From Department of Microbiology, Tumor and Cell biology
Karolinska Institutet, Stockholm, Sweden

MOLECULAR CHARACTERIZATION OF THE DYNAMICS AND DEVELOPMENT OF THE HUMAN MICROBIOTA

Hedvig Engström Jakobsson

Stockholm 2011
We are not alone.
ABSTRACT

The human body contains an enormous amount of bacteria, which are important in processes such as colonization resistance, digestion of food particles, and in the development of the immune system. Most host-microbe interactions are not harmful. Each individual harbors a unique and site-specific microbiota which is considered to be relatively stable within an individual over time. For many years, the indigenous microbiota was characterized using culture-based methods. The use of culture-based methods is quite time consuming and not completely comprehensive: a large fraction of the bacteria are not detected because of unknown growth conditions. Through the use of high-throughput sequencing technologies, based on analysis of the 16S rRNA gene, knowledge about the microbiota residing within human has increased.

In this thesis the 454 pyrosequencing technology was used in order to characterize the human indigenous microbiota in infants and adults. By developing primers specific for a certain region of the 16S rRNA gene and by the addition of a 4-5 nucleotide long barcode to each primer the 454 pyrosequencing was developed to fit multiple samples in a single run.

When the microbiota was analyzed in infants during the first two years of life, it was observed that mode of delivery had an impact on the early microbiota composition as well as chemokine levels. The diversity within the Bacteroidetes phylum was higher in vaginal delivered infants through the first six months of life and the major genus *Bacteroides* was detected significantly more frequently in the vaginally delivered infants. Moreover, a significant association between the presence of the genus *Bacteroides* in the infants’ stool at one and three months and high levels of the Th1-associated chemokines CXCL10 and CXCL11 was found. This thesis provide evidence that the Bacteroidetes are transferred from mother to child during vaginal delivery and that impaired colonization of this phylum may lead to an altered Th1/Th2 balance. The findings could help to explain the association between mode of delivery and allergy development in children. The etiology of allergy is, however, multifactorial, with many variables contributing to the final expression of atopic disease. The prevalence of allergic disease has increased markedly, especially in the Western world, and a combination of genetics and environmental factors has been proposed as a cause of this rise. Specific genera have been suggested to be part of this etiology, but with differing results. In this thesis, a specific genus was not correlated to allergic disease, but instead a low diversity of the total microbiota early in life was associated with development of allergic disease at two years of age.

In this thesis, the stability of the gastrointestinal microbiota in healthy adults and following perturbation with antibiotics was also analyzed. A relatively, but not completely stable microbiota was found in adults. Following antibiotic treatment, dramatic short-term effects were observed in throat and fecal samples. Long-term perturbations were also observed in the microbiota and also a dramatic increase and persistence in antibiotic resistance genes causing macrolide resistance was seen.
LIST OF PUBLICATIONS

This thesis is based on the following papers, which are referred to in the text according to their Roman numerals.

I. Comparative analysis of human gut microbiota by barcoded pyrosequencing.

II. Impact of delivery mode on development of the human intestinal microbiota and immune system.
    Hedvig E Jakobsson, Thomas R Abrahamsson, Maria C Jenmalm, Christopher Quince, Cecilia Jernberg, Bengt Björkstén, Lars Engstrand, and Anders F Andersson.
    *Submitted manuscript.*

III. Low diversity of the gut microbiota in infants developing atopic eczema.
    Thomas R Abrahamsson, Hedvig E Jakobsson, Anders F Andersson, Bengt Björkstén, Lars Engstrand, and Maria C Jenmalm.
    *The Journal of Allergy and Clinical Immunology* (2011) *In press.*

IV. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome.
    Hedvig E Jakobsson, Cecilia Jernberg, Anders F Andersson, Maria Sjölund-Karlsson, Janet K Jansson, Lars Engstrand.
CONTENTS

INTRODUCTION 1
  The gastrointestinal microbiota 2
  The oral microbiota 2
  The gastric microbiota 3
  The intestinal microbiota 3
  Other microbiota niches within the human being 4
  The functions of the human gut microbiota 4

AIMS OF THE THESIS 5

METHODOLOGICAL CONSIDERATIONS 6
  The 16S rRNA gene 7
  T-RFLP 9
  454 pyrosequencing 10
  Bioinformatic analyses 12

THE DEVELOPMENT OF THE HUMAN GUT MICROBIOTA 14
  Factors influencing the microbiota during infancy 15
    Mode of delivery, environment and transmission routes 16
    Gestational age 19
    Diet 19

THE ROLE OF THE GUT MICROBIOTA IN THE DEVELOPMENT OF ALLERGIC DISEASE 21
  The gut-associated immune system 21
  Allergy 22
  Neonatal and infant immune responses 24
  Gut microbiota and allergy development 25

THE STABILITY OF THE HUMAN MICROBIOTA 28
  The impact of antibiotics on the indigenous microbiota 29

CONCLUDING REMARKS 33

ACKNOWLEDGEMENTS 35

REFERENCES 37
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AMP</td>
<td>Antimicrobial peptide</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>BP</td>
<td>Base pair</td>
</tr>
<tr>
<td>CA</td>
<td>Correspondance analysis</td>
</tr>
<tr>
<td>CBMC</td>
<td>Cord Blood Mononuclear Cell</td>
</tr>
<tr>
<td>CoNS</td>
<td>Coagulase-negative staphylococci</td>
</tr>
<tr>
<td>CS</td>
<td>Caesarean section</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variance</td>
</tr>
<tr>
<td>CXCL10</td>
<td>CXC-chemokine ligand 10</td>
</tr>
<tr>
<td>CXCL11</td>
<td>CXC-chemokine ligand 11</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturating gradient gel electrophoresis</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ERM</td>
<td>Erythromycin resistance methylases</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescent in situ hybridization</td>
</tr>
<tr>
<td>GALT</td>
<td>Gut associated lymphoid tissue</td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>LAB</td>
<td>Lactic Acid Bacteria</td>
</tr>
<tr>
<td>LP</td>
<td>Lamina propria</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MAC</td>
<td>Microbiota-associated characteristics</td>
</tr>
<tr>
<td>MALT</td>
<td>Mucosal associated lymphoid tissues</td>
</tr>
<tr>
<td>OTU</td>
<td>Operational Taxonomic Unit</td>
</tr>
<tr>
<td>PCoA</td>
<td>Principal Coordinate Analysis</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PP</td>
<td>Peyer’s patches</td>
</tr>
<tr>
<td>PCA</td>
<td>Polysacharide A</td>
</tr>
<tr>
<td>RDP</td>
<td>Ribosomal Database Project</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>TGGE</td>
<td>Temperature gradient gel electrophoresis</td>
</tr>
<tr>
<td>Th</td>
<td>T helper</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor Necrosis Factor</td>
</tr>
<tr>
<td>Treg</td>
<td>Regulatory T cell</td>
</tr>
<tr>
<td>T-RFLP</td>
<td>Terminal-Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>TRF</td>
<td>Terminal restriction fragment</td>
</tr>
<tr>
<td>VD</td>
<td>Vaginal delivery</td>
</tr>
</tbody>
</table>
INTRODUCTION

For decades, the major aim within the field of microbiology was to develop antibiotics and vaccines to target pathogenic bacteria. However, most host-microbe interactions are not harmful. The human body contains a high density of microbial communities: it is estimated that the human body harbors about 100 trillion bacterial cells and that this number outnumbers the number of human cells by 10 to 1 (Savage 1977). The majority of microbes that reside within our bodies have remained largely unknown because of unknown growth requirements necessary for their cultivation. However, after introduction of molecular-based methods, especially recent high-throughput sequencing technologies, knowledge of the indigenous microbiota that resides within the human being has increased. Most studies have focused on the intestinal environment, although an increasing number of body sites are being investigated. Different molecular approaches have been applied to directly assess the diversity and composition of human-associated bacterial communities without the necessity for cultivation (Eckburg et al 2005, Suau et al 1999, Zoetendal et al 1998). By increasing our knowledge regarding the microbiota that resides at different locations in a healthy human we are able to better detect dysbiosis, which means a microbial imbalance, and relate it to disease development. A healthy gut microbiota is a metabolically active community and lives in symbiosis with its host protecting it from harmful pathogenic bacteria. However, dysbiosis of the indigenous microbiota that can occur, for example due to lack of exposure to bacteria early in life, may result in the development of disease. Indeed, several disease states have been associated with the indigenous microbiota, e.g. allergic disease (Forno et al 2008, Wang et al 2008), type 1 diabetes (Wen et al 2008), obesity (Ley et al 2005, Ley et al 2006, Turnbaugh et al 2009), stomach cancer (Dicksved et al 2009, Parsonnet et al 1991), atherosclerosis (Koren et al 2011), inflammatory bowel disease (Dicksved et al 2008, Ott et al 2004, Seksik et al 2003, Willing et al 2010), and necrotizing enterocolitis (Fell 2005).

