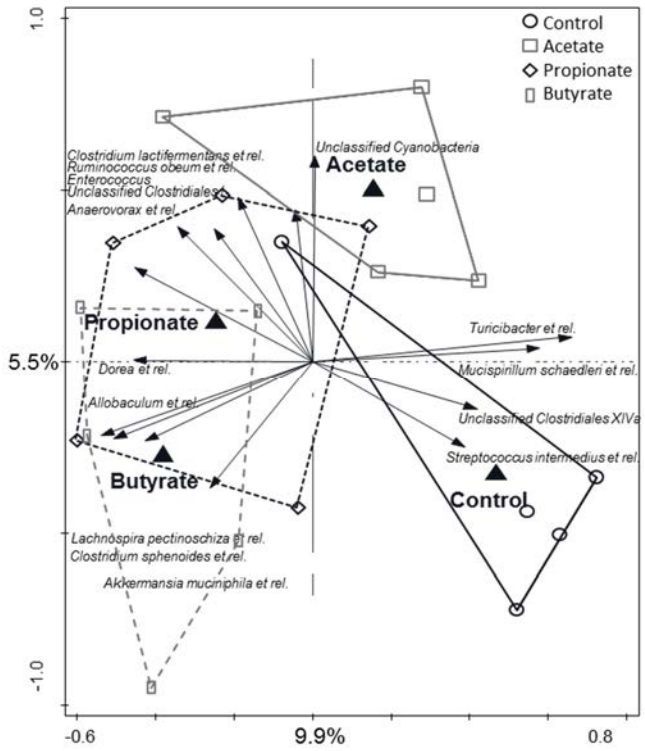


Figure 9.2 Redundancy analysis of murine microbial community, where the variables are control or the sodium supplements acetate, propionate or butyrate. The variables account for 18.1% of the total variation where in this plot 85.5% of the canonical variation is shown.

Host responses to SCFA

The dietary SCFA down-regulated a nuclear receptor, the peroxisome proliferator-activated receptor (PPAR)-γ, in the adipose tissue and liver (den Besten 2014). In these tissues, the repression of PPAR-γ induced a shift from lipogenesis to fatty-acid

oxidation, which probably explains how SCFA supplementation can prevent or reverse high-fat diet induced obesity. Intriguingly, in **Chapter 4** we proposed that fibre fermentation by the colon microbiota induces the activation of PPAR-γ in the colon epithelia, where different and shared subsets of PPAR-γ target genes were affected by the different fibre diets, which is most likely mediated through increased levels of SCFA exposure. Taken together PPAR-γ emerges as one of the key host response regulators that sense increased SCFA resulting from enhanced fermentation activity by the gut microbiota, leading to local modulation of the host's metabolism and potentially overall host physiology. To further explore the role of PPAR-γ in the host response to SCFA an intestine specific PPAR-γ knock-out (KO) mouse model was set up in the group of J.W. Jonker (University of Groningen, the Netherlands). The PPAR-γ KO mice were fed either the inulin or the control diet that were also used in the experiments described in **Chapter 4** and **6,** and the microbial composition changes elicited by this dietary intervention were compared to those seen in wild type mice (Figure 9.3).

The microbial composition of wild type and PPAR-γ KO mice appeared to be highly similar, whereas it was much more significantly affected by the diet intervention as compared to the host genetic background. These observations suggest that abolishing the host response to SCFA through intestinal PPAR-γ does not directly affect microbial community composition.

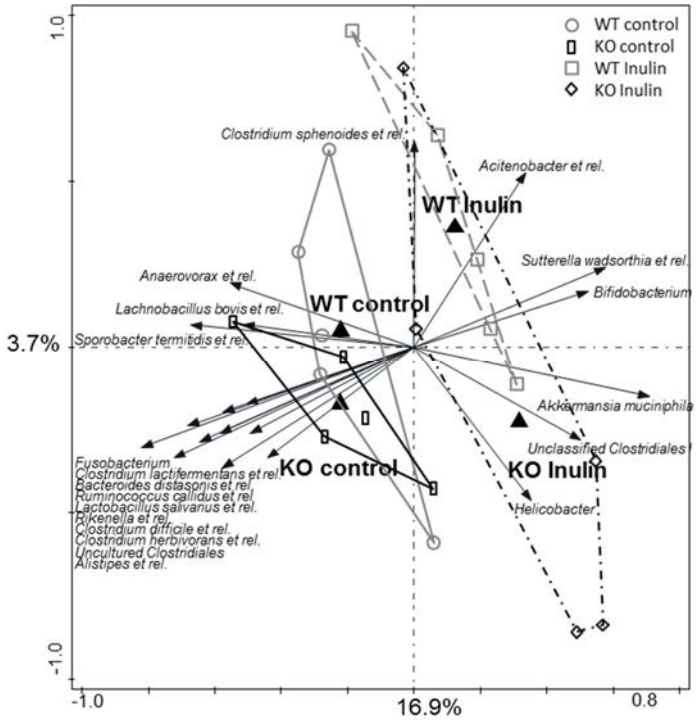


Figure 9.3 Redundancy analysis of colon microbiota from wildtype (WT) or PPAR-γ KO mice on a control or inulin diet. The variables were the different groups - WT on control diet (WT control); WT on inulin diet (WT Inulin); PPAR-γ KO on control diet and PPAR-γ KO on the inulin diet. These variables together accounted for 22.3% of the total variation, of which in this plot 92.2% is shown.

Overview and future perspectives

In this thesis protein and fibre dietary interventions in mice were studied to elucidate their impact on the fermentation activities of the gut microbiota, their SCFA production, and the mucosal responses to the SCFA. The fibres inulin, fructooligosaccharides, arabinoxylan and guar gum lead to increased SCFA concentrations and induced similar changes in relative abundance of microbial groups in the colon of mice. The relative abundance of bacteria belonging to *Clostridium* cluster XIVa correlated strong with expression of mucosal genes involved in energy metabolism of these mice. By using a *de novo* metatransciptome assembly pipeline microbial activities of protein and fibre fermentation in the caeca were studied. This new method showed that in fibre fermentation the *Bacteroidaceae, Verrucomicrobiaceae* and *Erysipelotrichaceae* increased, while *Lachnospiraceae* decreased in activity during inulin and guar gum dietary interventions. The *Erysipelotrichaceae* and *Lachnospiraceae* families displayed the same trend with high protein interventions, where also *Clostridiaceae* are important contributors to the protein fermentation, but don't show changed activity (Figure 9.4).

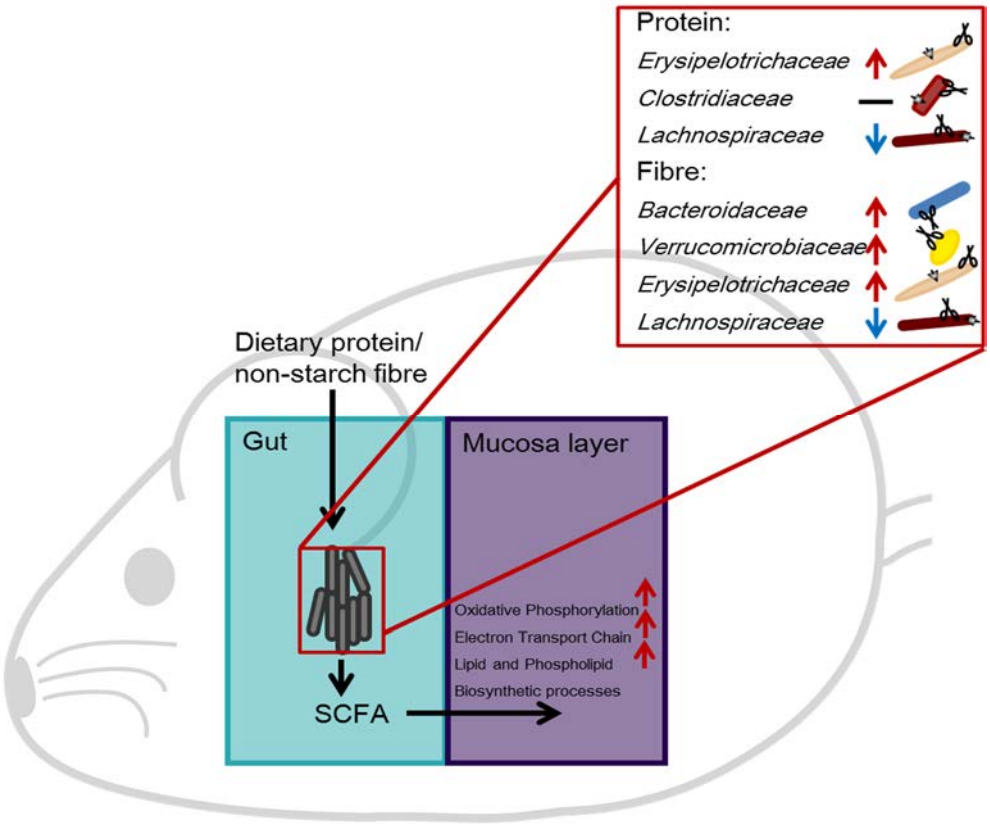


Figure 9.4. Overview of main results in this thesis.

The metatranscriptome enabled the detection of the major players in the fibre and protein fermentation. However we identified many gaps in the data, which make it challenging to identify the activity pattern of individual members within the community. Isolating and sequencing more members of the mouse gut microbiota will partially fill these gaps. These isolates also need to be further characterized in order to fully understand their metabolic capacities, and this information needs to be used to refine corresponding genome annotations. The annotation of many of the enzymes involved in SCFA conversions could be improved if the metabolic output of the bacterial species was characterized in more detail. Alternatively, as suggested by Vital *et al.*, 2014, for further annotations of genes involved in butyrate production, targeting the complete pathway is a more robust way to predict the function of individual genes (Vital *et al.* 2014). Genes that are predicted to be involved in this pathway can be identified using hidden Markov models (HMM) and, if the genome is accurately annotated, by the EC numbers. The identified genes could then be compared to the gene catalogue created by Vital *et al.*, which contains the pathways towards butyrate production (Vital *et al.* 2014). Creating gene catalogues for specialized pathways, e.g., butyrate production, could also support the annotation of

genes encoding enzymes of other SCFA production pathways, or the pathway intermediates like succinate, formate and oxaloacetate. Moreover, the protein catabolism capacity of gut bacteria has been poorly characterized to date. Our study reveals that only two members, the *Clostridiaceae* and *Erysipelotrichaceae* are involved in protein catabolism, but we were unable to accurately detect the transcripts that represent the complete pathway of amino acid degradation in these bacterial families, which would allow the confirmation of their use of dietary protein as energy source. The ability to use amino acids as carbon and energy source by *Clostridium* spp. has been known for some decades (Barker 1981, Macfarlane *et al.* 1986, Fonknechten *et al.* 2010). However in genome sequencing and subsequent analysis of GI tract microbiota, amino acids fermentation did not receive a lot of attention (Backhed *et al.* 2005, Ley *et al.* 2006, Qin *et al.* 2010, Arumugam *et al.* 2011, Rajilic-Stojanovic and de Vos 2014). The *Erysipelotrichaceae* were classified as a family in 2009 (Rajilic-Stojanovic and de Vos 2014), however, only 28 bacteria within this family have been sequenced to date and currently nine still need characterization and assignments within this family (NCBI genomes, September 2014). We showed that the *Erysipelotrichaceae* is an important member in both fibre and protein fermentation in the murine caecum, and further characterization of species within this family will help to understand its functional role within the microbial community. Moreover the *Erysipelotrichaceae* metabolic pattern as projected on iPATH patterns of Figure S 6.6 and Figure 7.5 appeared to be independent of the diet consumed by the mice, indicating that this family executes a similar functional role within the microbial community in all the different dietary interventions. The role of *Erysipelotrichaceae* within the human microbiota also deserves to be further investigated, especially if it appears to have similar functions and responses in human and could thus prove to be a good indicator for increased fermentation after a dietary intervention study.

The luminal SCFA concentration measurements in **Chapter 6** do not seem indicative of the fibre fermentation processes that we observed from the microbiota profiles and the host mucosal gene patterns. In contrast, the SCFA concentration measurements did appear to reflect the fermentation processes in the studies presented in **Chapter 4** and **7**. To obtain more robust indicators of increased fermentation, a combination of microbiota analysis and/or host responses should be done parallel to the SCFA measurements in future experiments. Indicator species in the large intestine microbiota could be found among members of *Clostridium* cluster XIVa and indicators in the mucosal transcriptome patterns, could include the genes regulated by PPAR-γ. However, such indicators still require validation in human trials, although several studies already showed increased *Clostridium* cluster XIVa related to fibre fermentation in human faeces (Hayashi *et al.* 2002, Hobden *et al.* 2013, Valdés *et al.* 2013). Increased *Clostridium* cluster XIVa and overall SCFA concentrations, especially butyrate are considered health biomarkers (Scheppach 1994, Louis and Flint 2009, Van den Abbeele *et al.* 2012a), implying that the consumption of inulin, fructooligosaccharides, arabinoxylan and guar gum can be

considered as beneficial for health.

Our project had the ambition to include systems biology approaches to better understand the interplay between microbiome and host. By using the mouse model we were able to perform site-specific measurements on fermentation, microbiota and mucosal gene expression in different regions of the intestine. This enabled the systematic evaluation of the interaction between microbes and the host, through combining multiple omics techniques using multivariate statistics. However to fully understand the microbe-diet and microbe-microbe interactions, and the consequences of these interactions for the SCFA rates in situ, further genomic and physiological characterization of intestinal isolates is crucial. This could open avenues towards the comprehensive description of the role of certain bacteria and their metabolic functions that are strongly correlated with host gene expression modulation. These descriptions could provide a foundation for the construction of (predictive) systems biology-based models for the interplay in the diet-microbiome-host interaction triangle.

References

Aas, J. A., B. J. Paster, L. N. Stokes, I. Olsen, and F. E. Dewhirst. 2005. Defining the normal bacterial flora of the oral cavity. Journal of clinical microbiology **43**:5721-5732.

Afman, L. and M. Muller. 2006. Nutrigenomics: from molecular nutrition to prevention of disease. J Am Diet Assoc **106**:569-576.

Aizawa, S. I. and T. Kubori. 1998. Bacterial flagellation and cell division. Genes to cells **3**:625-634.

Alex, S., K. Lange, T. Amolo, J. S. Grinstead, A. K. Haakonsson, E. Szalowska, A. Koppen, K. Mudde, D. Haenen, S. Al-Lahham*, et al.* 2013. Short-chain fatty acids stimulate angiopoietin-like 4 synthesis in human colon adenocarcinoma cells by activating peroxisome proliferator-activated receptor gamma. Mol Cell Biol **33**:1303-1316.

Al-Lahham, S. A., Roelofsen, H., Rezaee, F., Weening, D., Hoek, A., Vonk, R., & Venema, K. (2012). Propionic acid affects immune status and metabolism in adipose tissue from overweight subjects. European journal of clinical investigation, 42(4), 357-364.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**:3389-3402.

Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J. M. Batto*, et al.* 2011. Enterotypes of the human gut microbiome. Nature **473**:174-180.

Attebery, H. R., V. L. Sutter, and S. M. Finegold. 1972. Effect of a partially chemically defined diet on normal human fecal flora. Am J Clin Nutr **25**:1391-1398.

Aune, D., D. S. Chan, R. Lau, R. Vieira, D. C. Greenwood, E. Kampman, and T. Norat. 2011. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ **343**:d6617.

Azcarate-Peril, M. A., D. M. Foster, M. B. Cadenas, M. R. Stone, S. K. Jacobi, S. H. Stauffer, A. Pease, and J. L. Gookin. 2011. Acute necrotizing enterocolitis of preterm piglets is characterized by dysbiosis of ileal mucosa-associated bacteria. Gut microbes **2**:234-243.

Backhed, F., H. Ding, T. Wang, L. V. Hooper, G. Y. Koh, A. Nagy, C. F. Semenkovich, and J. I. Gordon. 2004. The gut microbiota as an environmental factor that regulates fat storage. Proceedings of the National Academy of Sciences of the United States of America **101**:15718-15723.

Backhed, F., R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon. 2005. Host-bacterial mutualism in the human intestine. Science **307**:1915-1920.

Balakrishnan, M. and M. H. Floch. 2012. Prebiotics, probiotics and digestive health. Curr Opin Clin Nutr Metab Care **15**:580-585.

Baldrian, P., M. Kolařík, M. Štursová, J. Kopecký, V. Valášková, T. Větrovský, L. Žifčáková, J. Šnajdr, J. Rídl, and Č. Vlček. 2011. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. The ISME journal **6**:248-258.

Barboza, M., D. A. Sela, C. Pirim, R. G. Locascio, S. L. Freeman, J. B. German, D. A. Mills, and C. B. Lebrilla. 2009. Glycoprofiling bifidobacterial consumption of galacto-oligosaccharides by mass spectrometry reveals strain-specific, preferential consumption of glycans. Applied and environmental microbiology **75**:7319-7325.

Barcenilla, A., S. E. Pryde, J. C. Martin, S. H. Duncan, C. S. Stewart, C. Henderson, and H. J. Flint. 2000. Phylogenetic relationships of butyrate-producing bacteria from the human gut. Applied and environmental microbiology **66**:1654-1661.

Barker, H. 1981. Amino acid degradation by anaerobic bacteria. Annu Rev Biochem **50**:23-40.

Beck, J. A., S. Lloyd, M. Hafezparast, M. Lennon-Pierce, J. T. Eppig, M. F. Festing, and E. M. Fisher. 2000. Genealogies of mouse inbred strains. Nature genetics **24**:23-25.

Belenguer, A., S. H. Duncan, A. G. Calder, G. Holtrop, P. Louis, G. E. Lobley, and H. J. Flint. 2006. Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Applied and environmental microbiology **72**:3593-3599.

Belobrajdic, D. P., G. H. McIntosh, and J. A. Owens. 2004. A high-whey-protein diet reduces body weight gain and alters insulin sensitivity relative to red meat in wistar rats. The Journal of nutrition **134**:1454-1458.

Bik, E. M., P. B. Eckburg, S. R. Gill, K. E. Nelson, E. A. Purdom, F. Francois, G. Perez-Perez, M. J. Blaser, and D. A. Relman. 2006. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A **103**:732-737.

Bird, A. R., M. A. Conlon, C. T. Christophersen, and D. L. Topping. 2010. Resistant starch, large bowel fermentation and a broader perspective of prebiotics and probiotics. Beneficial microbes **1**:423-431.

Bloemen, J. G., K. Venema, M. C. van de Poll, S. W. Olde Damink, W. A. Buurman, and C. H. Dejong. 2009. Short chain fatty acids exchange across the gut and liver in humans measured at surgery. Clin Nutr **28**:657-661.

Booijink, C. C., S. El-Aidy, M. Rajilic-Stojanovic, H. G. Heilig, F. J. Troost, H. Smidt, M. Kleerebezem, W. M. De Vos, and E. G. Zoetendal. 2010. High temporal and inter-individual variation detected in the human ileal microbiota. Environmental Microbiology **12**:3213-3227.

Booijink, C. C., E. G. Zoetendal, M. Kleerebezem, and W. M. de Vos. 2007. Microbial communities in the human small intestine: coupling diversity to metagenomics. Future microbiology **2**:285-295.

Boulesteix, A. L. and K. Strimmer. 2007. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief Bioinform **8**:32-44.

Bounous, G. and G. J. Devroede. 1974. Effects of an elemental diet on human fecal flora. Gastroenterology **66**:210-214.

Brooks, J. D., V. G. Paton, and G. Vidanes. 2001. Potent induction of phase 2 enzymes in human prostate cells by sulforaphane. Cancer Epidemiology Biomarkers & Prevention **10**:949-954.

Burrin, D. G., B. Stoll, R. Jiang, X. Chang, B. Hartmann, J. J. Holst, G. H. Greeley, and P. J. Reeds. 2000. Minimal enteral nutrient requirements for intestinal growth in neonatal piglets: how much is enough? The American journal of clinical

nutrition **71**:1603-1610.
Cani, P. D., R. Bibiloni, C. Knauf, A. Waget, A. M. Neyrinck, N. M. Delzenne, and R. Burcelin. 2008. Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet–induced obesity and diabetes in mice. Diabetes **57**:1470-1481.
Cani, P. D. and A. Everard. 2014. [Akkermansia muciniphila: a novel target controlling obesity, type 2 diabetes and inflammation?]. Med Sci (Paris) **30**:125-127.
Chacko, A. and J. H. Cummings. 1988. Nitrogen losses from the human small bowel: obligatory losses and the effect of physical form of food. Gut **29**:809-815.
Chang, D., Y. Zhu, L. An, J. Liu, L. Su, Y. Guo, Z. Chen, Y. Wang, L. Wang, and J. Wang. 2013. A multi-omic analysis of an Enterococcus faecium mutant reveals specific genetic mutations and dramatic changes in mRNA and protein expression. BMC microbiology **13**:304.
Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan. 2013a. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics **14**:128.
Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan. 2013b. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics **14**:128.
Cilieborg, M. S., M. Boye, L. Molbak, T. Thymann, and P. T. Sangild. 2011. Preterm Birth and Necrotizing Enterocolitis Alter Gut Colonization in Pigs. Pediatr Res **69**:10-16.
Claesson, M. J., I. B. Jeffery, S. Conde, S. E. Power, E. M. O'Connor, S. Cusack, H. M. Harris, M. Coakley, B. Lakshminarayanan, O. O'Sullivan*, et al.* 2012. Gut microbiota composition correlates with diet and health in the elderly. Nature **488**:178-184.
Clark, S., R. Egan, P. I. Frazier, and Z. Wang. 2013. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics:bts723.
Consortium, H. M. P. 2012. Structure, function and diversity of the healthy human microbiome. Nature **486**:207-214.
Cook, S. I. and J. H. Sellin. 1998. Review article: short chain fatty acids in health and disease. Alimentary pharmacology & therapeutics **12**:499-507.
Cummings, J. H. 1981. Short chain fatty acids in the human colon. Gut **22**:763-779.
Cummings, J. H. and G. T. Macfarlane. 1991. The control and consequences of bacterial fermentation in the human colon. The Journal of applied bacteriology **70**:443-459.
Cummings, J. H. and G. T. Macfarlane. 1997. Colonic microflora: nutrition and health. Nutrition **13**:476-478.
Cummings, J. H., G. T. Macfarlane, and H. N. Englyst. 2001. Prebiotic digestion and fermentation. The American journal of clinical nutrition **73**:415S-420S.
Cummings, J. H., E. W. Pomare, W. J. Branch, C. P. Naylor, and G. T. Macfarlane. 1987. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut **28**:1221-1227.
Cummings, J. H. and A. M. Stephen. 2007. Carbohydrate terminology and classification. Eur J Clin Nutr **61 Suppl 1**:S5-18.

Dai, M., P. Wang, A. D. Boyd, G. Kostov, B. Athey, E. G. Jones, W. E. Bunney, R. M. Myers, T. P. Speed, H. Akil*, et al.* 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res **33**:e175.

Darragh, A. J. and P. J. Moughan. 1995. The Three-Week-Old Piglet as a Model Animal for Studying Protein Digestion in Human Infants. Journal of pediatric gastroenterology and nutrition **21**:387-393.

De Preter, V. and K. Verbeke. 2013. Metabolomics as a diagnostic tool in gastroenterology. World J Gastrointest Pharmacol Ther **4**:97-107.

Decuypere, J., I. Vervaeke, H. Henderickx, and N. Dierick. 1977. Gastrointestinal cannulation in pigs: a simple technique allowing multiple replacements. Journal of animal science **45**.

den Besten, G. 2014. Elucidating the mechanisms of action of short-chain fatty acids, from dietary fiber to host metabolism.

den Besten, G., R. Havinga, A. Bleeker, S. Rao, A. Gerding, K. van Eunen, A. K. Groen, D.-J. Reijngoud, and B. M. Bakker. 2014. The Short-Chain Fatty Acid Uptake Fluxes by Mice on a Guar Gum Supplemented Diet Associate with Amelioration of Major Biomarkers of the Metabolic Syndrome. PloS one **9**:e107392.

den Besten, G., K. van Eunen, A. K. Groen, K. Venema, D. J. Reijngoud, and B. M. Bakker. 2013. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J Lipid Res **54**:2325-2340.

Deng, W., D. Xi, H. Mao, and M. Wanapat. 2008. The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. Mol Biol Rep **35**:265-274.

Derrien, M., M. C. Collado, K. Ben-Amor, S. Salminen, and W. M. de Vos. 2008. The Mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. Applied and environmental microbiology **74**:1646-1648.

Diamant, M., E. Blaak, and W. De Vos. 2011. Do nutrient–gut–microbiota interactions play a role in human obesity, insulin resistance and type 2 diabetes? Obesity reviews **12**:272-281.

Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K., & Bultman, S. J. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell metabolism, **13**(5), 517-526.

Douglas, W. R. 1972. Of pigs and men and research. Space life sciences **3**:226-234.

Dubuquoy, L., C. Rousseaux, X. Thuru, L. Peyrin-Biroulet, O. Romano, P. Chavatte, M. Chamaillard, and P. Desreumaux. 2006. PPARgamma as a new therapeutic target in inflammatory bowel diseases. Gut **55**:1341-1349.

Duncan, S. H., P. Louis, and H. J. Flint. 2004. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Applied and environmental microbiology **70**:5810-5817.

Dupont, C. L., D. B. Rusch, S. Yooseph, M.-J. Lombardo, R. A. Richter, R. Valas, M. Novotny, J. Yee-Greenbaum, J. D. Selengut, and D. H. Haft. 2011. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. The ISME journal **6**:1186-1199.

El Aidy, S., M. Derrien, C. A. Merrifield, F. Levenez, J. Dore, M. V. Boekschoten, J. Dekker, E. Holmes, E. G. Zoetendal, P. van Baarlen*, et al.* 2013a. Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon.

The ISME journal **7**:743-755.

El Aidy, S., C. A. Merrifield, M. Derrien, P. van Baarlen, G. Hooiveld, F. Levenez, J. Dore, J. Dekker, E. Holmes, S. P. Claus*, et al.* 2013b. The gut microbiota elicits a profound metabolic reorientation in the mouse jejunal mucosa during conventionalisation. Gut **62**:1306-1314.

Eswaran, S., J. Muir, and W. D. Chey. 2013. Fiber and functional gastrointestinal disorders. Am J Gastroenterol **108**:718-727.

Everard, A., C. Belzer, L. Geurts, J. P. Ouwerkerk, C. Druart, L. B. Bindels, Y. Guiot, M. Derrien, G. G. Muccioli, N. M. Delzenne*, et al.* 2013. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci U S A **110**:9066-9071.

Faith, J. J., J. L. Guruge, M. Charbonneau, S. Subramanian, H. Seedorf, A. L. Goodman, J. C. Clemente, R. Knight, A. C. Heath, R. L. Leibel*, et al.* 2013. The long-term stability of the human gut microbiota. Science **341**:1237439.