To date the human indigenous microbiota is not fully described, although we do know that within each anatomical niche in the human being, a complex, specialized microbiota is found (Figure 1) (Costello et al 2009, Dethlefsen et al 2007). Although there are more than 50 bacterial phyla on Earth, the human microbiota is dominated by only four phyla, namely the Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria (Dethlefsen et al 2007). However, other phyla are also represented, e.g. the Chlamydiae, Cyanobacteria, Deferribacteres, Deinococcus-Thermus, Fusobacteria, Spirochaetes, Verrucomicrobia, and the candidate phyla TM7 and SR1 (Dethlefsen et al 2007). Archaea are also found in the human microbiota and the most abundant species found in the colon is the Methanobrevibacter smithii (Eckburg et al 2005, Gill et al 2006). Methanobrevibacter oralis has also been detected in the oral cavity, but only in relation to disease (Lepp et al 2004).
The gastrointestinal microbiota

The oral microbiota

The oral cavity has a large surface area with several distinct niches, each suggested to harbor a specific bacterial community according to a recent study of the healthy microbiota in different body sites (Aas et al 2005). Up to 11 different phyla have been found in the oral cavity: Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Tenericutes, Fusobacteria, Proteobacteria, Spirochaetes, Synergistetes, and the two candidate phyla SR1 and TM7 (Dewhirst et al 2010). Saliva contains antimicrobial peptides (AMPs), enzymes, and mucins that together make up a barrier towards microbes. However, a highly diverse microbiota is still found in the oral cavity (Aas et al 2005). An attempt to define a healthy oral core microbiome recently showed that a major proportion of bacterial sequences of unrelated healthy individuals was identical, supporting the concept of a core microbiome in a healthy state (Zaura et al 2009). The predominant genera were Streptococcus, Gramlicatella, Neisseria, Haemophilus,
Corynebacterium, Rothia, Actinomyces, Prevotella, Capnocytophaga, Porphyromonas, and Fusobacterium. Bik et al. found similar results regarding the most abundant genera when 10 healthy individuals were analyzed (Bik et al 2010). They also found that at least 15 bacterial genera were conserved among all 10 individuals, with significant interindividual differences at the species and strain level. Interestingly, in a recent study several phylotypes found in the oral cavity were correlated with their presence in atherosclerotic plaques, and several bacteria were associated with disease biomarkers for atherosclerosis, suggesting that bacteria from the oral cavity could contribute to the development of cardiovascular disease (Koren et al 2011). The healthy microbiota has also been characterized in the distal oesophagus (Pei et al 2004) and showed a similar bacterial composition as the one found in the oral cavity with a dominance of Streptococcus, Rothia, Prevotella, Veillonellaceae, and Granulicatella.

The gastric microbiota
Because the environment in the stomach is harsh and acidic, the stomach has been considered to harbor few bacteria with approximate 10^0 to 10^4 cells per 1 gram gastric content (Martins dos Santos et al 2010). The role of Helicobacter pylori in gastric disease has been well documented (Marteau and Chaput 2011), however the role and presence of other gastric bacteria have not been extensively studied. A few studies have determined that the gastric microbiota composition is highly diverse (Bik et al 2006, Li et al 2009). Bik et al. (2006) found that the gastric microbiota differed significantly from the oral and esophagus flora and that it harbored 128 phylotypes including H. pylori (Bik et al 2006). This indicates a distinct microbiota in the stomach (Bik et al 2006). Other genera, besides H. pylori, that have been found to dominate the stomach include Streptococcus, Prevotella, Veillonella, and Rothia (Bik et al 2006, Li et al 2009). A gastric cancer microbiota has been found to harbor low abundances of H. pylori and a dominance of other bacteria, such as Streptococcus, Lactobacillus, Veillonella, and Prevotella (Dicksved et al 2009). See more about H. pylori in the chapter “The stability of the human microbiota”.

The intestinal microbiota
The intestinal microbiota is by far the most studied and also the gut is the most densely colonized habitat. The number of bacteria differs throughout the intestine with a gradual increase in bacterial counts from the duodenum (approximate 10^4-10^5/gram content) down to the colon (approximate 10^{11}/gram of feces) (Martins dos Santos et al 2010). Most studies conducted have analyzed the fecal content because it is easier to obtain a fecal sample than a biopsy. One issue often discussed is whether the fecal microbiota is truly representative of the gut microbiota as a whole. Although it has been postulated that the fecal microbiota represents the gut microbiota (Eckburg et al 2005), this does not necessarily reveal the mucosa-associated microbiota. The adult microbiota in fecal samples has been shown to be dominated by genera within the Firmicutes and Bacteroidetes phyla. A phylogenetic core gut microbiota has also been recently suggested (Tap et al 2009). It has been estimated that the core intestinal microbiota contains between 1000-1150 prevalent bacterial species, and around 160 species that are shared between individuals (Qin et al 2010).
Other microbiota niches within the human being

The skin is also a major microbial habitat that is colonized by bacteria along with fungi and viruses. The primary role of the skin is to act as a physical barrier against foreign organisms and other substances, such as toxic compounds (Grice 2011). Historically, *Staphylococcus epidermidis* and other coagulase-negative staphylococci (CoNS) have been regarded as the primary bacterial colonizers of the skin. Other skin colonizers include genera within the phylum Actinobacteria, such as *Corynebacterium*, *Propionibacterium* and *Brevibacterium* and the genus *Micrococcus* (Grice 2011). By 16S rRNA gene analysis the skin microbiota has been shown to have a high diversity (Gao et al 2007, Grice 2011).

The healthy vaginal microbiota is dominated by different species of lactic acid bacteria (LAB), such as lactobacilli (Vasquez et al 2002, Verhelst et al 2004). However, other bacteria such as *Atopobium* sp., *Megasphaera* sp., and *Leptotrichia* sp., can reside in the vagina as well (Zhou et al 2004). An imbalance in the vaginal microbiota composition is thought to lead to a syndrome of bacterial vaginosis, and different species have been associated with this imbalance for example *Gardnerella vaginalis*, *Mobiluncus* sp., *Mycoplasma hominis*, species of *Clostridiaceae*, and *Atopobium vaginae* (our own unpublished observation) (Verstraelen et al 2004, Zhou et al 2007).

The functions of the human gut microbiota

For the host, the indigenous microbiota serves several beneficial roles. The main functions of the intestinal microbiota can be divided into *trophic*, *protective* and *metabolic* functions (Guarner 2006). The indigenous microbiota aids in the control of epithelial cell proliferation and differentiation and also in the development and homoeostasis of the immune system (*Trophic functions*). The microbiota protects against pathogens through a mechanism called colonization resistance (*Protective functions*), a mechanism that is not well developed in infants, but increases with age (Adlerberth and Wold 2009). Finally, the microbiota aids in the break-down of food particles and fermentation of non-digestible dietary residue and endogenous mucus, salvage of energy as short-chain fatty acids (SCFAs), production of vitamin K, and the absorption of ions (*Metabolic functions*) (Guarner 2006). The microbiota has also been shown to promote angiogenesis (Stappenbeck et al 2002). In exchange the microbes are provided with an environment rich in nutrients, which are necessary for their survival. In other words the human microbiota maintains a symbiotic relationship with their host.
AIMS OF THE THESIS

This thesis focuses on characterization of the dynamics of the human indigenous microbiota during infancy as well as in adults and its relation to health and disease. This was addressed by studying the development of the indigenous intestinal microbiota in infants and how early microbiota disturbances may have the potential to affect the development of allergic disease. The stability of the human microbiota in adults and what happens when this stability is disturbed by antibiotic treatment was also investigated. The specific aims of the thesis were:

To develop a method based on 16S rRNA gene pyrosequencing for monitoring of microbial communities (*Paper I*).

To monitor the development of the intestinal microbiota in infants born vaginally and through caesarean section and relate delivery mode and specific bacteria to components of the immune system (*Paper II*).

To assess the microbial diversity and characterize the intestinal bacteria during the first year of life in infants who either developed atopic eczema or did not have any allergic manifestation up to two years of age (*Paper III*).

To assess the impact of a commonly used antibiotic treatment for *Helicobacter pylori* on the bacterial community composition in throat and stool samples over a four-year time period (*Paper IV*).
METHODOLOGICAL CONSIDERATIONS

Until recently, complex environments, like the human gut microbiota, have mainly been characterized using culture-based methods. However, it is estimated that more than 80% of the bacteria residing in the gut are missed by culture-based methods (Eckburg et al 2005). Also, the need for comparison of complex microbial communities in many different samples increased the need for rapid screening approaches. Therefore, several molecular screening approaches have recently been developed for determining microbial community composition and diversity. These include different microbial community fingerprinting techniques based on characterization of 16S rRNA genes, such as terminal-restriction fragment length polymorphism (T-RFLP) (Paper IV), denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE). A fingerprinting method generates a pattern of the microbial community within a sample. These relatively low resolution fingerprinting approaches are now becoming replaced by high-throughput DNA sequencing technologies such as 454 pyrosequencing, acquired by Roche Diagnostics (Indianapolis, USA). The method will be referred to as 454 pyrosequencing throughout (Paper I-IV).

454 pyrosequencing has added more knowledge about complex microbial environments through its ability to analyze the microbial composition rapidly in many samples with greater depth of sequence coverage. The advantage with sequencing compared to the other fingerprinting methods mentioned above is the ability to classify the sequences according to taxonomy. However, depending on the scientific question raised the use of a high-throughput sequencing technology might not always be necessary. For example in order to study the stability of the dominant microbial members and to cluster the dominant bacteria in a large number of samples, T-RFLP remains a cheaper and faster alternative. Also, traditional culturing approaches still have some advantages over high-throughput sequencing. For example, in a clinical setting, we are routinely able to culture clinically important species such as Bacteroides, Staphylococcus, and Enterococcus and once cultured we can determine their phenotypes of interest, for example antibiotic resistance. We also have the possibility to type clinically relevant isolates to the clonal level and study them specifically (Jernberg et al 2007, Sjölund et al 2003). However, as sequencing techniques become more rapid and less expensive other applications could be achievable based on sequencing alone. It is possible that in the future high-throughput sequencing technology can be used as a first screening tool in samples such as large patient materials. This can then be combined with other methods capable of characterization down to species and clonal level. Sequencing of genomes and total DNA including functional genes, a so-called metagenomics sequencing approach, could also give information about virulence, metabolic functions, antibiotic resistance genes and other key functions which is of importance in better understanding of the role of members of the community as well as providing better detection tools and treatment options. However, the most important issue is to choose the most relevant method to address a specific scientific question.