Falony, G., A. Vlachou, K. Verbrugghe, and L. De Vuyst. 2006. Cross-feeding between Bifidobacterium longum BB536 and acetate-converting, butyrate-producing colon bacteria during growth on oligofructose. Applied and environmental microbiology **72**:7835-7841.

Fava, F., J. A. Lovegrove, K. M. Tuohy, and G. R. Gibson. 2008. The potential role of the intestinal gut microbiota in obesity and the metabolic syndrome. Food Science and Technology Bulletin: Functional Foods **5**:71-92.

Ferguson, L. R., C. Tasman-Jones, H. Englyst, and P. J. Harris. 2000. Comparative effects of three resistant starch preparations on transit time and short-chain fatty acid production in rats. Nutr Cancer **36**:230-237.

Finegold, S. M., V. L. Sutter, and G. E. Mathisen. 1983. Normal indiggenous intestinal flora. Pages 3-31 Human intestinal microflora in health and disease.

Flint, H. J. and E. A. Bayer. 2008. Plant cell wall breakdown by anaerobic microorganisms from the Mammalian digestive tract. Annals of the New York Academy of Sciences **1125**:280-288.

Flint, H. J., E. A. Bayer, M. T. Rincon, R. Lamed, and B. A. White. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol **6**:121-131.

Flint, H. J., K. P. Scott, S. H. Duncan, P. Louis, and E. Forano. 2012a. Microbial degradation of complex carbohydrates in the gut. Gut Microbes **3**:289-306.

Flint, H. J., K. P. Scott, P. Louis, and S. H. Duncan. 2012b. The role of the gut microbiota in nutrition and health. Nat Rev Gastroenterol Hepatol **9**:577-589.

Fonknechten, N., S. Chaussonnerie, S. Tricot, A. Lajus, J. R. Andreesen, N. Perchat, E. Pelletier, M. Gouyvenoux, V. Barbe, M. Salanoubat*, et al.* 2010. Clostridium sticklandii, a specialist in amino acid degradation:revisiting its metabolism through its genome sequence. BMC genomics **11**:555.

Fraher, M. H., P. W. O'Toole, and E. M. Quigley. 2012. Techniques used to characterize the gut microbiota: a guide for the clinician. Nat Rev Gastroenterol Hepatol **9**:312-322.

Franzosa, E. A., X. C. Morgan, N. Segata, L. Waldron, J. Reyes, A. M. Earl, G. Giannoukos, M. R. Boylan, D. Ciulla, D. Gevers*, et al.* 2014. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A **111**:E2329-2338.

Frese, S. A., A. K. Benson, G. W. Tannock, D. M. Loach, J. Kim, M. Zhang, P. L. Oh, N. C. K. Heng, P. B. Patil, N. Juge*, et al.* 2011. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. PLoS Genet **7**:e1001314.

Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong. 2008. Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences **105**:3805-3810.

Fritz, J. V., M. S. Desai, P. Shah, J. G. Schneider, and P. Wilmes. 2013. From meta-omics to causality: experimental models for human microbiome research. Microbiome **1**:14.

Gaboriau-Routhiau, V., S. Rakotobe, E. Lécuyer, I. Mulder, A. Lan, C. Bridonneau, V. Rochet, A. Pisi, M. De Paepe, G. Brandi*, et al.* 2009. The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. Immunity **31**:677-689.

Gargallo, J. and D. Zimmerman. 1980. A simple intestinal cannula for swine. American journal of veterinary research **41**:618.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry*, et al.* 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol **5**:R80.

Geurts, L., V. Lazarevic, M. Derrien, A. Everard, M. Van Roye, C. Knauf, P. Valet, M. Girard, G. G. Muccioli, P. Francois*, et al.* 2011a. Altered gut microbiota and endocannabinoid system tone in obese and diabetic leptin-resistant mice: impact on apelin regulation in adipose tissue. Frontiers in microbiology **2**:149.

Geurts, L., V. Lazarevic, M. Derrien, A. Everard, M. Van Roye, C. Knauf, P. Valet, M. Girard, G. G. Muccioli, P. Francois*, et al.* 2011b. Altered gut microbiota and endocannabinoid system tone in obese and diabetic leptin-resistant mice: impact on apelin regulation in adipose tissue. Front Microbiol **2**:149.

Gibson, G. R. and M. B. Roberfroid. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. Journal of Nutrition **125**:1401-1412.

Gibson, G. R. and X. Wang. 1994. Enrichment of bifidobacteria from human gut contents by oligofructose using continuous culture. FEMS microbiology letters **118**:121-127.

Gibson, S. A., C. McFarlan, S. Hay, and G. T. MacFarlane. 1989. Significance of microflora in proteolysis in the colon. Applied and environmental microbiology **55**:679-683.

Gilbert, J. A., D. Field, Y. Huang, R. Edwards, W. Li, P. Gilna, and I. Joint. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PloS one **3**:e3042.

Gonzalez, I., K. A. Cao, M. J. Davis, and S. Dejean. 2012. Visualising associations between paired 'omics' data sets. BioData Min **5**:19.

Gosalbes, M. J., A. Durban, M. Pignatelli, J. J. Abellan, N. Jimenez-Hernandez, A. E. Perez-Cobas, A. Latorre, and A. Moya. 2011. Metatranscriptomic approach to analyze the functional human gut microbiota. PloS one **6**:e17447.

Greenblum, S., P. J. Turnbaugh, and E. Borenstein. 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci U S A **109**:594-599.

Grootaert, C., P. Van den Abbeele, M. Marzorati, W. F. Broekaert, C. M. Courtin, J.

A. Delcour, W. Verstraete, and T. Van de Wiele. 2009. Comparison of prebiotic effects of arabinoxylan oligosaccharides and inulin in a simulator of the human intestinal microbial ecosystem. FEMS Microbiol Ecol **69**:231-242.

Gross, G., J. Van Der Meulen, J. Snel, R. Van Der Meer, M. Kleerebezem, T. A. Niewold, M. M. Hulst, and M. A. Smits. 2008. Mannose-specific interaction of Lactobacillus plantarum with porcine jejunal epithelium. FEMS Immunology & Medical Microbiology **54**:215-223.

Guilloteau, P., R. Zabielski, H. M. Hammon, and C. C. Metges. 2010. Nutritional programming of gastrointestinal tract development. Is the pig a good model for man? Nutrition research reviews **23**:4-22.

Haange, S. B., A. Oberbach, N. Schlichting, F. Hugenholtz, H. Smidt, M. von Bergen, H. Till, and J. Seifert. 2012. Metaproteome analysis and molecular genetics of rat intestinal microbiota reveals section and localization resolved species distribution and enzymatic functionalities. Journal of proteome research **11**:5406-5417.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, and M. Lieber. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols **8**:1494-1512.

Haenen, D., C. Souza da Silva, J. Zhang, S. J. Koopmans, G. Bosch, J. Vervoort, W. J. Gerrits, B. Kemp, H. Smidt, M. Muller*, et al.* 2013a. Resistant starch induces catabolic but suppresses immune and cell division pathways and changes the microbiome in the proximal colon of male pigs. J Nutr **143**:1889-1898.

Haenen, D., J. Zhang, C. Souza da Silva, G. Bosch, I. M. van der Meer, J. van Arkel, J. J. van den Borne, O. Perez Gutierrez, H. Smidt, B. Kemp*, et al.* 2013b. A diet high in resistant starch modulates microbiota composition, SCFA concentrations, and gene expression in pig intestine. The Journal of nutrition **143**:274-283.

Haenen, D., J. Zhang, C. Souza da Silva, G. Bosch, I. M. van der Meer, J. van Arkel, J. J. van den Borne, O. Perez Gutierrez, H. Smidt, B. Kemp*, et al.* 2013c. A diet high in resistant starch modulates microbiota composition, SCFA concentrations, and gene expression in pig intestine. J Nutr **143**:274-283.

Haenen, D., J. Zhang, C. Souza da Silva, G. Bosch, I. M. van der Meer, J. van Arkel, J. J. G. C. van den Borne, O. Pérez Gutiérrez, H. Smidt, B. Kemp*, et al.* 2013d. A Diet High in Resistant Starch Modulates Microbiota Composition, SCFA Concentrations, and Gene Expression in Pig Intestine. The Journal of Nutrition **143**:274-283.

Hayashi, H., M. Sakamoto, and Y. Benno. 2002. Fecal microbial diversity in a strict vegetarian as determined by molecular analysis and cultivation. Microbiology and immunology **46**:819-831.

Heber, S. and B. Sick. 2006. Quality assessment of Affymetrix GeneChip data. OMICS **10**:358-368.

Hedemann, M. S., P. K. Theil, and K. Bach Knudsen. 2009. The thickness of the intestinal mucous layer in the colon of rats fed various sources of non-digestible carbohydrates is positively correlated with the pool of SCFA but negatively correlated with the proportion of butyric acid in digesta. British journal of nutrition **102**:117-125.

Hildebrand, F., T. L. Nguyen, B. Brinkman, R. G. Yunta, B. Cauwe, P. Vandenabeele, A. Liston, and J. Raes. 2013. Inflammation-associated enterotypes, host

genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. Genome Biol **14**:R4.

Hobden, M., A. Martin-Morales, L. Guérin-Deremaux, E. Deaville, A. Costabile, G. Walton, I. Rowland, O. Kennedy, and G. Gibson. 2013. In vitro fermentation of NUTRIOSE® soluble fibre in a continuous culture human colonic model system. Proceedings of the Nutrition Society **72**:E186.

Hooper, L. V., T. Midtvedt, and J. I. Gordon. 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu Rev Nutr **22**:283-307.

Hoskins, L. C. and E. T. Boulding. 1976. Degradation of Blood Group Antigens in Human Colon Ecosystems I. *In vitro* production of ABO blood group-degrading enzymes by enteric bacteria. The Journal of Clinical Investigation **57**:63-73.

Hou, Y., F. Moreau, and K. Chadee. 2012. PPARgamma is an E3 ligase that induces the degradation of NFkappaB/p65. Nat Commun **3**:1300.

Hudson, M., S. Borriello, and M. Hill. 1981. Elemental diets and the bacterial flora of the gastrointestinal tract. CRC Press, Boca Raton, FL.

Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. Genome research **17**:377-386.

Huson, D. H., S. Mitra, H.-J. Ruscheweyh, N. Weber, and S. C. Schuster. 2011. Integrative analysis of environmental sequences using MEGAN4. Genome research **21**:1552-1560.

Hyatt, D., G.-L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics **11**:119.

IJssennagger, N., M. Derrien, G. M. van Doorn, A. Rijnierse, B. van den Bogert, M. Muller, J. Dekker, M. Kleerebezem, and R. van der Meer. 2012. Dietary heme alters microbiota and mucosa of mouse colon without functional changes in host-microbe cross-talk. PloS one **7**:e49868.

Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4**:249-264.

Ivanov, I. I., K. Atarashi, N. Manel, E. L. Brodie, T. Shima, U. Karaoz, D. Wei, K. C. Goldfarb, C. A. Santee, S. V. Lynch*, et al.* 2009. Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell **139**:485-498.

Jansman, A., J. Zhang, S. Koopmans, R. Dekker, and H. Smidt. 2012. Effects of a simple or a complex starter microbiota on intestinal microbiota composition in caesarean derived piglets. Journal of animal science **90**:433-435.

Jeffery, I. B., M. J. Claesson, P. W. O'Toole, and F. Shanahan. 2012. Categorization of the gut microbiota: enterotypes or gradients? Nat Rev Microbiol **10**:591-592.

Johansen, T., H. S. Hansen, B. Richelsen, and K. Malmlof. 2001. The obese Gottingen minipig as a model of the metabolic syndrome: dietary effects on obesity, insulin sensitivity, and growth hormone profile. Comparative medicine **51**:150-155.

Keenan, M. J., R. J. Martin, A. M. Raggio, K. L. McCutcheon, I. L. Brown, A. Birkett, S. S. Newman, J. Skaf, M. Hegsted, R. T. Tulley*, et al.* 2012. High-amylose resistant starch increases hormones and improves structure and function of the gastrointestinal tract: a microarray study. J Nutrigenet Nutrigenomics **5**:26-44.

Kenagy, G. J. and D. F. Hoyt. 1980. Reingestion of Feces in Rodents and Its Daily Rhythmicity. Oecologia **44**:403-409.

Kerckhoffs, A. P., M. Samson, G. P. van Berge Henegouwen, L. M. Akkermans, V. B. Nieuwenhuijs, and M. R. Visser. 2006. Sampling microbiota in the human gastrointestinal tract. Gastrointestinal microbiology:25-50.

Kleerebezem, M. and E. E. Vaughan. 2009. Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity. Annual review of microbiology **63**:269-290.

Koirala, S., P. Mears, M. Sim, I. Golding, Y. R. Chemla, P. D. Aldridge, and C. V. Rao. 2014. A Nutrient-Tunable Bistable Switch Controls Motility in Salmonella enterica Serovar Typhimurium. MBio **5**:e01611-01614.

Kolenbrander, P. E. 2000. Oral microbial communities: biofilms, interactions, and genetic systems. Annual review of microbiology **54**:413-437.

Kopylova, E., L. Noé, and H. Touzet. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics **28**:3211-3217.

Kovatcheva-Datchary, P., M. Egert, A. Maathuis, M. Rajilic-Stojanovic, A. A. de Graaf, H. Smidt, W. M. de Vos, and K. Venema. 2009. Linking phylogenetic identities of bacteria to starch fermentation in an in vitro model of the large intestine by RNA-based stable isotope probing. Environmental Microbiology **11**:914-926.