In this thesis, microbial communities in human samples were analyzed using the fingerprinting techniques T-RFLP and 454 pyrosequencing, based on 16S rRNA
gene analysis. When the projects included in this thesis were started 454 pyrosequencing was just becoming available. Today there are many additional sequencing technologies available e.g. the Illumina Genome Analyzer (Bennett 2004), the SOLiD system (Metzker 2010), and Ion Torrent (Rothberg et al 2011). Those methods employ different sequencing technologies and generate shorter sequence read lengths compared to 454 pyrosequencing. When the number of sequence reads obtained from the 454 pyrosequencing, Illumina, and SOLiD were compared it was found that the fraction of junk data, which could not be aligned was largest when using the SOLiD system and that less sequence coverage was required for detection when using the 454 pyrosequencing platform (Suzuki et al 2011). In summary, these platforms are useful for many different applications, including amplicon sequencing, transcriptome sequencing and metagenomics.

The 16S rRNA gene
The 16S rRNA is a component of the 30S subunit of the ribosome in bacteria and archaea. It is a suitable gene to use as a phylogenetic target since it is evolutionarily conserved and does not transfer between organisms (Olsen et al 1986, Woese 1987). It can therefore provide a taxonomic identification tool ranging from the domain level to approximately the species level. The 16S rRNA gene is about 1600 nucleotides in length (Olsen et al 1986) and contains 9 variable regions with interspersed conserved regions (Figure 2 and Paper I) (Baker et al 2003). A limiting factor to be considered when using the 16S rRNA gene is that it is present in 1-15 copies per bacterial genome and that there might be a divergence in gene copies within a single organism (Klappenbach et al 2000). However, this can be avoided by analyzing a gene that exists in a single copy instead, e.g. the rpoB gene (Klappenbach et al 2000). Certain very diverse bacteria might be complemented with analysis of a single copy gene because of the high divergence within the organism.

Characterization of unknown organisms by rRNA gene sequences requires a collection of reference sequences of known organisms. The two largest 16S rRNA gene databases are the Ribosomal Database Project (RDP) (Cole et al 2009) at

![Figure 2. The variability within the 16S rRNA gene (Paper I).](image)
http://rdp.cme.msu.edu/ and Greengenes (DeSantis et al 2006) at http://greengenes.lbl.gov/cgi-bin/nph-index.cgi. These databases provide 16S rRNA gene sequences, alignments, analysis services and other related products. The RDP database is currently comprised of more than 1.9 million 16S rRNA gene sequences (September 2011).

Many different primer pairs covering different regions of the 16S rRNA gene have been designed and evaluated. Primer pairs specific for certain taxa have been designed as well as primers that are more universal. Since the 16S rRNA gene databases are continually expanding there is a constant need for development and re-evaluation of the primers that are used and which variable regions of the 16S rRNA gene are the most suitable to use. Because the sequence length using high-throughput sequencing is usually short and does not cover the whole 16S rRNA gene it is important to choose a region covering as many taxa as possible. There is still no consensus regarding which region to use, which makes comparisons between studies more difficult. A few regions, such as the regions around V2, V4, and V6 have been more popular to use and a few studies have assessed region suitability (Chakravorty et al 2007, Liu et al 2008, Sundquist et al 2007, Wang et al 2007). The V2 and V4 regions have been described to generate the lowest error rate when assigning taxonomy (Liu et al 2008, Wang et al 2007). Recently it was suggested to use fragments covering different regions (V4, V5+V6, and V6+V7) for richness estimations in a microbial community (Youssef et al 2009). Another study suggested that an average of the V1-V3 region and the V7-V9 region would generate results similar to Sanger sequencing, however with a much deeper coverage depth (Kumar et al 2011). In this thesis two different primer pairs were used in the 454 pyrosequencing analysis: the forward primer 784 and the reverse primer 1061 (primer pair 1), covering the V6 region (developed in Paper I and used in Papers I and IV) and the forward primer 341 and reverse primer 805 (primer pair 2) covering the V3 and V4 regions (Herlemann et al 2011) (used in Papers II-III). The primers were complemented with adapters needed for the 454 sequencing procedure and also sample-specific barcodes. In Paper I, 59 bp of the V6 region was sequenced and it was concluded in this paper that this was the most variable region. In Paper IV a longer sequence of the V6 region (about 200 bp) was analyzed. In Papers II and III, where the primer pair 2 was used, a 200 bp of the V4 region was analyzed instead. The two primer pairs were compared regarding coverage to 16S rRNA gene sequences (the sequences were downloaded from the RDP database and covered both primer regions). Primer pair 1 failed in covering several phyla including the Verrucomicrobia phylum, which is one common phylum found in human microbiota analysis. Another advantage using the second primer pair is the generation of longer sequence reads (~450 bp), so those primers could be used using the GS FLX Titanium (see below) in future analysis.

The length of the sequence needed, how many sequences necessary, and which region to use depends on the scientific question raised though (Hamady and Knight 2009). For example, it has been estimated that for defining the major phyla found in a sample, relatively few sequences are needed (Hamady and Knight 2009). Sequencing a short region such as 100 bp of the 16S rRNA gene targeting a specific bacterial community might be as good choice as covering the entire 16S rRNA gene (Liu et al 2007, Sundquist et al 2007). However, analyzing around 1000 sequences/samples
has been suggested to provide a good balance between the number of samples and the sampling depth (Hamady and Knight 2009). Of course many rare species will be missed, but this sequence number will probably cover species that are found ≥ 1% (Hamady and Knight 2009). The taxonomic assignment is influenced by the region of the 16S rRNA gene, which assignment method used and the length of the 16S rRNA gene (Liu et al 2008). Analyzing variable regions of the 16S rRNA gene might not be the best choice for community characterization, however an excellent choice for OTU diversity estimations.

A bacterial species has for long time been defined practically as a group of strains, including a type strain, that have at least 70% overall sequence identity (traditionally defined by DNA-DNA hybridization kinetics) and at least 97% in 16S rRNA gene sequence identity (Gevers et al 2005). However, the most recent recommendations for 16S rRNA gene sequence similarity is a threshold range of 98.7-99% when testing a novel isolate (Stackebrandt 2006). Within the 16S rRNA gene analysis field, the term operational taxonomic unit (OTU) or phylotype have been used to describe organisms based on their phylogenetic relationships to other organisms.

**T-RFLP**

T-RFLP is a robust and rapid molecular method that has been used to study microbial community composition and diversity in complex communities, e.g. soil and feces. T-RFLP has for example successfully been used to monitor shifts in the microbial composition following antibiotic treatment (Paper IV (Jernberg et al 2005, Jernberg et al 2007)). T-RFLP is a PCR based method where one of the 16S rRNA gene primers is fluorescently labeled. By digesting the PCR product using different restriction enzymes different terminal restriction fragments (TRFs) are generated. These fluorescently labeled fragments are separated either by polyacrylamide gel (Paper IV) or by capillary electrophoresis in an automated sequencer. By doing in silico digestion of 16S rRNA genes it is possible to get putative identifications for the microbial species corresponding to specific terminal restriction fragments. Also, by using several restriction enzymes it is easier to discriminate between different bacteria (Dunbar et al 2001). The length or peak height and area of the TRFs are calculated and by including a size standard in every lane it is possible to compare samples from different runs. From these results it is then possible to calculate the richness and evenness of dominant species of a microbial community (see more regarding richness and evenness below). The most common way to analyze T-RFLP data is by multivariate statistical analyses methods such as principal component analysis (PCA), correspondence analysis (CA) or cluster analysis. T-RFLP is a good method for comparing different communities, not necessarily for species identification. For more detailed information regarding species it is preferable to combine T-RFLP with cloning and sequencing (Dunbar et al 2001). In Paper IV T-RFLP was used in combination with 454 pyrosequencing and the results for the dominant microbial community members corresponded well when comparing the methods.
454 pyrosequencing

The 454 pyrosequencing platform was described in 2005 for whole genome sequencing (Margulies et al 2005). It then continued to be further developed and by including sequence barcodes it became possible to analyze multiple samples in a single run (Paper I (Birnaden et al 2007, Dethlefsen et al 2008, McKenna et al 2008, Meyer et al 2008, Sogin et al 2006)). The need for cloning is here eliminated because the method is employing single molecule amplification before sequencing (Margulies et al 2005) thereby increasing the speed of the process. The basics of the 454 pyrosequencing approach are illustrated in Figure 3. For amplicon sequencing, the 16S rRNA genes are first PCR amplified. The DNA templates are then mixed with beads and amplified in a water-in-oil-emulsion, and then the DNA-bound beads are placed on a fiber-optic slide. The slide is placed in the 454-pyrosequencing system and a pyrosequencing reaction (Ronaghi et al 1998) occurs in each well of the slide. Each well is read separately and in the 454 sequencing instrument, sensors are in contact with the bottom of the fiber-optic slide that detect the photons released in the pyrosequencing reaction.