Kroon, P. A., M. N. Clifford, A. Crozier, A. J. Day, J. L. Donovan, C. Manach, and G. Williamson. 2004. How should we assess the effects of exposure to dietary polyphenols in vitro? Am J Clin Nutr **80**:15-21.

Krych, L., C. H. Hansen, A. K. Hansen, F. W. van den Berg, and D. S. Nielsen. 2013. Quantitatively different, yet qualitatively alike: a meta-analysis of the mouse core gut microbiome with a view towards the human gut microbiome. PloS one **8**:e62578.

Kverka, M., Z. Zakostelska, K. Klimesova, D. Sokol, T. Hudcovic, T. Hrncir, P. Rossmann, J. Mrazek, J. Kopecny, E. F. Verdu*, et al.* 2011. Oral administration of Parabacteroides distasonis antigens attenuates experimental murine colitis through modulation of immunity and microbiota composition. Clin Exp Immunol **163**:250-259.

Lacroix, M., C. Gaudichon, A. Martin, C. Morens, V. Mathe, D. Tome, and J. F. Huneau. 2004. A long-term high-protein diet markedly reduces adipose tissue without major side effects in Wistar male rats. Am J Physiol Regul Integr Comp Physiol **287**:R934-942.

Lahti, L., L. L. Elo, T. Aittokallio, and S. Kaski. 2011. Probabilistic analysis of probe reliability in differential gene expression studies with short oligonucleotide arrays. IEEE/ACM Trans Comput Biol Bioinform **8**:217-225.

Lahti, L., A. Torrente, L. L. Elo, A. Brazma, and J. Rung. 2013. A fully scalable online pre-processing algorithm for short oligonucleotide microarray atlases. Nucleic Acids Res **41**:e110.

Langlands, S. J., M. J. Hopkins, N. Coleman, and J. H. Cummings. 2004. Prebiotic carbohydrates modify the mucosa associated microflora of the human large bowel. Gut **53**:1610-1616.

Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357-359.

Lapthorne, S., P. M. Pereira-Fantini, F. Fouhy, G. Wilson, S. L. Thomas, N. L. Dellios, M. Scurr, O. O'Sullivan, R. P. Ross, and C. Stanton. 2013. Gut microbial diversity is reduced and is associated with colonic inflammation in a piglet model of short bowel syndrome. Gut microbes **4**:0-9.

Larsen, F. M., P. J. Moughan, and M. N. Wilson. 1993. Dietary fiber viscosity and endogenous protein excretion at the terminal ileum of growing rats. Journal of Nutrition **123**:1898-1904.

Larsen, M., C. Juhl, N. Pørksen, C. Gotfredsen, R. Carr, U. Ribel, M. Wilken, and B. Rolin. 2005. β-Cell function and islet morphology in normal, obese, and obese β-cell mass-reduced Göttingen minipigs. American Journal of Physiology-Endocrinology And Metabolism **288**:E412-E421.

Larsen, M. O., B. Rolin, M. Wilken, R. D. Carr, and O. Svendsen. 2002. High-Fat High-Energy Feeding Impairs Fasting Glucose and Increases Fasting Insulin Levels in the Göttingen Minipig. Annals of the New York Academy of Sciences **967**:414-423.

Laycock, G., L. Sait, C. Inman, M. Lewis, H. Smidt, P. van Diemen, F. Jorgenson, M. Stevens, and M. Bailey. 2012. A defined intestinal colonization microbiota for gnotobiotic pigs. Veterinary Immunology and Immunopathology.

Lê Cao, K.-A., I. González, and S. Déjean. 2009. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics **25**:2855-2856.

Le Cao, K. A., P. G. Martin, C. Robert-Granie, and P. Besse. 2009b. Sparse canonical methods for biological data integration: application to a cross-platform study. BMC Bioinformatics **10**:34.

Leimena, M. M., J. Ramiro-Garcia, M. Davids, B. van den Bogert, H. Smidt, E. J. Smid, J. Boekhorst, E. G. Zoetendal, P. J. Schaap, and M. Kleerebezem. 2013. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC genomics **14**:530.

Leimena, M. M., M. Wels, R. S. Bongers, E. J. Smid, E. G. Zoetendal, and M. Kleerebezem. 2012. Comparative analysis of Lactobacillus plantarum WCFS1 transcriptomes by using DNA microarray and next-generation sequencing technologies. Applied and environmental microbiology **78**:4141-4148.

Lepš, J. and P. Šmilauer. 2003. Multivariate Analysis of Ecological Data using CANOCO. Cambridge University Press, Cambridge, UK.

Lerat, E., V. Daubin, and N. A. Moran. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the γ-Proteobacteria. PLoS biology **1**:e19.

Leser, T. D. and L. Molbak. 2009. Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host. Environmental Microbiology **11**:2194-2206.

Leung, H. C., S.-M. Yiu, J. Parkinson, and F. Y. Chin. 2013. IDBA-MT: de novo assembler for metatranscriptomic data generated from next-generation sequencing technology. Journal of Computational Biology **20**:540-550.

Ley, R. E., F. Backhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon. 2005. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A **102**:11070-11075.

Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight*, et al.* 2008a. Evolution of mammals and their gut microbes. Science **320**:1647-1651.

Ley, R. E., C. A. Lozupone, M. Hamady, R. Knight, and J. I. Gordon. 2008b. Worlds within worlds: evolution of the vertebrate gut microbiota. Nature reviews. Microbiology **6**:776-788.

Ley, R. E., P. J. Turnbaugh, S. Klein, and J. I. Gordon. 2006. Microbial ecology: human gut microbes associated with obesity. Nature **444**:1022-1023.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078-2079.

Li, S., W. Sauer, and M. Fan. 1993. The effect of dietary crude protein level on amino acid digestibility in early-weaned pigs. J Anim Physiol Anim Nutr (Berl) **70**:26-37.

Lin, H. V., A. Frassetto, E. J. Kowalik Jr, A. R. Nawrocki, M. M. Lu, J. R. Kosinski, J. A. Hubert, D. Szeto, X. Yao, G. Forrest*, et al.* 2012. Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. PLoS ONE **7**:e35240.

Lin, K., H. Kools, P. J. de Groot, A. K. Gavai, R. K. Basnet, F. Cheng, J. Wu, X. Wang, A. Lommen, G. J. Hooiveld*, et al.* 2011. MADMAX - Management and analysis database for multiple ~omics experiments. J Integr Bioinform **8**:160.

Liu, C., S. M. Finegold, Y. Song, and P. A. Lawson. 2008. Reclassification of Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus and Ruminococcus schinkii as Blautia coccoides gen. nov., comb. nov., Blautia hansenii comb. nov., Blautia hydrogenotrophica comb. nov., Blautia luti comb. nov., Blautia producta comb. nov., Blautia schinkii comb. nov. and description of Blautia wexlerae sp. nov., isolated from human faeces. International journal of systematic and evolutionary microbiology **58**:1896-1902.

Liu, X., J. M. Blouin, A. Santacruz, A. Lan, M. Andriamihaja, S. Wilkanowicz, P. H. Benetti, D. Tome, Y. Sanz, F. Blachier*, et al.* 2014. High-protein diet modifies colonic microbiota and luminal environment but not colonocyte metabolism in the rat model: the increased luminal bulk connection. Am J Physiol Gastrointest Liver Physiol **307**:G459-470.

Louis, P., S. H. Duncan, S. I. McCrae, J. Millar, M. S. Jackson, and H. J. Flint. 2004. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. Journal of bacteriology **186**:2099-2106.

Louis, P. and H. J. Flint. 2007. Development of a semiquantitative degenerate real-time pcr-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples. Applied and environmental microbiology **73**:2009-2012.

Louis, P. and H. J. Flint. 2009. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett **294**:1-8.

Louis, P., K. P. Scott, S. H. Duncan, and H. J. Flint. 2007. Understanding the effects of diet on bacterial metabolism in the large intestine. J Appl Microbiol **102**:1197-1208.

Louis, P., P. Young, G. Holtrop, and H. J. Flint. 2010. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. Environmental Microbiology **12**:304-314.

Lozupone, C., K. Faust, J. Raes, J. J. Faith, D. N. Frank, J. Zaneveld, J. I. Gordon,

and R. Knight. 2012a. Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles of human gut symbionts. Genome research **22**:1974-1984.

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

Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, and Y. Liu. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience **1**:18.

Lyon, M. F. and A. G. Searle. 1989. Genetic variants and strains of the laboratory mouse. Oxford University Press.

Macfarlane, G. T., C. Allison, S. A. Gibson, and J. H. Cummings. 1988. Contribution of the microflora to proteolysis in the human large intestine. The Journal of applied bacteriology **64**:37-46.

Macfarlane, G. T., J. H. Cummings, and C. Allison. 1986. Protein degradation by human intestinal bacteria. Journal of general microbiology **132**:1647-1656.

Macfarlane, G. T., G. R. Gibson, and J. H. Cummings. 1992. Comparison of fermentation reactions in different regions of the human colon. The Journal of applied bacteriology **72**:57-64.

Macfarlane, G. T., S. Hay, and G. R. Gibson. 1989. Influence of mucin on glycosidase, protease and arylamidase activities of human gut bacteria grown in a 3-stage continuous culture system. J Appl Microbiol **66**:407-417.

Macfarlane, S. and G. T. Macfarlane. 2003. Regulation of short-chain fatty acid production. Proc Nutr Soc **62**:67-72.

Macfarlane, S. and G. T. Macfarlane. 2006. Composition and metabolic activities of bacterial biofilms colonizing food residues in the human gut. Applied and environmental microbiology **72**:6204-6211.

Macfarlane, S., G. T. Macfarlane, and J. H. Cummings. 2006. Review article: prebiotics in the gastrointestinal tract. Alimentary pharmacology & therapeutics **24**:701-714.

Makivuokko, H., S. Lahtinen, P. Wacklin, E. Tuovinen, H. Tenkanen, J. Nikkila, M. Bjorklund, K. Aranko, A. Ouwehand, and J. Matto. 2012. Association between the ABO blood group and the human intestinal microbiota composition. BMC Microbiology **12**:94.

Mäkivuokko, H. and P. Nurminen. 2006. In Vitro Methods to Model the Gastrointestinal Tract. Pages 237-250 *in* A. C. Ouwehand and D. E. Vaughan, editors. Gastrointestinal microbiology. Informa; Taylos & Francis Group, New York.

Mallett, A., C. Bearne, I. Rowland, M. Farthing, C. Cole, and R. Fuller. 1987. The use of rats associated with a human faecal flora as a model for studying the effects of diet on the human gut microflora. Journal of Applied Microbiology **63**:39-45.

Manach, C., A. Scalbert, C. Morand, C. Rémésy, and L. Jiménez. 2004. Polyphenols: food sources and bioavailability. Am J Clin Nutr **79**:727-747.

Marco, M. L., M. C. de Vries, M. Wels, D. Molenaar, P. Mangell, S. Ahrne, W. M. de Vos, E. E. Vaughan, and M. Kleerebezem. 2010. Convergence in probiotic Lactobacillus gut-adaptive responses in humans and mice. The ISME journal

4:1481-1484.

Marco, M. L., T. H. Peters, R. S. Bongers, D. Molenaar, S. van Hemert, J. L. Sonnenburg, J. I. Gordon, and M. Kleerebezem. 2009. Lifestyle of Lactobacillus plantarum in the mouse caecum. Environmental Microbiology **11**:2747-2757.

Markle, J. G., D. N. Frank, S. Mortin-Toth, C. E. Robertson, L. M. Feazel, U. Rolle-Kampczyk, M. von Bergen, K. D. McCoy, A. J. Macpherson, and J. S. Danska. 2013. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science **339**:1084-1088.

Martens, E. C., R. Roth, J. E. Heuser, and J. I. Gordon. 2009. Coordinate Regulation of Glycan Degradation and Polysaccharide Capsule Biosynthesis by a Prominent Human Gut Symbiont. Journal of Biological Chemistry **284**:18445-18457.

Martin, F.-P. J., Y. Wang, N. Sprenger, I. K. S. Yap, T. Lundstedt, P. Lek, S. Rezzi, Z. Ramadan, P. van Bladeren, L. B. Fay*, et al.* 2008a. Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model. Molecular systems biology **4**.

Martin, F. P., M. E. Dumas, Y. Wang, C. Legido-Quigley, I. K. Yap, H. Tang, S. Zirah, G. M. Murphy, O. Cloarec, J. C. Lindon*, et al.* 2007. A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. Mol Syst Biol **3**:112.

Martin, F. P., N. Sprenger, I. K. Yap, Y. Wang, R. Bibiloni, F. Rochat, S. Rezzi, C. Cherbut, S. Kochhar, J. C. Lindon*, et al.* 2009. Panorganismal gut microbiome-host metabolic crosstalk. J Proteome Res **8**:2090-2105.

Martin, F. P., Y. Wang, N. Sprenger, I. K. Yap, S. Rezzi, Z. Ramadan, E. Pere-Trepat, F. Rochat, C. Cherbut, P. van Bladeren*, et al.* 2008b. Top-down systems biology integration of conditional prebiotic modulated transgenomic interactions in a humanized microbiome mouse model. Molecular systems biology **4**:205.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal **17**:pp. 10-12.

Martinez, R. C., A. E. Aynaou, S. Albrecht, H. A. Schols, E. C. De Martinis, E. G. Zoetendal, K. Venema, S. M. Saad, and H. Smidt. 2011. In vitro evaluation of gastrointestinal survival of *Lactobacillus amylovorus* DSM 16698 alone and combined with galactooligosaccharides, milk and/or *Bifidobacterium animalis* subsp. lactis Bb-12. International Journal Of Food Microbiology **149**:152-158.