Figure 3. Overview of the 454 pyrosequencing technology. (A) PCR amplification of 16S rRNA genes (B) The amplicons are mixed with beads and amplified in a water-in-oil-emulsion under conditions favoring binding of a single DNA fragment per bead. (C) The DNA-bounded beads are placed on a fiber-optic slide containing wells that are designed to fit only one bead per well. (D) Smaller beads carrying the reagents needed for the pyrosequencing reaction are added to the plate. (E) The chip is then placed in the sequencing system and a pyrosequencing reaction occurs in each well of the chip. Each well is read separately and in the 454 sequencing instrument, sensors are in contact with the bottom of the fiber-optic slide and detects the photons released in the pyrosequencing reaction. (F) Microscope photograph of emulsion showing droplets containing a bead (thin arrow) and empty droplets (thick arrow). (G) The fibre-optic slide. Adapted and were reprinted by permission from Macmillan Publishers Ltd. [Nature] 2005;437:376-380.
The 454 pyrosequencing technology has developed rapidly and is constantly upgraded with the possibility to generate more sequences in each upgraded version. In 2005 the first pyrosequencing machine, the GS20, was released with the possibility to read 100 bp long sequences. This was followed by the GS FLX instrument, which can read sequences that are 200-400 bp long. In 2008 the GS FLX Titanium was launched with the ability to generate 400-600 million bp per run with 400-500 base pair read lengths. In the present thesis, the 454 pyrosequencing technology was developed for amplicon sequencing of several samples in a single run using the GS20 sequencing machine (Paper I). 16S rRNA gene primers were designed with the addition of 4-5 nucleotide long sequence tag at the 5' end of one of the primers (Figure 4). This made it possible to trace back to the original sample and thereby analyze multiple samples in a single run. In Paper II-III the GS FLX instrument was used for amplicon sequencing.

Figure 4. The strategy of using sample specific sequence tags in the 454-pyrosequencing. The PCR primers are indicated as black arrows in the amplification of the 16S rRNA gene (PCR1). In the emulsion-PCR (PCR2) the primer pair is indicated as grey and blue primers. The primer indicated by the blue arrow is also used in the sequencing reaction. C = conserved 16S rRNA gene region. V = variable 16S rRNA gene region. The figure was kindly provided by Anders Andersson.
**Bioinformatic analyses**

Considering the large amount of sequence data that is generated by new high-throughput sequence technologies and that most available analysis tools are not capable of handling all those sequences, the need for bioinformatics tools that can analyze all the data is necessary. Several steps are involved in the processing of sequencing data generated from a 454 pyrosequencing run. Immediately following a run, very low quality sequences need to be directly removed (Margulies et al 2005). Other basic ways to filter out noise include checking for correct barcodes and primer sequences. After that the sequences proceed through different steps as described in the following sections.

The error rate per base for 454 pyrosequencing has been shown to be similar to that for Sanger sequencing (Huse et al 2007). Noise occurs due to biases that are introduced during the PCR amplification and sequencing steps. This can include the formation of chimeras, which are misjoinings of unrelated sequences, or generation of insertions or deletions within homopolymers (see below). If the noise is not removed this leads to inflated estimates of the number of OTUs and thereby an overestimation of diversity (Kunin et al 2010, Quince et al 2009). One way to address this problem is to use specialized software tailored to remove pyrosequencing noise. One example is the software package called AmpliconNoise (Quince et al 2011), capable of separately removing 454 sequencing errors and PCR single base errors (Quince et al 2011) (used in **Paper II** and **III**).

Different aspects of diversity can provide important information about the dynamics and structure of an ecological community. The diversity in a community can be assigned using different methods and which method to use depends on what is being measured. $\alpha$-diversity measures the diversity within a sample in terms of e.g. richness or evenness, while $\beta$-diversity measures diversity between samples (differences in communities). Species richness is a measure of the number of e.g. OTUs or species and can either be observed or estimated (usually larger than the observed). In **Paper II**, the diversity index “Chao1” was used for estimated richness (Figure 5) (Chao 1984). Evenness, reports how even abundance the different components of the community have. Pielou’s evenness index was used in **Paper II** (Figure 5) (Pielou 1966). A commonly used method to use when measuring $\alpha$-diversity is the Shannon-Weiner index (Magurran 2005) (**Paper I-IV**), which is a combined richness-evenness measurement, although other approaches including the Simpson method (Magurran 2005) and rarefaction curves may also be used. When measuring $\beta$-diversity, methods such as UniFrac (Lozupone and Knight 2005) (**Paper I-IV**), Bray-Curtis (**Paper IV**), Pearson correlation coefficient, or Spearman’s rank correlation coefficient (**Paper II**) are commonly used. One limitation of diversity measurements is that there is no consensus regarding the species or OTU definition, which makes it difficult to compare data between different studies. Using different diversity measurements results in different results and it is important to carefully consider and define what each method generates.
Another aspect to consider when analyzing sequencing data is that there is no consensus regarding how taxa are identified. Different algorithms have been used for clustering OTUs, with varying results (Hamady and Knight 2009). For example, two different methods for selecting OTUs are the nearest-neighbor algorithm and the furthest-neighbor algorithm, but the two methods can generate very different results (Hamady and Knight 2009). In the nearest-neighbor algorithm a sequence is added to an OTU if it is similar (above a certain threshold) to any sequence already present in that OTU, whereas in the furthest-neighbor algorithm a sequence is added to an OTU if it is similar to all sequences within that (Hamady and Knight 2009). Using nearest-neighbor algorithm in 454 pyrosequencing analysis usually then generates fewer OTUs. This is especially important to think about when analyzing diversity. In this thesis furthest-neighbor algorithms were employed for selecting OTUs (Paper I-IV).

In Papers II and III the Bioconductor R package EdgeR (Robinson et al 2010) was used to statistically test for over or under-representation of bacterial lineages among sample groups. This was done at the phylum, class, genus, and OTU (3% dissimilarity) levels. The employed test corrects for multiple testing, and the P-values are converted to False Discovery Rate values (Q-values), so the test measures significance in terms of false discovery rate (meaning fraction of positives expected to be false positives) instead of the false positive rate. This test automatically accounts for differences in sequencing depths for the different samples. When analyzing complex microbial communities many features are analyzed in parallel and this could lead to false positive findings, therefore it has been suggested to use this type of measurement and include Q-values (Storey and Tibshirani 2003).

Figure 5. Pielou’s evenness index and Chao1 index values for all 24 infants as well as for the mothers. Statistical significance was measured using Wilcoxon signed rank test and * indicates \( P < 0.05 \), **indicates \( P < 0.01 \), *** indicates \( P < 0.001 \). Paper II.
THE DEVELOPMENT OF THE HUMAN GUT MICROBIOTA

The GI-tract of a newborn healthy infant is considered sterile (Mackie et al 1999) and during the first years of life the infant GI-tract microbiota develops towards that of an adult. The infantile gut microbiota has been well characterized using culture-based studies, which reveal that the first colonizers of the infantile gut are facultative anaerobes such as *Escherichia coli*, enterococci, streptococci, and CoNS (Benno et al 1984, Rotimi and Duerten 1981, Stark and Lee 1982, Yoshioka et al 1983). The early facultative anaerobes create a reduced environment by consuming oxygen and thereby favoring the growth of *Bifidobacterium*, *Bacteroides* and *Clostridium* sp. (Mackie et al 1999, Stark and Lee 1982). Studies based on 16S rRNA gene methodology confirm the early culture-based studies regarding early colonizers and a dominance of *Bifidobacterium* and *Bacteroides* (Paper II Figure 6A, B (Eggesbo et al 2011, Favier et al 2002, Hong et al 2010, Hopkins et al 2005, Park et al 2005, Penders et al 2006b, Vaishampayan et al 2010, Wang et al 2004), although they also reveal new information regarding bacteria that are difficult to culture, such as the genus *Ruminococcus*, and also the time of acquisition. A few studies have reported low abundances of the genus *Bifidobacterium* in the infant microbiota (Hall et al 1990, Palmer et al 2007).

![Figure 6](image.png)

Figure 6. The mean relative abundance of the most dominant genera found in the twelve vaginal delivered infants (A) and twelve caesarean section delivered infants (B) over time as well as in their mothers. Ve, Verrucomicrobia; Pr, Proteobacteria; Fi, Firmicutes; Ba, Bacteroides; Ac, Actinobacteria. Paper II.
Swedish infants have been shown to have a delayed colonization with *E. coli* today (Adlerberth et al. 2006, Nowrouzian et al. 2003). Skin staphylococci has been found in higher abundances in the infant gut earlier today, suggested to be a consequence of the reduced colonization by *E. coli* (Adlerberth et al. 2006, Eggesbo et al. 2011, Lindberg et al. 2011). Recently infants were found to be initially colonized with staphylococci and high abundances of *E. coli* in their gut microbiota similar to the classical culture studies (Eggesbo et al. 2011). In **Paper II**, unclassified *Enterobacteriaceae* was found as 1 week; however whether these OTUs correspond to *E. coli* or not remains to be defined (Figure 6A, B).

Over time the anaerobic bacteria outnumber the aerobic bacteria by approximate 1000:1. Palmer *et al.* found that during the first week to the first month of life the microbiota was very variable from individual to individual (Palmer et al. 2007). In concordance with recent studies (Eggesbo et al. 2011, Palmer et al. 2007) **Paper II** demonstrate that microbial succession differs markedly between individuals during the first year of life and that the inter-subject differences declined over time, probably as a result of differences in time of weaning and incidental exposure of bacteria from the environment. The community composition converges to an adult-like state within two years. However, even at 2-2.5 years of age the diversity is still lower than in adults, indicating that the flora is still not fully developed (Figure 7) (**Paper II** (Koenig et al. 2011)).