Martinez, R. C., H. R. Cardarelli, W. Borst, S. Albrecht, H. Schols, O. P. Gutierrez, A. J. Maathuis, B. D. de Melo Franco, E. C. De Martinis, E. G. Zoetendal*, et al.* 2013. Effect of galactooligosaccharides and Bifidobacterium animalis Bb-12 on growth of Lactobacillus amylovorus DSM 16698, microbial community structure, and metabolite production in an in vitro colonic model set up with human or pig microbiota. FEMS microbiology ecology **84**:110-123.

Martins dos Santos, V., M. Muller, and W. M. de Vos. 2010. Systems biology of the gut: the interplay of food, microbiota and host at the mucosal interface. Current opinion in biotechnology **21**:539-550.

Matamoros, S., C. Gras-Leguen, F. Le Vacon, G. Potel, and M.-F. de La Cochetiere. 2013. Development of intestinal microbiota in infants and its impact on health. Trends in Microbiology **21**:167-173.

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

McNeil, N. I. 1984. The contribution of the large intestine to energy supplies in man. The American journal of clinical nutrition **39**:338-342.

McOrist, A. L., R. B. Miller, A. R. Bird, J. B. Keogh, M. Noakes, D. L. Topping, and M. A. Conlon. 2011. Fecal Butyrate Levels Vary Widely among Individuals but Are Usually Increased by a Diet High in Resistant Starch. Journal of Nutrition **141**:883-889.

Merico, D., R. Isserlin, O. Stueker, A. Emili, and G. D. Bader. 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One **5**:e13984.

Miller, E. and D. Ullrey. 1987. The pig as a model for human nutrition. Annual review of nutrition **7**:361-382.

Miller, T. L. and M. J. Wolin. 1979. Fermentations by saccharolytic intestinal bacteria. Am J Clin Nutr **32**:164-172.

Moore, W. E. and L. V. Holdeman. 1974. Human fecal flora: the normal flora of 20 Japanese-Hawaiians. Appl Microbiol **27**:961-979.

Morgulis, A., G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala, and A. A. Schaffer. 2008. Database indexing for production MegaBLAST searches. Bioinformatics **24**:1757-1764.

Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res **35**:W182-W185.

Morse III, H. C. 1978. Origins of inbred mice. Academic Press.

Mortensen, P. B. and M. R. Clausen. 1996. Short-chain fatty acids in the human colon: relation to gastrointestinal health and disease. Scand J Gastroenterol Suppl **216**:132-148.

Muegge, B. D., J. Kuczynski, D. Knights, J. C. Clemente, A. Gonzalez, L. Fontana, B. Henrissat, R. Knight, and J. I. Gordon. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science **332**:970-974.

Mulder, I., B. Schmidt, C. Stokes, M. Lewis, M. Bailey, R. Aminov, J. Prosser, B. Gill, J. Pluske, and C.-D. Mayer. 2009. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC biology **7**:79.

Mullaney, J. A. 2013. The biotransformation of glucosinolates from a bacterial perspective. Massey, Palmerston North.

Mullaney, J. A., J. Ansell, W. J. Kelly, and J. A. Heyes. 2013. The biotransformation of glucosinolates from a bacterial perspective. CAB Reviews.

Müller, M. and S. Kersten. 2003. Nutrigenomics: goals and strategies. Nat Rev Genet **4**:315-322.

Muñoz-Tamayo, R., B. Laroche, É. Walter, J. Doré, S. H. Duncan, H. J. Flint, and M. Leclerc. 2011. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. FEMS Microbiol Ecol **76**:615-624.

Musso, G., R. Gambino, and M. Cassader. 2010. Obesity, diabetes, and gut microbiota the hygiene hypothesis expanded? Diabetes care **33**:2277-2284.

Nabuurs, M., A. Hoogendoorn, F. Van Zijderveld, and J. Van der Klis. 1993. A long-term perfusion test to measure net absorption in the small intestine of weaned pigs. Research in veterinary science **55**:108-114.

Nadkarni, P. M. and C. R. Parikh. 2012. An eUtils toolset and its use for creating a pipeline to link genomics and proteomics analyses to domain-specific biomedical literature. Journal of clinical bioinformatics **2**:1-7.

Nagarajan, N. and M. Pop. 2013. Sequence assembly demystified. Nat Rev Genet **14**:157-167.

Namiki, T., T. Hachiya, H. Tanaka, and Y. Sakakibara. 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res **40**:e155-e155.

Nedrud, J. G. 2006. Animal models for gastric Helicobacter immunology and vaccine studies. FEMS Immunology & Medical Microbiology **24**:243-250.

Nielsen, H. B., M. Almeida, A. S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, D. R. Plichta, L. Gautier, A. G. Pedersen, and E. Le Chatelier. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol **32**:822-828.

Niewold, T. A., H. H. D. Kerstens, J. van der Meulen, M. A. Smits, and M. M. Hulst. 2005. Development of a porcine small intestinal cDNA micro-array: characterization and functional analysis of the response to enterotoxigenic E. coli. Veterinary Immunology and Immunopathology **105**:317-329.

Niewold, T. A., J. Van der Meulen, H. H. Kerstens, M. A. Smits, and M. M. Hulst. 2010. Transcriptomics of enterotoxigenic *Escherichia coli* infection. Individual variation in intestinal gene expression correlates with intestinal function. Veterinary microbiology **141**:110-114.

Niewold, T. A., E. J. A. Veldhuizen, J. van der Meulen, H. P. Haagsman, A. A. C. de Wit, M. A. Smits, M. H. G. Tersteeg, and M. M. Hulst. 2007. The early transcriptional response of pig small intestinal mucosa to invasion by Salmonella enterica serovar typhimurium DT104. Molecular Immunology **44**:1316-1322.

Norat, T., D. Chan, R. Lau, D. Aune, and R. Vieira. 2010. WCRF/AICR Systematic Literature Review Continuous Update Project Report. Pages 1-855.

Nyangale, E. P., D. S. Mottram, and G. R. Gibson. 2012. Gut microbial activity, implications for health and disease: the potential role of metabolite analysis. J Proteome Res **11**:5573-5585.

O'Hara, A. M. and F. Shanahan. 2006. The gut flora as a forgotten organ. EMBO reports **7**:688-693.

O'Brien, C. L., G. E. Allison, F. Grimpen, and P. Pavli. 2013. Impact of colonoscopy bowel preparation on intestinal microbiota. PLoS ONE **8**:e62815.

Ouwehand, A. C., M. Derrien, W. de Vos, K. Tiihonen, and N. Rautonen. 2005. Prebiotics and other microbial substrates for gut functionality. Current opinion in biotechnology **16**:212-217.

Pace, N. R., J. Sapp, and N. Goldenfeld. 2012. Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. Proc Natl Acad Sci U S A **109**:1011-1018.

Pang, X., X. Hua, Q. Yang, D. Ding, C. Che, L. Cui, W. Jia, P. Bucheli, and L. Zhao. 2007. Inter-species transplantation of gut microbiota from human to pigs. The ISME Journal **1**:156-162.

Parter, M., N. Kashtan, and U. Alon. 2007. Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol **7**:169.

Pedersen, R., H.-C. Ingerslev, M. Sturek, M. Alloosh, S. Cirera, B. Ø. Christoffersen,

S. G. Moesgaard, N. Larsen, and M. Boye. 2013. Characterisation of Gut Microbiota in Ossabaw and Göttingen Minipigs as Models of Obesity and Metabolic Syndrome. PloS one **8**:e56612.

Peng, Y., H. C. Leung, S.-M. Yiu, and F. Y. Chin. 2010. IDBA–a practical iterative de Bruijn graph de novo assembler. Pages 426-440 *in* Research in Computational Molecular Biology. Springer.

Piccolo, S. R., M. R. Withers, O. E. Francis, A. H. Bild, and W. E. Johnson. 2013. Multiplatform single-sample estimates of transcriptional activation. Proc Natl Acad Sci U S A **110**:17778-17783.

Pittet, V., T. G. Phister, and B. Ziola. 2013. Transcriptome Sequence and Plasmid Copy Number Analysis of the Brewery Isolate Pediococcus claussenii ATCC BAA-344T during Growth in Beer. PloS one **8**:e73627.

Pluznick, J. L., R. J. Protzko, H. Gevorgyan, Z. Peterlin, A. Sipos, J. Han, I. Brunet, L.-X. Wan, F. Rey, T. Wang*, et al.* 2013. Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. Proceedings of the National Academy of Sciences of the United States of America **110**:4410-4415.

Prakash, O., P. K. Pandey, G. J. Kulkarni, K. N. Mahale, and Y. S. Shouche. 2014. Technicalities and Glitches of Terminal Restriction Fragment Length Polymorphism (T-RFLP). Indian J Microbiol **54**:255-261.

Puiman, P. and B. Stoll. 2008. Animal models to study neonatal nutrition in humans. Current Opinion in Clinical Nutrition & Metabolic Care **11**:601-606 610.1097/MCO.1090b1013e32830b32835b32815.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada*, et al.* 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464**:59-65.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. 2005. InterProScan: protein domains identifier. Nucleic Acids Res **33**:W116-W120.

Quinlan, A. R. and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**:841-842.

Rajilic-Stojanovic, M. and W. M. de Vos. 2014. The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev.

Rajilic-Stojanovic, M., H. G. Heilig, D. Molenaar, K. Kajander, A. Surakka, H. Smidt, and W. M. de Vos. 2009. Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. Environmental Microbiology **11**:1736-1751.

Rajilic-Stojanovic, M., H. Smidt, and W. M. de Vos. 2007. Diversity of the human gastrointestinal tract microbiota revisited. Environmental Microbiology **9**:2125-2136.

Rawls, J. F., M. A. Mahowald, R. E. Ley, and J. I. Gordon. 2006. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. Cell **127**:423-433.

Reikvam, D. H., M. Derrien, R. Islam, A. Erofeev, V. Grcic, A. Sandvik, P. Gaustad, L. A. Meza-Zepeda, F. L. Jahnsen, H. Smidt*, et al.* 2012. Epithelial-microbial crosstalk in polymeric Ig receptor deficient mice. Eur J Immunol **42**:2959-2970.

Reyes, A., M. Haynes, N. Hanson, F. E. Angly, A. C. Heath, F. Rohwer, and J. I. Gordon. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466**:334-338.

Richards, V. P., S. C. Choi, P. D. Pavinski Bitar, A. A. Gurjar, and M. J. Stanhope. 2013. Transcriptomic and genomic evidence for Streptococcus agalactiae adaptation to the bovine environment. BMC genomics **14**:920.

Ridaura, V. K., J. J. Faith, F. E. Rey, J. Cheng, A. E. Duncan, A. L. Kau, N. W. Griffin, V. Lombard, B. Henrissat, J. R. Bain*, et al.* 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science **341**:1241214.

Riviere, A., F. Moens, M. Selak, D. Maes, S. Weckx, and L. De Vuyst. 2014. The ability of bifidobacteria to degrade arabinoxylan oligosaccharide constituents and derived oligosaccharides is strain dependent. Applied and environmental microbiology **80**:204-217.

Rodenburg, W., J. Keijer, E. Kramer, C. Vink, R. van der Meer, and I. M. Bovee-Oudenhoven. 2008. Impaired barrier function by dietary fructo-oligosaccharides (FOS) in rats is accompanied by increased colonic mitochondrial gene expression. BMC Genomics **9**:144.

Rodewald, R. 1976. pH-dependent binding of immunoglobulins to intestinal cells of the neonatal rat. The Journal of cell biology **71**:666-669.

Rosendale, D. I., A. Cookson, N. Roy, and I. Vetharaniam. 2011. Opportunities for predictive modelling and gut health. Conceptually exploring an *in silico* tool for quantifying the benefits of dietary fibre consumption. AgroFOOD industry hi-tech **22**:18-22.

Round, J. L. and S. K. Mazmanian. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol **9**:313-323.

Russell, W. R., L. Hoyles, H. J. Flint, and M. E. Dumas. 2013. Colonic bacterial metabolites and human health. Curr Opin Microbiol **16**:246-254.

Sakamoto, M. and Y. Benno. 2006. Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. Int J Syst Evol Microbiol **56**:1599-1605.

Salzman, N. H., H. de Jong, Y. Paterson, H. J. Harmsen, G. W. Welling, and N. A. Bos. 2002. Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. Microbiology **148**:3651-3660.

Sangild, P. T., R. H. Siggers, M. Schmidt, J. Elnif, C. R. Bjornvad, T. Thymann, M. L. Grondahl, A. K. Hansen, S. K. Jensen, and M. Boye. 2006. Diet-and colonization-dependent intestinal dysfunction predisposes to necrotizing enterocolitis in preterm pigs. Gastroenterology **130**:1776-1792.

Sartor, M. A., C. R. Tomlinson, S. C. Wesselkamper, S. Sivaganesan, G. D. Leikauf, and M. Medvedovic. 2006. Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments. BMC Bioinformatics **7**:538.

Sauer, W. and K. De Lange. 1992. Novel methods for determining protein and amino acid digestibilities in feedstuffs. Modern methods in protein nutrition and metabolism:87-120.

Savage, D. C. 1977. Microbial ecology of the gastrointestinal tract. Annual review of microbiology **31**:107-133.

Scanlan, P. D. and J. R. Marchesi. 2008. Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. The ISME journal **2**:1183-1193.

Scanlan, P. D., C. R. Stensvold, M. Rajilic-Stojanovic, H. G. Heilig, W. M. De Vos, P. W. O'Toole, and P. D. Cotter. 2014. The microbial eukaryote Blastocystis is a prevalent and diverse member of the healthy human gut microbiota. FEMS microbiology ecology.