**Figure 7.** Increase in fecal microbiota $\alpha$-diversity over time. Distributions of Shannon diversity indices displayed for all 24 infants (VD and CS infants) at one week, one, three, six, twelve, and 24 months, and for their 24 mothers. *** indicates $P < 0.001$, and ** indicates $P < 0.05$. Fifty percent of the data points reside within boxes, 75% within whiskers, and medians are indicated by horizontal lines within boxes (circles indicate individual values). **Paper II.**

**Factors influencing the microbiota during infancy**

The origin of the different aerobic and anaerobic bacteria successively colonizing the infant is not certain, but common external factors investigated include mode of delivery, the environment including family members, and the diet. These factors will be briefly discussed in the following sections.
Mode of delivery, environment and transmission routes

One important factor reported to shape the infant microbiota is the mode of delivery. The incidence of caesarean section (CS) has increased from 5% in the 1970s to more than 60% in some hospitals in China from recent reports (Sufang et al 2007). The percentage of CS in Sweden during 2000-2010 has been reported to be 17% according to the World Health Organization. Since a dysbiosis in the microbiota has been linked to for example energy balance (Bäckhed et al 2004, Martens et al 2008, Samuel et al 2008) and maturation of the immune system (Are et al 2008, Mazmanian et al 2005), early changes in the microbiota might predispose the infant to disease later in life. An association between CS delivery and the development of atopic disease has been shown in several studies (Bager et al 2008, Salam et al 2006, Thavagnanam et al 2008).

Caesarean section delivered infants have a different gut bacterial composition early in life compared to vaginally delivered infants (VD), including less abundance of Bacteroides and bifidobacteria (Fanaro et al 2003, Penders et al 2006b) and a delayed colonization of Bacteroides, E. coli, and bifidobacteria (Adlerberth et al 2006, Adlerberth et al 2007, Bennet and Nord 1987, Grönlund et al 1999, Hall et al 1990, Hallström et al 2004, Neut et al 1987). In Paper II, the most striking difference between VD and CS infants was the impact on the Bacteroidetes phylum (Figure 8A, B). The VD infants were significantly more frequently colonized by the genus Bacteroides. Because of reduced competition other bacteria than Bacteroides are able to colonize and grow in CS delivered infants. Other such bacteria that have been shown to colonize are members of the Enterobacteriaceae such as Klebsiella, Enterobacter, and clostridia (Adlerberth et al 2006, Adlerberth et al 2007). Colonization of the genus Enterococcus has previously not been shown to depend on delivery mode, suggesting other sources in addition to the maternal intestinal microbiota. In Paper II, however, a higher relative abundance of Enterococcus was found in the CS infants at one month of age, suggesting that the reduced competition favors the growth and colonization of enterococci (Figure 6B).

Figure 8. Development of the infant fecal microbiota at the phylum level. The mean relative abundance (%) of the most dominant bacterial phyla found in the twelve VD infants (A) and twelve CS infants (B) over time, as well as in their mothers. Paper II.
The establishment of the oral microbiota also seem to differ between CS and VD infants with a higher number of taxa found in the three-month-old VD infants (Lif Holgerson et al 2011). That the environment is a potential source for infant gut bacteria is shown in studies from developing countries, usually with poor sanitary and crowded conditions, where infants are earlier colonized with Enterobacteriaceae, especially *E. coli*, lactobacilli and enterococci (Bennet et al 1991, Mata 1971, Rotimi et al 1985). Other bacteria such as *Staphylococcus aureus* and CoNS have been found in low abundances in the infant gut in developing countries (Adlerberth et al 2007, Bennet et al 1991, Lindberg et al 2000, Mata 1971). When the intestinal microbiota in Estonian and Swedish infants was compared it was found that lactobacilli and eubacteria colonize earlier in Estonian infants suggesting geographical differences in infant microbiota composition (Sepp et al 1997).

It is speculated that the earliest colonization events for an infant are to a large extent determined by the opportunistic colonization by bacteria to which a baby is exposed in its environment (Palmer et al 2007). Parallel temporal patterns between a twin pair suggest that the environment is a factor strongly contributing to the microbiota composition (Palmer et al 2007). Bacterial communities at certain body sites have been shown to be more similar between human family members than unrelated individuals (Dethlefsen et al 2007). Also, it is suggested that single children not growing up with any siblings harbor a slightly different microbiota composition (Adlerberth and Wold 2009). The human gut microbiota has co-evolved with its host, as mirrored in the relationships of gut communities of different primates (Ochman et al 2010). Dominant members of anaerobic Firmicutes and Bacteroidetes of the gut microbiota do not appear to grow outside this environment and hence need to be transmitted between human hosts (Ley et al 2006a). Also, studies on mice show that the microbiota composition is inherited from the mother and that mice that were kept within the same household had a more similar microbiota composition than mice from different households (Ley et al 2005).

Common environmental exposures are likely to include the maternal vaginal, fecal, or skin microbiota, as suggested by the observed similarity of some infants’ early stool microbiota to these samples (Paper II (Dominguez-Bello et al 2010, Palmer et al 2007, Vaishampayan et al 2010)), which is consistent with evidence of vertical transmission of microbes (Caufield et al 2007, Linz et al 2007, Mandar and Mikelsaar 1996). The best studied strain for vertical transmission is *E. coli* and it has been shown that most of the *E. coli* strains colonizing the infant gut originate from the mothers gut (Adlerberth et al 1998, Fryklund et al 1992), but other contamination sources could be the hospital staff (Bettelheim and Lennox-King 1976, Fryklund et al 1992). Infantile intestinal *S. aureus* colonization has been strongly linked to the parental skin *S. aureus* carriage (Lindberg et al 2004) and lactobacilli and group B streptococci are suggested to come from the maternal vaginal flora (Matsumiya et al 2002, Tsolia et al 2003). The overall intestinal microbiota composition has been compared in mothers and their children, however with no consensus regarding bacterial transmission (Paper II (Dominguez-Bello et al 2010, Palmer et al 2007, Turnbaugh et al 2009, Vaishampayan et al 2010)). To what extent the transmission occurs from mother to offspring is not clear though. Differences in microbiota composition depending on
delivery mode indicate a mother-child transmission. The genus *Bacteroides* has been proposed to be transmitted from the maternal gut (Adlerberth et al 2007, Vaishampayan et al 2010). A recent study based on pyrosequencing of 16S rRNA genes demonstrated that the microbiota of VD neonates (<24 hours post delivery) resembled the vaginal microbiota of their own mother and was undifferentiated across multiple body habitats (skin, oral, nasopharynx, and feces), while in CS it resembled the mother’s skin microbiota (Dominguez-Bello et al 2010). While this study provided evidence that microbiota from the birth channel is transferred from mother to child providing an inoculum for the initial microbiota, it remains to be shown to what extent specific gut-specialist members are successfully transmitted during vaginal delivery. In Paper II, specific lineages of the intestinal microbiota, as defined by 16S rRNA gene sequences, were transmitted from mother to child during vaginal delivery. For the phylum Bacteroidetes a greater overlap in 16S rRNA gene sequences between babies and their own mothers were found than between babies and other mothers for up to six months (Figure 9); this could not be shown within other phyla such as the Firmicutes and Actinobacteria. *Bifidobacterium* have been suggested to be transmitted from the mother (Tannock et al 1990) and recently a mother-to-child transmission of *B. longum* was shown supporting gut transmission (Makino et al 2011). However, in Paper II a significant overlap in the mothers’ and babies’ gut bifidobacteria rRNA sequences were not found regardless of delivery mode. Hence, *Bifidobacterium* might mainly be transmitted from the breast milk, and to a lesser extent from the intestinal microbiota, as suggested but not confirmed previously (Tannock et al 1990). The reason for this could be that different subsets of the microbial community utilize different transmission routes. It is possible that late colonizers of the infant intestinal microbiota to a higher extent are taken up from other sources such as family members and food. The lack of significant overlap in Bacteroidetes sequences between mothers and their babies after six months of age may indicate a transient colonization of the maternal strains. This corroborates a recent study that compared intestinal *Bacteroides* sequences of a single mother and her infant, where the phylotypes of the mother were present in the infant at one month but had been replaced by other phylotypes at 11 months of age (Vaishampayan et al 2010).

**Figure 9.** Mother-child overlap in Bacteroidetes sequence clusters for vaginally delivered infants. The fraction of Bacteroidetes OTUs found in the infant that are also found in the mother are displayed for all infants positive for Bacteroidetes at three (A) and six (B) months. Black bars indicate comparisons with the infant’s own mother; grey bars average of comparisons with the other mothers. **Paper II.**
Gestational age

Preterm delivery causes considerable morbidity and mortality. It is unclear what factors contribute to preterm delivery, although intra-uterine infections have been shown to play a major role (Goldenberg et al 2000) and might be caused by uncultured microbes. The microbiota in preterm infants has been investigated and has been shown to differ from healthy full-term infants (Hallström et al 2004, Millar et al 1996, Sakata et al 1985, Schwieritz et al 2003). Acute chorioamnionitis is the most common lesion reported in the placenta after spontaneous preterm birth and is thought to be caused by microbial invasion of the amniotic cavity (DiGiulio et al 2008). Neonatal necrotizing enterocolitis (NEC) causes morbidity and mortality among very low birth weight infants and a difference in bacterial composition has been proposed to cause NEC, although this has not been proven (Morowitz et al 2010).

Diet

There seems to be a minor difference in microbiota composition between breast-fed and formula-fed infants (Adlberth and Wold 2009). There are some bacteria, such as Clostridium, Streptococcus, Bacteroides and Enterobacteriaceae that have been found more prevalent in formula-fed babies (Adlberth and Wold 2009, Fallani et al 2010, Mackie et al 1999).

The human breast milk has been shown to contain several different bacteria such as the genera Staphylococcus, Streptococcus, Micrococcus, Lactobacillus, Enterococcus, Lactococcus and Bifidobacterium and it is suggested that these bacteria might be endogenously derived, not contaminated from the breast skin (Albesharat et al 2011). A vertical transfer of maternal lactic acid bacteria (LAB) from the maternal gut to the breast milk and further from the breast milk to the infant gut has been suggested (Albesharat et al 2011). The origin of these bacteria is debated however. Similar strains of enterococci, lactobacilli and S. aureus have been found in breast milk and the infant gut (Kawada et al 2003, Martin et al 2003).