Scharlau, D., A. Borowicki, N. Habermann, T. Hofmann, S. Klenow, C. Miene, U. Munjal, K. Stein, and M. Glei. 2009. Mechanisms of primary cancer prevention by butyrate and other products formed during gut flora-mediated fermentation of dietary fibre. Mutation Research **682**:39-53.

Scheppach, W. 1994. Effects of short chain fatty acids on gut morphology and function. Gut **35**:S35-38.

Scheppach, W., J. G. Müller, F. Boxberger, G. Dusel, F. Richter, H. P. Bartram, S. U. Christl, C. E. Dempfle, and H. Kasper. 1997. Histological changes in the colonic mucosa following irrigation with short-chain fatty acids. European Journal of Gastroenterology and Hepatology **2**:163-168.

Scheppach, W., H. Sommer, T. Kirchner, G. M. Paganelli, P. Bartram, S. Christl, F. Richter, G. Dusel, and H. Kasper. 1992. Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. Gastroenterology **1**:51-56.

Schmidt, B., I. E. Mulder, C. C. Musk, R. I. Aminov, M. Lewis, C. R. Stokes, M. Bailey, J. I. Prosser, B. P. Gill, and J. R. Pluske. 2011. Establishment of normal gut microbiota is compromised under excessive hygiene conditions. PLoS One **6**:e28284.

Schmidt, D., M. D. Wilson, C. Spyrou, G. D. Brown, J. Hadfield, and D. T. Odom. 2009. ChIP-seq: Using high-throughput sequencing to discover protein–DNA interactions. Methods **48**:240-248.

Schmieder, R. and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**:863-864.

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics **28**:1086-1092.

Schwarz, J., D. Tome, A. Baars, G. J. Hooiveld, and M. Muller. 2012. Dietary protein affects gene expression and prevents lipid accumulation in the liver in mice. PloS one **7**:e47303.

Scott, K. P., S. W. Gratz, P. O. Sheridan, H. J. Flint, and S. H. Duncan. 2013. The influence of diet on the gut microbiota. Pharmacol Res **69**:52-60.

Seedorf, H., W. F. Fricke, B. Veith, H. Bruggemann, H. Liesegang, A. Strittmatter, M. Miethke, W. Buckel, J. Hinderberger, F. Li*, et al.* 2008. The genome of Clostridium kluyveri, a strict anaerobe with unique metabolic features. Proc Natl Acad Sci U S A **105**:2128-2133.

Shen, J., B. Zhang, H. Wei, C. Che, D. Ding, X. Hua, P. Bucheli, L. Wang, Y. Li, and X. Pang. 2010. Assessment of the modulating effects of fructo-oligosaccharides on fecal microbiota using human flora–associated piglets. Archives of microbiology **192**:959-968.

Shulman, R. J., S. J. Henning, and B. L. Nichols. 1988. The Miniature Pig as an Animal Model for the Study of Intestinal Enzyme Development. Pediatr Res

23:311-315.
Siggers, R. H., J. Siggers, T. Thymann, M. Boye, and P. T. Sangild. 2011. Nutritional modulation of the gut microbiota and immune system in preterm neonates susceptible to necrotizing enterocolitis. The Journal of nutritional biochemistry **22**:511-521.
Slavin, J. L. 1987. Dietary fiber: classification, chemical analyses, and food sources. J Am Diet Assoc **87**:1164-1171.
Slavin, J. L. 2013. Fiber and prebiotics: mechanisms and health benefits. Nutrients **5**:1417-1435.
Smith, E. A. and G. T. Macfarlane. 1997. Formation of Phenolic and Indolic Compounds by Anaerobic Bacteria in the Human Large Intestine. Microbial ecology **33**:180-188.
Smoot, L. M., J. C. Smoot, H. Smidt, P. A. Noble, M. Konneke, Z. A. McMurry, and D. A. Stahl. 2005. DNA microarrays as salivary diagnostic tools for characterizing the oral cavity's microbial community. Advances in dental research **18**:6-11.
Smyth, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3**:Article3.
Sonnenburg, J. L., J. Xu, D. D. Leip, C. H. Chen, B. P. Westover, J. Weatherford, J. D. Buhler, and J. I. Gordon. 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science **307**:1955-1959.
Stein, H., C. Shipley, and R. Easter. 1998. Technical note: a technique for inserting a T-cannula into the distal ileum of pregnant sows. Journal of animal science **76**:1433-1436.
Stevens, C. E. 1977. Comparative physiology of the digestive system.*in* M. J. Swenson, editor. Dukes Physiology of Domestic Animals, Comstock, Ithaca, NY and London.
Stevens, C. E. and I. D. Hume. 1998. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. Physiological Reviews **78**:393-427.
Su, W., C. R. Bush, B. M. Necela, S. R. Calcagno, N. R. Murray, A. P. Fields, and E. A. Thompson. 2007. Differential expression, distribution, and function of PPAR-gamma in the proximal and distal colon. Physiol Genomics **30**:342-353.
Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander*, et al.* 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A **102**:15545-15550.
Sukemori, S., S. Ikeda, Y. Kurihara, and S. Ito. 2003. Amino acid, mineral and vitamin levels in hydrous faeces obtained from coprophagy-prevented rats. J Anim Physiol Anim Nutr (Berl) **87**:213-220.
Suzuki, K., B. Meek, Y. Doi, M. Muramatsu, T. Chiba, T. Honjo, and S. Fagarasan. 2004. Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. Proceedings of the National Academy of Sciences, USA **101**:1981-1986.
Suzuki, M. T., L. T. Taylor, and E. F. DeLong. 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Applied and environmental microbiology **66**:4605-4614.
Tasse, L., J. Bercovici, S. Pizzut-Serin, P. Robe, J. Tap, C. Klopp, B. L. Cantarel, P.

M. Coutinho, B. Henrissat, M. Leclerc, *et al.* 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome research **20**:1605-1612.

Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. Science **278**:631-637.

ter Braak, C. 1987. Ordination. Pages 91-173 *in* C. J. F. ter Braak and O. van Tongeren, editors. Data analysis in community and landscape ecology. Unipub.

ter Braak, C. J. F. and S. P. 2012. Canoco reference manual and user's guide: software for ordination, version 5.0. Microcomputer Power.

Thompson-Chagoyán, O., J. Maldonado, and A. Gil. 2007. Colonization and Impact of Disease and Other Factors on Intestinal Microbiota. Digestive Diseases and Sciences **52**:2069-2077.

Thompson, C. L., B. Wang, and A. J. Holmes. 2008. The immediate environment during postnatal development has long-term impact on gut community structure in pigs. The ISME Journal **2**:739-748.

Tirosh, B. and A. Rubinstein. 1998. Migration of adhesive and nonadhesive particles in the rat intestine under altered mucus secretion conditions. J Pharm Sci **87**:453-456.

Topping, D. L. and P. M. Clifton. 2001. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiological Reviews **81**:1031-1064.

Tringe, S. G. and P. Hugenholtz. 2008. A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol **11**:442-446.

Turnbaugh, P. J., F. Backhed, L. Fulton, and J. I. Gordon. 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe **3**:213-223.

Turnbaugh, P. J. and J. I. Gordon. 2009. The core gut microbiome, energy balance and obesity. J Physiol **587**:4153-4158.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. 2007. The human microbiome project. Nature **449**:804-810.

Turnbaugh, P. J., R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444**:1027-1031.

Turnbaugh, P. J., C. Quince, J. J. Faith, A. C. McHardy, T. Yatsunenko, F. Niazi, J. Affourtit, M. Egholm, B. Henrissat, and R. Knight. 2010. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proceedings of the National Academy of Sciences **107**:7503-7508.

Turnbaugh, P. J., V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight, and J. I. Gordon. 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med **1**:6ra14.

Turroni, F., M. Ventura, L. F. Butto, S. Duranti, P. W. O'Toole, M. O. Motherway, and D. van Sinderen. 2014. Molecular dialogue between the human gut microbiota and the host: a Lactobacillus and Bifidobacterium perspective. Cell Mol Life Sci **71**:183-203.

Urich, T., A. Lanzén, J. Qi, D. H. Huson, C. Schleper, and S. C. Schuster. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PloS one **3**:e2527.

Valdés, L., P. Gullón, N. Salazar, D. Rios-Covián, M. J. González-Muñoz, J. C. Parajó, P. Ruas-Madiedo, M. Gueimonde, and G. Clara. 2013. Population Dynamics of Some Relevant Intestinal Microbial Groups in Human Fecal Batch Cultures with Added Fermentable Xylooligosaccharides Obtained from Rice Husks. BioResources **8**:2429-2441.

Van den Abbeele, P., C. Belzer, M. Goossens, M. Kleerebezem, W. M. De Vos, O. Thas, R. De Weirdt, F.-M. Kerckhof, and T. Van de Wiele. 2012a. Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. The ISME journal **7**:949-961.

Van den Abbeele, P., C. Belzer, M. Goossens, M. Kleerebezem, W. M. De Vos, O. Thas, R. De Weirdt, F. M. Kerckhof, and T. Van de Wiele. 2013a. Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. ISME J **7**:949-961.

Van den Abbeele, P., P. Gérard, S. Rabot, A. Bruneau, S. El Aidy, M. Derrien, M. Kleerebezem, E. G. Zoetendal, H. Smidt, and W. Verstraete. 2011. Arabinoxylans and inulin differentially modulate the mucosal and luminal gut microbiota and mucin-degradation in humanized rats. Environmental Microbiology **13**:2667-2680.

Van den Abbeele, P., C. Grootaert, M. Marzorati, S. Possemiers, W. Verstraete, P. Gerard, S. Rabot, A. Bruneau, S. El Aidy, M. Derrien*, et al.* 2010. Microbial community development in a dynamic gut model is reproducible, colon region specific, and selective for Bacteroidetes and Clostridium cluster IX. Applied and environmental microbiology **76**:5237-5246.

Van den Abbeele, P., S. Roos, V. Eeckhaut, D. A. MacKenzie, M. Derde, W. Verstraete, M. Marzorati, S. Possemiers, B. Vanhoecke, F. Van Immerseel*, et al.* 2012b. Incorporating a mucosal environment in a dynamic gut model results in a more representative colonization by lactobacilli. Microb Biotechnol **5**:106-115.

Van den Abbeele, P., K. Venema, T. Van de Wiele, W. Verstraete, and S. Possemiers. 2013b. Different human gut models reveal the distinct fermentation patterns of Arabinoxylan versus inulin. J Agric Food Chem **61**:9819-9827.

van den Bogert, B., W. M. de Vos, E. G. Zoetendal, and M. Kleerebezem. 2011. Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. Applied and environmental microbiology **77**:2071-2080.

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

Van den Bogert, B., M. M. Leimena, W. M. De Vos, E. G. Zoetendal, and M. Kleerebezem. 2011. Functional Intestinal Metagenomics. In F. J. De Bruin (ed.), Handbook of Molecular Microbial Ecology **Vol II: Metagenomics in Different Habitats. Wiley-Blackwell.**

Van der Meulen, J., M. Hulst, M. Smits, and T. Schuurman. 2010. Small intestinal segment perfusion test in piglets: future applications in studying probiotics-gut crosstalk in infectious diarrhoea? Beneficial Microbes **1**:439-445.

van Zanten, G. C., A. Knudsen, H. Röytiö, S. Forssten, M. Lawther, A. Blennow, S. J. Lahtinen, M. Jakobsen, B. Svensson, and L. Jespersen. 2012. The Effect of

Selected Synbiotics on Microbial Composition and Short-Chain Fatty Acid Production in a Model System of the Human Colon. PLoS ONE **7**:e47212.

Vanhoutvin, S. A., F. J. Troost, H. M. Hamer, P. J. Lindsey, G. H. Koek, D. M. Jonkers, A. Kodde, K. Venema, and R. J. Brummer. 2009. Butyrate-induced transcriptional changes in human colonic mucosa. PLoS One **4**:e6759.

Variyam, E. and L. Hoskins. 1983. In vitro degradation of gastric mucin. Carbohydrate side chains protect glycopeptides core from pancreatic proteases. Gastroenterology **84**:533-537.

Veldhuizen, E. J. A., A. van Dijk, M. H. G. Tersteeg, S. I. C. Kalkhove, J. van der Meulen, T. A. Niewold, and H. P. Haagsman. 2007. Expression of β-defensins pBD-1 and pBD-2 along the small intestinal tract of the pig: Lack of upregulation in vivo upon Salmonella typhimurium infection. Molecular Immunology **44**:276-283.

Vijay-Kumar, M., J. D. Aitken, F. A. Carvalho, T. C. Cullender, S. Mwangi, S. Srinivasan, S. V. Sitaraman, R. Knight, R. E. Ley, and A. T. Gewirtz. 2010. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science **328**:228-231.

Vital, M., A. C. Howe, and J. M. Tiedje. 2014. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. MBio **5**:e00889.

Vital, M., C. R. Penton, Q. Wang, V. B. Young, D. A. Antonopoulos, M. L. Sogin, H. G. Morrison, L. Raffals, E. B. Chang, G. B. Huffnagle*, et al.* 2013. A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community. Microbiome **1**:8.

Walker, W., G. Morgan, and C. Maxwell. 1986. Ileal cannulation in baby pigs with a simple T-cannula. Journal of animal science **62**:407-411.

Walter, J. and R. Ley. 2011. The human gut microbiome: ecology and recent evolutionary changes. Annual review of microbiology **65**:411-429.

Wang, Y., X. Li, and H. P. Blaschek. 2013. Effects of supplementary butyrate on butanol production and the metabolic switch in Clostridium beijerinckii NCIMB 8052: genome-wide transcriptional analysis with RNA-Seq. Biotechnology for biofuels **6**:138.