The microbiota composition seems to change following weaning (Ahrne et al 2005, Fallani et al 2011, Hopkins et al 2005, Wang et al 2004). For example, the lactobacilli population has been shown to differ before and after weaning with dominant strains before weaning similar to strains found in vaginal and oral flora, and dominant strains following weaning that are commonly found in food (Ahrne et al 2005). This indicates that the diet may have an impact on the intestinal microbiota composition. This is further suggested in a study were the microbiota composition changed in infants who consumed cow’s milk that was supplemented with fish oil (Nielsen et al 2007). However, the effect of the diet on the infant microbiota composition is controversial (Balmer and Wharton 1989, Benno et al 1984, Favier et al 2002, Harmsen et al 2000, Hopkins et al 2005, Lundequist et al 1985, Martin et al 2003, Penders et al 2006b, Stark and Lee 1982, Yoshioka et al 1983). In Paper II, the difference in presence of Bacteroides between VD and CS could not be explained by differences in nutrition since most infants were breastfed until three months of age. Interestingly, a recent study (Koenig et al 2011) showed that the Bacteroidetes phylum increased dramatically following the introduction of peas and other table foods. This indicates that this phylum and perhaps other phylum as well may be boosted and increase in abundance by the diet.
Children in Burkina Faso, who consume a more fiber rich diet, have been shown to have a different bacterial gut composition, especially an enrichment of Bacteroidetes, when compared to European children, probably as a consequence of differences in diet habits (De Filippo et al 2010). Another factor linking the Bacteroidetes phylum to diet is that the genus Bacteroides has been shown also to be implicated in obesity, with both a low total bacterial diversity and depletion of Bacteroidetes (Ley et al 2006, Turnbaugh et al 2009). Lifestyle factors related to the anthroposophic way of life have been shown to influence the composition of the gut flora (Alm et al 2002, Dicksved et al 2007). For example, children raised with an anthroposophic lifestyle had a higher gut microbial diversity when compared to children raised in a farm (Dicksved et al 2007). This is probably due to a difference in diet or a low consumption of antibiotics, which are characteristics of an anthroposophic lifestyle (Alm et al 1999).
THE ROLE OF THE GUT MICROBIOTA IN THE DEVELOPMENT OF ALLERGIC DISEASE

Atopic allergy is the most common chronic disease among children in the developed world. It is estimated that more than 25% of the population in industrialized countries suffers from IgE-mediated allergies (Valenta 2002). Exposure to environmental antigens, such as dust mite and animal and food proteins, can cause allergic disease. However, it is not known why some individuals develop disease while others do not. Early life events occurring during critical windows of immune vulnerability can have a long-term impact on immune development and immune diseases such as allergy (Björkstén 1999, Jenmalm 2011, Prescott 2003). Microbial exposure is a major driving factor for the development of the immune system and immune regulatory responses (Okada et al 2010).

The gut-associated immune system

As the GI-tract harbors more bacteria than any other site of the body, the intestinal immune system is exposed to a high density of bacteria. The immune response toward mucosal encountered foreign antigens are induced in the mucosal associated lymphoid tissues (MALT), including the gut associated lymphoid tissue (GALT) (Nagler-Anderson 2001). The GALT consists of Peyer’s patches (PP) in the small intestine and lymphoid follicles in the colon (Nagler-Anderson 2001). Immune cells such as T-cells, B-cells, dendritic cells (DCs) and macrophages are present in GALT (Iweala and Nagler 2006). Different factors aid in the prevention of antigen entry through the epithelial barrier. The single layer of mucosal epithelial cells is the main site for antigen entry, but also acts as a barrier. Between the cells are tight junctions, allowing only ions to enter (Nagler-Anderson 2001). In this cell layer, different cell types are found, namely enterocytes, goblet cells, enteroendocrine cells, Microfold (M) cells and Paneth cells (Winkler et al 2007). The goblet cells are responsible for the production of mucus, creating an effective barrier (Deplancke and Gaskins 2001). In this mucus, several components, such as mucins, defensins, and secretory IgA (sIgA) antibodies, prevent bacteria from adhering (Brandtzaeg 1995, Deplancke and Gaskins 2001). The mucus layer is composed of two sub-layers; an inner, stratified mucus layer that is firmly adherent to the epithelial cells and an outer, nonattached layer (Johansson et al 2011). The inner mucus layer is dense and normally does not allow bacteria to penetrate (Johansson et al 2008). These two mucus layers are organized around the highly glycosylated MUC2 mucin, which forms a large, net-like polymer. On the MUC2 mucin, numerous O-glycans are found, and these glycans serve as nutrients for the bacteria but may also act as attachment sites. This is suggested to contribute to the selection of the species-specific colon flora (Johansson et al 2011). Low pH in the stomach, peristalsis, and the indigenous microbiota are other barriers. Antimicrobial peptides produced by the epithelial cells include enzymes, e.g. lysozyme, as well as C-type lectins, and defensins that are capable of killing bacteria. The connective tissue, the lamina propria (LP), is located beneath the epithelial cells. M cells and DCs transfer luminal antigens to cells in the lamina propria. In the next step, the DCs can present the
antigens to lymphocytes in PPs. The adaptive immune system is characterized by two major types of T lymphocytes that express either CD4 or CD8 on their surface. The CD4+ cells function generally as T helper cells (Th) and CD8+ cells function as T cytotoxic (Tc). When a Th cell is presented to an antigen, different kinds of cytokines are produced, leading to for example a Th1, Th2, Th9 or a Th17 response (Akdis and Akdis 2009). These Th cell responses are important in infection control.

Normally, ingested or inhaled antigens should not give rise to a strong inflammatory immune response, a phenomenon called oral tolerance (Weiner 2000). However, failure to induce oral tolerance may initiate hypersensitive responses to food antigens (van Wijk and Knippels 2007). The indigenous microbiota seems to be important in developing oral tolerance. For example, germ free animals, which lack microbes, have difficulties in developing oral tolerance (Sudo et al 1997). How this tolerance is achieved is, however, not clear, although it is known that regulatory T cells (Treg cells), IL-10, and TGF-β are important for down regulating responses towards autoantigens and harmless environmental antigens (Akdis and Akdis 2009). The presence of a microbiota seem to be important in the development of a Treg cells, as it has been shown that germ-free mice have been shown to have a lower Treg cell activity compared to conventionally raised mice (Östman et al 2006).

Allergy
Common allergic manifestations during the first year of life include eczema and food allergies. There are different factors that are important in the development of allergic disease, such as time of exposure, the dose, and the allergen route, as well as other environmental factors. According to the phenomenon called the “atopic march” (Spergel and Paller 2003), eczema and food allergies in children eventually disappear and are replaced at school age by allergic asthma and rhinoconjunctivitis to inhaled allergens (Wood 2003). Allergy is defined as a hypersensitivity reaction initiated by specific immunological mechanisms (Johansson et al 2004). Hypersensitivity reactions are divided into four types (Type I-IV) and they develop either in the course of humoral or cell-mediated responses (Goldsby 2003). Each type of response involves certain mechanisms, cell types and mediator molecules. Atopy is defined as a personal and/or familial tendency to become sensitized and produce IgE antibodies in response to ordinary exposures to allergens (Johansson et al 2004). Atopic allergy is a hypersensitivity reaction type I. The atopic manifestation focused on in this thesis is atopic eczema (Paper III). The incidence of infant eczema is 20-30% in affluent countries, and in a population predisposed to allergic disease this number increases to 40-50% (Bohme et al 2003, Lowe et al 2007).

When a hypersensitivity reaction type I (allergic reaction) occurs, an allergen (antigen) induces a humoral response, resulting in the generation of IgE antibody-secreting plasma cells and memory cells (Gould and Sutton 2008). When the same antigen is encountered again, the immediate allergic response is thought to occur when the allergen cross-links IgE bound to high affinity IgE receptors (FcerRI) on mast cells (Gould and Sutton 2008). A degranulation then occurs with the release of inflammatory mediators such as histamine and leukotrienes. These mediators cause itching, smooth-muscle contraction (e.g. bronchoconstriction), vascular leakage from
blood vessels, and tissue damage. This effect could be either localized or systemic, causing symptoms such as rhinitis, conjunctivitis, and asthma (Valenta 2002). Cytokines and chemokines are secreted by e.g. mast cells and the late-phase reaction is initiated (Gould and Sutton 2008). This phase is characterized by a strong T-cell infiltration and eosinophil activation hours to days afterwards and is typical in allergic individuals suffering from chronic disease (e.g. atopic dermatitis) (Valenta 2002).

It has been discussed whether allergic infants harbor a different cytokine and chemokine profile and abundance than non-atopic infants. Atopy has been associated with a Th2 deviated cytokine response to allergens, with high levels of IL-4, IL-5, IL-9, and IL-13 (Jenmalm et al 2001). It has also been associated with low or equal levels of Th1 cytokines such as IFN-γ and IL-12 (Imada et al 1995, Jenmalm et al 2001). Children that later develop disease also seem to have a delayed maturation of their immune system and a prolonged Th2-deviation during childhood (Böttcher et al 2002, Prescott et al 1999). However, the proposed Th1/Th2 deviation during allergic disease is rather simplified. Other T cells have been shown to be implicated in allergic disease as well, such as Treg cells, complicating the immune response during allergic disease.