Wannemuehler, M. J., A.-M. Overstreet, D. V. Ward, and G. J. Phillips. 2014. Draft genome sequences of the altered Schaedler flora, a defined bacterial community from gnotobiotic mice. Genome Announc **2**:e00287-00214.

Westerterp-Plantenga, M. S., S. G. Lemmens, and K. R. Westerterp. 2012. Dietary protein - its role in satiety, energetics, weight loss and health. The British journal of nutrition **108 Suppl 2**:S105-112.

Westerterp-Plantenga, M. S., A. Nieuwenhuizen, D. Tome, S. Soenen, and K. R. Westerterp. 2009. Dietary protein, weight loss, and weight maintenance. Annu Rev Nutr **29**:21-41.

Windey, K., V. De Preter, T. Louat, F. Schuit, J. Herman, G. Vansant, and K. Verbeke. 2012a. Modulation of protein fermentation does not affect fecal water toxicity: a randomized cross-over study in healthy subjects. PloS one **7**:e52387.

Windey, K., V. De Preter, and K. Verbeke. 2012b. Relevance of protein fermentation to gut health. Molecular nutrition & food research **56**:184-196.

Winitz, M., R. F. Adams, D. A. Seedman, P. N. Davis, L. G. Jayko, and J. A. Hamilton.

1970. Studies in metabolic nutrition employing chemically defined diets. II. Effects on gut microflora populations. Am J Clin Nutr **23**:546-559.

Woese, C. R. and G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A **74**:5088-5090.

Wong, J. M., R. de Souza, C. W. Kendall, A. Emam, and D. J. Jenkins. 2006. Colonic health: fermentation and short chain fatty acids. Journal of Clinical Gastroenterology **40**:235-243.

Wu, G. D., J. Chen, C. Hoffmann, K. Bittinger, Y. Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight*, et al.* 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science **334**:105-108.

Xiong, X., D. N. Frank, C. E. Robertson, S. S. Hung, J. Markle, A. J. Canty, K. D. McCoy, A. J. Macpherson, P. Poussier, J. S. Danska*, et al.* 2012. Generation and analysis of a mouse intestinal metatranscriptome through Illumina based RNA-sequencing. PloS one **7**:e36009.

Yamada, T., I. Letunic, S. Okuda, M. Kanehisa, and P. Bork. 2011. iPath2. 0: interactive pathway explorer. Nucleic Acids Res **39**:W412-W415.

Yang, Y. and S. A. Smith. 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. BMC genomics **14**:328.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC bioinformatics **13**:134.

Yin, Y., Y. Wang, L. Zhu, W. Liu, N. Liao, M. Jiang, B. Zhu, H. D. Yu, C. Xiang, and X. Wang. 2013. Comparative analysis of the distribution of segmented filamentous bacteria in humans, mice and chickens. ISME Journal **7**:615-621.

Yoder-Himes, D. R., P. S. Chain, Y. Zhu, O. Wurtzel, E. M. Rubin, J. M. Tiedje, and R. Sorek. 2009. Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. Proc Natl Acad Sci U S A **106**:3976-3981.

Yu, Z. and M. Morrison. 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. BioTechniques **36**:808-812.

Ze, X., S. H. Duncan, P. Louis, and H. J. Flint. 2012. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. The ISME journal **6**:1535-1543.

Zerbino, D. R. and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research **18**:821-829.

Zhang, C., M. Zhang, S. Wang, R. Han, Y. Cao, W. Hua, Y. Mao, X. Zhang, X. Pang, and C. Wei. 2009. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. The ISME journal **4**:232-241.

Zoetendal, E. G., C. C. Booijink, E. S. Klaassens, H. G. Heilig, M. Kleerebezem, H. Smidt, and W. M. de Vos. 2006. Isolation of RNA from bacterial samples of the human gastrointestinal tract. Nature protocols **1**:954-959.

Zoetendal, E. G. and W. M. de Vos. 2014. Effect of diet on the intestinal microbiota and its activity. Curr Opin Gastroenterol **30**:189-195.

Zoetendal, E. G., J. Raes, B. van den Bogert, M. Arumugam, C. C. Booijink, F. J. Troost, P. Bork, M. Wels, W. M. de Vos, and M. Kleerebezem. 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. The ISME journal **6**:1415-1426.

Appendices

Nederlandse samenvatting	188
Curriculum vitae	191
List of publications	192
Co-author affiliations	194
Overview of completed training activities	195
Dankwoord	197

Nederlandse samenvatting

Ons darmstelsel wordt bevolkt door micro-organismen, gezamenlijk microbiota genoemd. Deze microbiota bestaan uit organismen uit alle drie de domeinen van het leven (Eukarya, Bacteria en Archaea). Hun totale aantal in ons darmstelsel is tien maal groter dan het aantal cellen in ons lichaam. De bacteriën vormen verreweg de grootste groep, en deze zijn in deze thesis verder bestudeerd. Het woord "microbiota" wordt verder in deze thesis gebruikt om de bacteriële groep microbiota aan te duiden binnen de microbiota in het darmstelsel. Het bestuderen van microbiota kan gedaan worden door middel van isolatie en kweken van de individuele soorten. Echter, slechts een kwart van deze soorten kan op dit moment gekweekt worden. Hierdoor worden met name technieken gebruikt die niet afhankelijk zijn van het isoleren en kweken van deze organismen. Dit kan omdat het RNA en DNA van deze organismen geïsoleerd en gesequenced (de volgorde van de nucleotiden, ATCG, bepalen) kan worden. Ook kunnen unieke sequenties bepaald worden met behulp van micro-array methodes. Met deze sequenties kunnen soorten en genen via grote databanken getraceerd worden. Op deze wijze is het mogelijk te bepalen welke organismen er in de darmmonsters aanwezig zijn.
Op dit moment is het bekend dat er in een gram fecaal materiaal uit de dikke darm 10^{11} bacteriën voorkomen, die behoren tot negen verschillende phyla, waarbij de phyla Firmicutes en Bacteroidetes het talrijkst zijn. Deze microbiota hebben de functie het onverteerd materiaal uit onze voeding, en de afgescheiden lichaamsproducten te fermenteren (het omzetten van organisch materiaal) tot stoffen, die voor de mens wel bruikbaar zijn. Deze stoffen bestaan voornamelijk uit korte-keten vetzuren (KKV). De meest geproduceerde KKV zijn acetaat, propionaat en butyraat. Al deze KKV kunnen door onze darmwand opgenomen worden, waarbij met name butyraat als energie bron voor de darmwandcellen fungeert. De overblijvende KKV worden via de darmwand opgenomen in het bloed en getransporteerd naar de lever, waarna de KKV in het lichaam terecht komen. Van met name butyraat is bekend dat het positieve effecten heeft in het lichaam (zie Tabel 3.1). Het is op dit moment al bekend dat de KKV productie beïnvloed kan worden door het gebruik van vezels in het dieet. In Hoofdstuk 3 wordt samengevat wat er bekend is van de veranderingen in KKV concentraties in de darm door de aanwezigheid van verschillende vezels in het dieet. Vooral veranderingen in de butyraatproductie worden veroorzaakt door secundaire effecten. De microbiota die de vezels afbreken zijn vaak niet degene die het butyraat aanmaken. Zij produceren voornamelijk acetaat en/of lactaat. Acetaat en lactaat kunnen door een andere groep microbiota gebruikt worden als energie bron en omgezet tot butyraat. De butyraat producenten behoren tot het phylum Firmicutes en bevinden zich in de *Clostridium* klas IV en XIVa.
In het project waarbinnen deze thesis gedaan is waren wij geïnteresseerd in de KKV productie door de microbiota en de opname en verwerking van deze KKV door de gastheer (zie Figuur 1.2). De onderzoeken die in deze thesis beschreven staan

hebben zich geconcentreerd op veranderingen, door middel van verschillende diëten, van de microbiota, hun KKV productie en de reactie van de gastheer op de KKV. Om dit te bestuderen hebben we gebruik gemaakt de muis als proefdier model. De beweegredenen om een diermodel te gebruiken in plaats van de mens zelf staan beschreven in Hoofdstuk 2. In de onderzoeken in deze thesis hadden we door het gebruik van muizen de mogelijkheid om op de plaats van de KKV productie door de microbiota en de opname ervan door de gastheer metingen te doen. Er is nagegaan welke microbiota KVV in het lichaam produceerden en hoe de gastheer, de muis, op deze KKV reageerde.

In Hoofdstuk 4 wordt beschreven hoe verschillende vezels in het colon (het laatste deel van de dikke darm) de microbiota (via het bacterieel DNA), de KKV concentraties en de expressiepatronen (patroon van het RNA in cellen, dit geeft aan welke genen er actief zijn) van de darmwandcellen van de gastheer veranderen. Uit dit onderzoek bleek dat er vier soorten vezels (fructo-oligosacchariden, inuline, arabinoxylaan en guar gum) zorgden voor een vergelijkbare verhoging van: de totale KKV concentratie, de hoeveelheid *Clostridium* klas XIVa (de butyraat producenten) en de gen expressie van processen gerelateerd aan de energie verwerking in het darmepitheel. De laatste twee toonden een ook sterke correlatie.

Vervolgens werd het RNA van de microbiota bestudeerd om een beter inzicht te krijgen in welke organismen actief bijdragen aan de KKV productie. RNA wordt gebruikt voor de productie van eiwitten en geeft daarom aan welke microbiota daar actief mee waren, in tegenstelling tot DNA waar alleen uit blijkt welke microbiota aanwezig zijn. Bestaande technieken om RNA mee te analyseren zijn alleen bruikbaar voor het sequencen van soorten uit de humane microbiota. Dit komt doordat de microbiota in de darm van de muis verschilt van die in de menselijke dikke darm.

In Hoofdstuk 5 wordt de ontwikkeling van een nieuwe techniek beschreven om RNA sequenties van de microbiota van de muis wel te kunnen bepalen. Bij het sequencen van het RNA van de muis worden stukjes van 100 nucleotiden gegenereerd. Dit is voor de muis-organismen te kort om goed te kunnen zeggen bij welke functie en welk organisme dat stukje hoort. De in Hoofdstuk 5 beschreven techniek maakt gebruik van deze stukjes om langere stukken van te bouwen. Deze langere fragmenten geven meer informatie waardoor er nauwkeuriger vastgesteld kan worden bij welke functie en welke soort de fragmenten horen. De grotere lengte zorgt dat, ondanks dat de microbiota anders is, het muizen microbiota RNA toch gematched kan worden met de databases van menselijke microbiota. Deze techniek is daarna in Hoofdstuk 6 en 7 toegepast.

Om te bepalen welke organismen actief bijdragen aan de KKV productie, is in Hoofdstuk 6 en 7 het RNA van de microbiota uit het cecum (het eerste deel van de dikke darm) van muizen gesequenced. In Hoofdstuk 6 waren dit dezelfde muizen als in het experiment uit Hoofdstuk 4. In Hoofdstuk 6 zijn opnieuw de de KKV concentratie, de microbiota samenstelling en de gen expressie van het darmepitheel bestudeerd. De resultaten van de cecum resulteren in dezelfde conclusies als van

de colon in Hoofdstuk 4: de vier vezels zorgen voor een verhoging in KKV, de butyraat producenten en de gen expressie van processen gerelateerd aan de energie verwerking in de epitheel laag. Het sequencen van het microbiota RNA gaf echter een breder scala aan microbiota voor de fermentatie van vezels, dan wat uit de DNA analyse bleek. De actieve bacteriële families waren: de *Bifidobacteriaceae*, de *Lachnospiraceae* (*Clostridium* klas XIVa), de *Clostridiaceae*, de *Bacteroidaceae*, de *Erysipelotrichaceae* en *Ruminococcaceae*. Van al deze families zijn sequenties terug gevonden die betrokken zijn bij de vorming van KKV. Een deel van deze families lijkt belangrijk te zijn bij de vezelafbraak, terwijl andere families alleen gebruik maken de producten (KKV of losse suikers) van de al afgebroken vezels.
Naast dat fermentatie van vezels kan leiden tot de productie van KKV, kunnen eiwitten ook door de microbiota worden gefermenteerd tot KKV. In Hoofdstuk 7 is bestudeerd hoe melkeiwit door microbiota wordt afgebroken. Hierbij werd bevestigd dat inderdaad een verhoogd gehalte aan melkeiwit in het dieet leidt tot een verhoogde KKV concentratie in de cecum van muizen. De microbiota samenstelling verandert hierbij ook. Uit de RNA sequenties hiervan bleek dat twee families, *Erysipelotrichaceae* en *Clostridiaceae,* belangrijk zijn in de eiwit afbraak en waarschijnlijk dus ook de KKV productie.
Concluderend toont dit proefschrift aan dat door middel van veranderingen in het dieet de microbiota en het epitheel in de darmwand beïnvloed worden. Deze informatie vergroot de kennis over microbiota, diëten en de reactie van de gastheer hierop. Ook kunnen door de verkregen data uit dit onderzoek, beschreven in dit proefschrift, toekomstige studies naar de humane darm microbiota verder optimaliseren.

Curriculum vitae

Floor Hugenholtz was born on March 8th 1986 in Groningen, the Netherlands. After completing high school in 2004, she started her BSc studies in Biology at the University of Groningen. She continued with the MSc molecular biology and biotechnology, where she worked on several projects: 'Heterogeneity of Bacillus subtilis' at Molecular Genetics, RUG; 'iGEM, international Genetically Engineered Machine competition', RUG and 'Engineering lantibiotics', University College Cork. She obtained her MSc degree in 2009. In 2010, she started her PhD at the Laboratory of Microbiology in Wageningen under the supervision of Prof. Hauke Smidt and Prof. Michiel Kleerebezem. The PhD thesis entitled 'Mouse gut microbiomics of short chain fatty acid metabolism and mucosal responses' will be defended in January 2015. In May 2014 she started a Postdoc under supervision of Prof. Willem de Vos on 'synthetic gut communities' and a TIFN project on 'gender and age related changes of the mouse gut microbiota'.