IgE antibodies to allergens, secreted by B cells, can be detected with a skin prick test or by analyzing circulating IgE antibodies (Johansson et al 2004). Th1 and Th2 cytokines have traditionally been analyzed to assess immune responses. However, they are usually detected in very low amounts in the circulatory system and sometimes very close to detection limit. Analyses of circulating chemokines, on the other hand, have been shown to be an alternative, since they are more easily detected in peripheral blood (Abrahamsson et al 2011, Sandberg et al 2009). Chemokines are a family of around 50 small proteins, that play a major role in chemotraction of leukocytes to the inflammation site (Mantovani et al 2004, Pease and Williams 2006). The chemokines are divided into four groups and the majority of chemokines are members of the CC or CXC families. The other two families described are the C and the CX3C family (Pease and Williams 2006). Chemokines are produced by several cell types, e.g. epithelial cells and macrophages (Mantovani et al 2004). The receptors for chemokines are expressed on the surface of several cell types involved in inflammation, such as lymphocytes, monocytes, DCs, and natural killer cells (Pease and Williams 2006). In this thesis, the Th2-associated chemokines CCL17, CCL22 and the Th1-associated chemokines CXCL10 and CXCL11 were measured and associated to different birth modes and bacteria (Paper II). Different disease states have been associated with increased chemokine levels. For example, atopic dermatitis has been associated with high CCL17 and CCL22 levels (Fujisawa et al 2002, Jahnz-Rozyk et al 2005, Kakinuma et al 2002).

Genetic factors seem to be highly correlated with development of allergic disease. Complete chromosome regions and polymorphisms have been correlated to allergy development (Vercelli 2008). Furthermore, parental allergy is a strong risk factor for developing allergic disease (Bergmann et al 1997). Already in utero the fetus could be exposed to allergens that may be able to cross the placenta (Loibichler et al 2002, Salvatore et al 2005). IgE antibodies are also detected in fetal cells in the placenta and most probably originate from the mother (Holt 2008, Joerink et al 2009, Sverremark Ekström et al 2002). The incidence of allergy shows worldwide variation and has increased during the last century (Burr et al 1989, Butland et al 1997, Åberg et al 1995)
especially in industrialized countries (Bråbäck et al 1995). Even though there is a strong genetic link to allergic disease, rapid fast increase in allergic disease cannot be explained by only genetic changes, and must be considered in combination with changing exposure to environmental factors. The gut microbiota is considered an important environmental factor in this aspect (Björkstén 1999, Jenmalm 2011, Wold 1998).

**Neonatal and infant immune responses**

When an infant is born, it enters a world full of microbes, a process that requires that the immune system is able to control them. Transfer of maternal immunological memory seems to be an important factor for this control. The infant manages the first bacterial inoculums probably through a certain tolerance passed on via the mother through the induction of Treg cells (Mold et al 2008) and by passive immunization by first IgG and later IgA (Hanson and Silfverdal 2009). That there is some sort of transfer of immunological memory from the mother to the fetus is supported in several studies. Maternal antibodies, in particular IgG antibodies, are actively transferred through the placenta to the fetus and slgA is transferred through maternal milk to the neonate (Böttcher et al 2000, Hanson and Silfverdal 2009, Jenmalm and Björkstén 2000).

Allergen responses have been shown from gestational week 22, suggesting that in utero fetal exposure to an allergen may result in primary sensitization to that allergen (Jones et al 1996). This allergen responsiveness could also be due to non-specific cross reactivity, however (Thornton et al 2004). Cord blood mononuclear cells (CBMCs) from infants that later have developed allergic disease have been shown to have an altered cytokine profile following allergen stimulation (Nilsson et al 2004, van der Velden et al 2001). The levels of IgG subclasses are even higher in newborn infants with an atopic mother, as compared to babies with non-atopic mothers, supporting the idea that the infants are influenced by the maternal immune profile (Jenmalm and Björkstén 2000). Another factor supporting this idea is that the maternal, but not paternal, total IgE levels correlate with elevated infant IgE levels and infant atopy (Johnson et al 1996, Liu et al 2003, Magnusson 1988). Several studies have shown evidence of transplacental transfer of allergens (Casas and Björkstén 2001, Holloway et al 2000, Szepfalusi et al 2000). In a farm study, environmental exposures during pregnancy protected against allergic sensitization, but had less effect during infancy, suggesting that allergic manifestations may be developed during pregnancy (Ege et al 2006). Also, the neonatal immune system is suggested to be affected by the maternal microbial environment through epigenetic mechanisms (Jenmalm 2011). All this indicates that the infant already in utero may be predisposed to develop allergic disease. The question whether sensitization occurs prenatally or postnatally has been intensely debated during recent years, however (Bonnelykke et al 2008, Ege et al 2008, Rowe et al 2007).

The immune system in infants is immature and the absolute and relative numbers of the cells of the innate and adaptive immune systems; DCs, B- and T-cells, are low (Willems et al 2009). Neonatal B cells are of an immature phenotype, as evidenced by cell-surface marker characteristics and increased susceptibility to tolerance induction (Landers et al 2005, Siegrist and Aspinall 2009). CBMCs have also been shown to produce fewer Th1 associated cytokines (Yerkovich et al 2007). Th1 functions
are dampened during fetal life (Holt and Jones 2000) and it is debated whether infants are born somewhat Th2-skewed and that this predisposes to allergy development (Romagnani 2004). A Th2-skewing during infancy has been suggested in several studies (Abrahamsson et al 2011, Prabhudass et al 2011, Prescott et al 1998, Saito et al 2010), but not all (Halonen et al 2009). The majority of studies indicate that infants are born Th2-skewed, however. Also, during pregnancy, the fetomaternal interface is surrounded by high levels of Th2 cytokines probably in order to redirect the maternal immune system so it does not react against the fetus (Sandberg et al 2009). If the in utero milieu is of importance, this Th2 profile during pregnancy might stimulate the infant immune system. Appropriate microbial stimulation may be required to develop a more balanced immune phenotype, including maturation of Th1-like responses (Vuillermin et al 2009) and appropriate development of Treg cell responses (Lloyd and Hawrylowicz 2009, McLoughlin and Mills 2011). Microbial stimulation has been shown to be important for immune maturation (Jenmalm et al 2011, Marchini et al 2005). A failure of Th2-silencing during maturation of the immune system may underlie development of Th2-mediated allergic disease (Böttcher et al 2002). Furthermore, sensitization at 2 years of age was correlated to high amount of Th2 cytokines in 6 month-old infants after allergen stimulation (Rowe et al 2007). Also, an enhanced development of IFN-gamma-producing capacity during the first 3 months of life has been associated with farming, endotoxins in house dust, and cat and dog exposure (Roponen et al 2005). This may indicate that some microbial exposure early in life is capable of driving the developing immune system toward Th1 responses (Roponen et al 2005).

**Gut microbiota and allergy development**

The so-called hygiene hypothesis was described in 1989 by David Strachan that presumed that an increased number of siblings protects against allergy development (Strachan 1989), which has been confirmed in several epidemiological studies (Karmaus and Botezan 2002). According to this hypothesis, allergic disease is caused by altered microbial exposure during childhood. This changed microbial exposure may affect the maturation of the immune system. Furthermore, children growing up on a farm have a lower prevalence of atopic allergy (Kilpelainen et al 2000, Riedler et al 2001). Presumably they are exposed to microbial products such as endotoxin and bacterial DNA. Children who are frequently exposed to high levels of endotoxin are less allergic (Braun-Fahrlander et al 2002, Böttcher et al 2003, Roy et al 2003). Low exposure to endotoxin has been associated with increased risk of atopic eczema (Gehring et al 2001). Thus, exposure to these microbial products may prevent from developing allergic disease. Several studies support the hygiene hypothesis, however with inconsistent results (Penders et al 2007a).

The gut microbiota differs between allergic and non-allergic infants during the first month of life (Björkstén et al 2001, Kalliomäki et al 2001, Penders et al 2007b, Sjögren et al 2009, van Nimwegen et al 2011), and allergic disease has been correlated with higher and lower abundances of specific taxa. For example, allergic infants were colonized less often with *Bacteroides* and bifidobacteria (Björkstén et al 1999, Hong et al 2010, Sepp et al 2005, Watanabe et al 2003) and more often with *S. aureus* (Watanabe et al 2003) and with a lower ratio of bifidobacteria to clostridia (Kalliomäki et al 2001).
Colonization with *Clostridium difficile* at one month of age was shown in a recent study to be associated with an increased risk of asthma at 6 to 7 years of age (van Nimwegen et al 2011). However, there have been contradictory results in more recent studies as well. Two large European prospective studies did not confirm any relationship with allergy to any particular bacterial group (Adlerberth et al 2007, Penders et al 2006a). Also, in *Paper III*, no association with the presence of any of these bacteria previously found and atopic disease was found. However a lower diversity of the phylum Bacteroidetes and its genus *Bacteroides* was found in infants who developed atopic eczema. Low levels of these bacteria have in previous literature been reported to be associated with allergic disease (Watanabe et al 2003). However, for other genera, such as *Coprobacillus* and *Peptostreptococcus*, a higher abundance was found in the atopic versus the non-atopic infants (*Paper III*). Whether these genera actually cause allergic disease or if they merely reflect a disturbance of the intestinal microbiota is unclear.

It is, however, debated whether low diversity of the gut microbiota in infancy is more important than the prevalence of specific bacterial taxa when trying to explain why the prevalence of allergic disease is increasing in affluent countries. A low diversity of the total intestinal microbiota has in a few studies been correlated to allergic disease (*Paper III* (Bisgaard et al 2011, Forno et al 2008, Wang et al 2008)). The theory is that the gut immune system reacts to exposure to new bacterial antigens and repeated exposure would enhance the development of immune regulation (Holt 1995, Wold 1998). In *Paper III*, the bacterial phylum, Proteobacteria, appeared to be less abundant in the atopic infants. This phylum comprises Gram negative bacteria, typically with endotoxin (LPS) incorporated in their cell walls. Endotoxin elicits a Th1 response via the innate immune system by enhancing IL-12 production from monocytes and DCs (Doreswamy and Peden 2010). Low endotoxin levels were associated with an increased risk of atopic disease as described above.