List of publications

Hugenholtz F*, Lange K*, Davids M, Schaap P, Müller M, Hooiveld GJEJ, Kleerebezem M, Smidt H. Linking the fate of dietary fibres in the murine caecum to microbial transcriptome patterns. In preparation.

Hugenholtz F*, Davids M*, Schwarz J, Müller M, Tomé D, Schaap P, Hooiveld GJEJ, Smidt H, Kleerebezem M. Fermentation of dietary milk protein by murine gut microbiota. In preparation.

Lange K*, **Hugenholtz F***, Schols H, Kleerebezem M, Smidt H, Müller M, Hooiveld GJEJ. Distinct responses of mucosal transcriptional profiles and luminal microbiota composition and SCFA concentrations to dietary fibers in murine colon. In submission.

Davids M*, **Hugenholtz F***, Martins dos Santos V, Smidt H, Kleerebezem M, Schaap PJ. Functional profiling of unfamiliar microbial communities using a validated *de novo* assembly metatranscriptome pipeline. In submission.

Oberbach A, Haange SB, Schlichting N, Heinrich M, Lehmann S, **Hugenholtz F**, Kullnick Y, Schmidt H, Till H, Seifert J, Jehmlich N, von Bergen M. *In vivo* labelling of the mucosal microbiome reveals effects of a high fat diet at the metabolic synapse between microbiome and host. In preparation.

Sovran B, Lu P, Loonen LMP, **Hugenholtz F**, Belzer C, Kranenbarg EH, Boekschoten M, van Baarlen P, Kleerebezem M, de Vos P, Wells JM, Renes IB, Dekker J. Minor changes in Muc2 production leads to the development of inflammation in mice. In preparation.

Zhang J, Hornung B, **Hugenholtz F**, Ramiro Garcia J, Souza da Silva C, Bosch G, Schaap PJ, Smidt H. Metatranscriptomic analysis of the microbial composition and function in pig gut. In preparation.

Kiilerich P, Secher Myrmel L, Fjære E, Hao Q, **Hugenholtz F**, Brask Sonne S, Derrien M, Møller Pedersen L, Koefoed Pedersen R, Mortensen A, Rask Licht T, Unni Rømer M, Vogel U, Waagbø LJ, Feng Q, Xiao L, Liu C, Liaset B, Kleerebezem M, Wang J, Madsen L, Kristiansen K. Effect of a long-term high-protein diet on obesity development and gut microbiota in mice. In preparation.

Versluis D, Leimena MM, Ramiro Garcia J, D'Andrea MM, **Hugenholtz F**, Zhang J, Öztürk B, Nylund L, Sipkema D, van Schaik W, de Vos WM, Kleerebezem M, Smidt H, van Passel MWJ. Mining microbial metatranscriptomes for expression of antibiotic resistance genes under natural conditions. In submission

Sovran B, Loonen LMP, Lu P, **Hugenholtz F**, Belzer C, Kranenbarg EH, Boekschoten M, van Baarlen P, Kleerebezem M, de Vos P, Dekker J, Renes IB, Wells JM. The IL-22-1 STAT3 pathway plays a key role in the maintenance of ileal homeostasis in mice lacking secreted mucus. Accepted in IBD.

Hugenholtz F*, Zhang J*, O'Toole PW, Smidt H. Studying the mammalian intestinal microbiome using animal models. Accepted for publication in Manual of Environmental Microbiology, 4th Edition, ASM Press.

Lu Y, Fu L, Lu YH, **Hugenholtz F**, Ma F. (2014) Structure, activity and dynamics of the methanogenic archaeal community during rice straw decomposition under different temperatures in a rice field soil of the Sanjiang Plain. Accepted for publication in Soil Biology & Biochemistry.

Hugenholtz F, Mullaney JA, Kleerebezem M, Smidt H, Rosendale DI (2013) Modulation of the microbial fermentation in the gut by fermentable carbohydrates. Bioactive Carbohydrates and Dietary Fibre.

Haange SB, Oberbach A, Schlichting N, **Hugenholtz F**, Smidt H, von Bergen M, Till H, Seifert J (2012) Metaproteome analysis and molecular genetics of rat intestinal microbiota reveals section and localization resolved species distribution and enzymatic functionalities. J Proteome Res.

iGEM:

Ryback BM, Odoni DI, van Heck RG, van Nuland Y, Hesselman MC, Martins Dos Santos VA, van Passel MW, **Hugenholtz F**. (2013) Design and analysis of a tunable synchronized oscillator. J Biol Eng.

Hesselman MC, Odoni DI, Ryback BM, de Groot S, van Heck RG, Keijsers J, Kolkman P, Nieuwenhuijse D, van Nuland YM, Sebus E, Spee R, de Vries H, Wapenaar MT, Ingham CJ, Schroën K, Martins dos Santos VA, Spaans SK, **Hugenholtz F**, van Passel MW (2012) A multi-platform flow device for microbial (co-) cultivation and microscopic analysis. PLoS One.

Hesselman MC, Koehorst JJ, Slijkhuis T, Odoni DI, **Hugenholtz F**, van Passel MW (2012) The Constructor: a web application optimizing cloning strategies based on modules from the registry of standard biological parts. J Biol Eng.

Master:

Field D, Begley M, O'Connor PM, Daly KM, **Hugenholtz F**, Cotter PD, Hill C, Ross RP (2012) Bioengineered nisin A derivatives with enhanced activity against both Gram positive and Gram negative pathogens. PLoS One.

Deegan LH, Suda S, Lawton EM, Draper LA, **Hugenholtz F**, Peschel A, Hill C, Cotter PD, Ross RP (2010) Manipulation of charged residues within the two-peptide lantibiotic lacticin 3147. Microb Biotechnol.

*Contributed equally

Co-author affiliations

Jing Zhang[1]
Katja Lange[2,3]
Mark Davids[3,4]
Paul W. O'Toole[5]
Jane A. Mullaney[6,7,8]
Douglas I. Rosendale[6]
Jessica Schwarz[2]
Henk Schols[9,10]
Vitor Martins dos Santos[3,4]
Daniel Tomé[11]
Michael Müller[2,3,£]
Peter J. Schaap[3,4]
Guido JEJ Hooiveld[2,3]
Michiel Kleerebezem[3,10,12,13]
Hauke Smidt[1,3,10]

[1]Laboratory of Microbiology, Wageningen University, Wageningen, the Netherlands
[2] Nutrition, Metabolism and Genomics group, Division of Human Nutrition, Wageningen University, Wageningen, the Netherlands
[3] Netherlands Consortium for Systems Biology, University of Amsterdam, Amsterdam, the Netherlands.
[4]Laboratory of Systems and Synthetic Biology, Wageningen University, Wageningen, the Netherlands
[5]School of Microbiology & Alimentary Pharmabiotic Centre, University College Cork, Ireland
[6] Food and Nutrition Group, Food Innovation, The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand
[7] AgResearch, Animal Nutrition & Health, Grasslands Research Centre, Palmerston North, New Zealand
[8] Riddet Institute, Massey University, Palmerston North, New Zealand
[9]Laboratory of Food Chemistry, Wageningen University, Wageningen, the Netherlands
[10]TI Food and Nutrition, Wageningen, the Netherlands.
[11] AgroParisTech, INRA, UMR914 Nutrition Physiology and Ingestive Behavior, Paris, France
[12]Host-Microbe Interactomics, Wageningen University, Wageningen, the Netherlands
[13]NIZO Food Research B.V, Ede, the Netherlands
[£]present address: Norwich Medical School, University of East Anglia, Norwich Research Park, United Kingdom

Overview of completed training activities

Discipline specific activities

courses

System biology course Statistics of ~omics data analysis	2010
Genetics and physiology of food-associated micro-organisms	2010
Modelling of Biological Processes	2011

Conferences

Gut day 2010	2010
NCSB conference (**oral presentation**)	2010
System biology FEBS Innsbruck (**poster presentation**)	2011
LAB10 (**poster presentation**)	2011
Gut day 2011 (**poster presentation;** 1st prize)	2011
NCSB conference (**poster presentation**)	2011
5th International Dietary Fibre Conference 2012 (**oral presentation**)	2012
ISME14 (**oral presentation**)	2012
Gut Microbiota in Health and Disease 3rd International Workshop (**poster presentation**)	2012
Gut day 2013 (**poster presentation**)	2013
NCSB conference (**poster presentation**)	2013
Rowett-INRA Symposium (**oral presentation**)	2014
LAB11 (**poster presentation**; 1st prize)	

General courses

VLAG PhD week	2010
Competence assesment	2011
Science, the press and the general public; communication and interaction	2011
Scientific writing	2011
Writing grant proposals	2013

Optional activities	
Preparing PhD research proposal	2010
PhD/postdoc meetings	2010-2014
Lab meetings	2010-2014
TIFN meetings	2010-2014
NCSB meetings	2010-2014
PhD excursions; PhD trip 2011 & PhD trip 2013	2011 & 2013
Organisation of PHD-trip 2013	2012-2013

Dankwoord

Most people read this section of a thesis first, and of course I left it for the last moment, precisely one day before I have to print this... So I'd like to thank here everyone who has helped me during my PhD and with finishing my PhD and thesis. Pff, so that is written down and this way no one will be forgotten in my hurry.
There are of course some people I will thank here more personally, sorry for the switches between languages, but that's what I have been doing the last 4 years.

Hauke en Michiel: om te beginnen, bedankt dat jullie mij hebben aangenomen voor deze AIO positie. Ik vond het een enorm leuke baan, en dat heb ik toch met name aan jullie te danken. Ook wil ik jullie bedanken voor jullie tijd tijdens meetings, vooral tijdens de eindeloze discussies die we hebben gehad hoe we verder met het project moesten. Ik ben jullie ook erg dankbaar dat jullie me veel andere projecten hebben laten doen, ondanks dat mijn eigen werk daar af en toe wat minder van opschoot.

My office friends: Mauricio, Jing and Yue, thanks for all the fun conversations and dinners we had. Mauricio and Jing, it has been a great experience going through the same stages in our PhD careers. Starting with plenty of paperwork, halfway where we didn't know where our projects were going and stressing out in the last year to finish everything in time. Yue, even though you started later, you were a great addition to our team (not to forget your cooking skills for our dinners)! I'm really happy that you can be my paranymph.

The moleco group: Thanks everyone for all the help, collaborations and fun in the last 4 years! I would like to mention some of you specifically. Coline, thanks for finishing your PhD together with me, it was fun to do this together and I hope that all our plans for January will work out! Hans, Ineke, and Philippe thanks for all the help in the lab. Milkha, for your help in showing me around for the RNAseq protocols. Janneke, Maria, Susana, and Tom thanks for chatting with me when you or I needed it and all the tips and tricks to get things done.

Everyone within the TIFN-NCSB project; Thanks for your collaboration and all the fun (although some long) meetings. Aat, bedankt voor het organiseren van de meetings en de herinnering die je ons op tijd gaf dat er een meeting aan kwam en er presentaties nodig waren. Mark, bedankt voor de samenwerking en jouw creativiteit met de RNAseq data zodat we toch genoeg met de data hebben kunnen doen binnen onze projecten. Katja, it has been great working together with you, with both our somewhat chaotic, too relaxed way, we might not have been a good match for collaboration, but somehow it worked out fine. I'm really grateful that you will be my paranymph and am happy to return the favour in a couple of months.

Everyone else in the Microbiology lab, it was great working here and hope it will stay

such a great atmosphere for the coming two year as well. The co-organizers of the 2013 PhD trip, it was great organizing this together!

Mijn vrienden buiten het werk (ik ga jullie niet pp noemen, maar voel je vooral aangesproken!) bedankt voor alle etentjes en andere activiteiten in de afgelopen 4 jaar, het waren positieve afleidingen!

Mijn familie wil ik bedanken voor alle steun en interesse die jullie hebben gehad tijdens mijn AIO werk. Vooral ook dat we het er vaak niet over hebben, waardoor ik bij jullie er goed 'uit' ben. Tommer, fijn dat je mee naar Wageningen wilde verhuizen en bedankt voor de hulp maar vooral ook de afleiding van de afgelopen maanden.

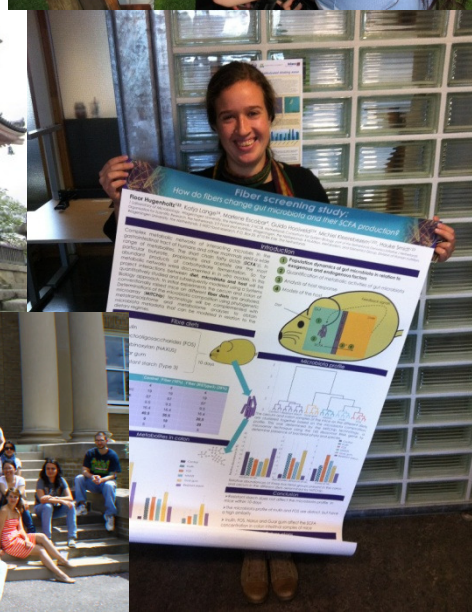
Floor











Fiber screening study
How do fibers change gut microbiota and their SCFA production?


PhD trip 2013!!!

This work was co-financed by TI Food and Nutrition and the Netherlands Consortium for Systems Biology, which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

Thesis design and layout: Floor Hugenholtz
Printed by: Gildeprint Drukkerijen – the Netherlands