Differences in the postnatal microbial colonization may explain the association between CS delivery and the development of allergic disease (Salam et al 2006, Thavagnanam et al 2008). A recent study showed that mode of delivery and birthplace had an impact on the development of atopic disease (van Nimwegen et al 2011). Children who were born vaginally at home and had at least one atopic parent had lower odds to develop asthma and sensitization to food allergens compared to children who were born vaginally at the hospital (van Nimwegen et al 2011). It has also been found that CS delivered eczema infants had significantly higher abundance of especially *Enterobacteriaceae* compared to CS delivered non-eczema infants (Hong et al 2010).

The influence of CS on immune development is largely unknown (Huurre et al 2008). *Paper II* provides evidence that CS delivery may impact the immune system. A significant association between the presence of the genus *Bacteroides* in the infants’ stool at one and three months of age and high levels of the Th1-associated chemokines CXCL10 and CXCL11 was found in *Paper II*. The Th1-chemokines CXCL10, and CXCL11 were analyzed with an in-house multiplexed Luminex assay (*Paper II*) (Abrahamsson et al 2011, de Jager et al 2005). There are unfortunately no appropriate chemokines associated with Treg cell responses, although this would have been interesting to analyze as well. Although the effect of *Bacteroides* on the chemokine levels seen in *Paper II* is hard to isolate from the effect of birth mode, which could
potentially involve differences in other genera, the observation that *Bacteroides* abundance at three months was directly associated with CXCL10 levels indicates that this genus may play an important role in the development of immune regulation (Figure 10).

As discussed above, appropriate microbial stimulation during infancy is required for the development of a more balanced immune phenotype, including maturation of Th1-like responses and appropriate development of regulatory T cell responses (Lloyd and Hawrylowicz 2009, McLoughlin and Mills 2011, Vuillermin et al 2009). The presence of *Bacteroides* may have an effect on the immune maturation and the lack of certain bacteria may prolong the possible Th2-deviation during infancy. *Bacteroides* species have in other studies been demonstrated to have an impact on the immune system via anti-inflammatory properties. Thus, *B. fragilis* prevented the induction of colitis via suppression of the pro-inflammatory cytokines TNF and IL-23 in an experimental colitis model (Mazmanian et al 2008) and also mediated a conversion from CD4+ T cells into IL-10 producing Foxp3 T regulatory cells during commensal colonization, eliciting mucosal tolerance in another mouse model (Round and Mazmanian 2010). Furthermore, *B. thetaiotaomicron* modulates the expression of a large quantity of genes involved in mucosal barrier reinforcement (Freitas et al 2005, Hooper et al 2001). As suggested, both bacteria and their eukaryotic host benefit from one another. This suggests that co-evolution has selected mechanisms promoting associations between bacteria and eukaryotic hosts. In a healthy individual, intestinal colonization of certain bacteria stimulates host production of antimicrobial peptides and secretory IgA and in turn protects the host from systemic translocation. For example, *B. fragilis* exerts strong effects on the immune system, mediated by the capsular polysaccharide (PSA), influencing T-cell mediated immune response and the Th1/Th2 balance (Mazmanian et al 2005, Mazmanian et al 2008).

![Figure 10](image)

**Figure 10.** Linear regressions of log (CXCL10) (mean of 6, 12 and 24 months) against log (relative *Bacteroides* abundance at three months) and birth mode. Circles represent samples for which CXCL10 was measured in at least one time point. *P*-values for the associations between *Bacteroides* abundance and birth mode with CXCL10 levels are 0.026 and 0.079, respectively.

**Paper II.**
THE STABILITY OF THE HUMAN MICROBIOTA

The human adult microbiota has been shown to host-specific (Dicksved et al 2007, Jernberg et al 2007) and is considered to be relatively stable within one individual and over time (Costello et al 2009, Jernberg et al 2007, Matsuki et al 2004, Savage 1977, Vanhoutte et al 2004, Zoetendal et al 1998). Among individuals, a high variability in different body sites, such as the gut, skin and oral cavity, has been found (Costello et al 2009). The stable microbiota may, however, be disturbed and altered by several factors such as diet, drug intake, disease and aging. Also, host genotype has been suggested to determine the microbiota composition (Stewart et al 2005, Turnbaugh et al 2009, Zoetendal et al 1998). In Paper IV it was shown that the microbial communities in throat and fecal were more similar within individuals than between individuals, at least up to the one year sampling period. This unique individual community composition was particularly evident for the fecal samples where larger differences were observed between individuals. There was also a pronounced temporal change in fecal community structure and after four years the gut communities had diverged in both the control group (Figure 11) and the treated group. The majority of individuals analyzed in Paper II were above 70 years old; it is possible that the microbiota composition gradually changes with age and that the microbial diversity is higher in elderly individuals, but few studies have examined this issue. One study recently showed that subjects over 100 years of age seem to have a slightly different microbiota composition (Biagi et al 2010).

![Figure 11](image-url)

**Figure 11.** Correlation plots showing OTU frequency at day 0 (x-axis), and day 8–13, 1 and 4 years (y-axis) in throat (A) and fecal (B) samples in the controls (A, B, and C). Bray-Curtis values are indicated as numbers in the figure as a number. A Bray Curtis value of 0 suggest the two sites have the same composition and 1 means the two sites do not share any species. The color of the dots represent different phyla: yellow, Actinobacteria; green, Bacteroidetes; blue, Firmicutes; red, Proteobacteria; grey, other phyla. Percentages of inter-sample variation explained by the two axes are shown in the figures. Paper IV.
No specific grouping with respect to geographical origin and gender and the microbiota in adults have been revealed, however a trend towards differences between countries are suggested (Lay et al 2005, Mueller et al 2006). However, differences in the microbiota composition seen in infants from different geographical regions, such as Estonia and Sweden suggests that geographical differences in microbiota composition exist (Sepp et al 1997). Also, a more frequent gut colonization with *S. aureus* in Swedish infants suggests that lifestyle could lead to differences in microbiota composition (Lindberg et al 2011). For example, an anthroposophic lifestyle was correlated with a higher fecal microbial diversity (Dicksved et al 2007). Even though specific factors could not be pin pointed in that study, one could speculate that the low intake of antibiotics, that are part of an anthroposophic lifestyle, might have an impact on the microbiota composition.

In **Paper IV** a highly diverse microbiota was found in the throat with 152 different phylotypes. In addition, our results from **Paper IV** indicate that the throat microbiota was more similar between individuals and more stable after antibiotic treatment and over long periods than the intestinal microbiota, suggesting that the throat environment is more selective for a specific host microbiota. That the same microbial habitat from different individuals shows great variability has been proposed in other studies as well (Dethlefsen et al 2007, Palmer et al 2007). An even higher diversity was found within the stomach with 262 phylotypes, however it could not be defined to what extent this represented resident or transient microbes (**Paper I**). The adult intestinal microbiota was characterized in **Paper I, II and IV**. In **Paper I** and IV the predominant phylum found in feces was Firmicutes, interestingly followed by Actinobacteria, Bacteroidetes, and Proteobacteria. The low abundance of the Bacteroidetes phylum found in our studies could be a result of strong inter-subject variability, which previously has been shown for this phylum (Eckburg et al 2005). The most dominant taxonomic groups found in feces were *Lachnospiraceae Incertae Sedis*, unclassified *Lachnospiraceae*, *Bifidobacterium*, *Collinsella*, and unclassified *Ruminococcaceae*.

The differences in microbiota composition seen between individuals may be partly explained by differences in diet. That the diet has a large influence on differences in microbiota composition between individuals has already been demonstrated in infants whose microbial composition changed following weaning. The microbiota has also been linked to obesity in mice and humans (Ley et al 2005, Ley et al 2006). Obese individuals that were restricted to either a fat- or carbohydrate diet demonstrated a shift in their microbiota to resemble that found in lean controls (Ley et al 2006). In a recent study, specific factors in the diet were correlated to changes in abundance of specific species in the gut microbiota ( Faith et al 2011).

Probably, the factor that probably has the strongest impact on the indigenous microbiota composition is antibiotics. This was studied in **Paper IV** and will be reviewed in more detail in the following section.

**The impact of antibiotics on the indigenous microbiota**

One concern with the administration of antibiotics is the possibility of selection of antibiotic resistant strains of bacteria; not only in those which the antibiotic is directed towards but also among the indigenous microbiota (Sjölund et al 2003). Depending on
which antibiotics are used, since different antibiotics have different modes of action, the drug volume, the route of administration, and the degree of resistance in the community, the use of antibiotic could have different impacts on the microbial community (Jernberg et al 2010). The effect of short-term administration has been explored in several studies using culture-based methods (Sullivan et al 2001) and also more recently using molecular-based methods (Paper IV (Dethlefsen et al 2008, Jernberg et al 2005)). The use of antibiotics could lead to short-term disturbances with decreased colonization resistance (Vollaard and Clasener 1994) and sometimes the development of antibiotic-associated diarrhea (AAD) (McFarland 1998). The decreased colonization resistance seen could lead to growth of potentially pathogenic bacteria already present in the gut, such as C. difficile (Sullivan et al 2001).

*Helicobacter pylori* is a Gram negative rod colonizing the gastric mucosa in half of the world’s population (Torres et al 2000). Infection with *H. pylori* is usually asymptomatic, however few individuals develop gastric disease such as peptic ulcer disease and gastric cancer (Suerbaum and Michetti 2002). The treatment for *H. pylori* usually consists of two antibiotics in combination with a proton pump inhibitor. In this thesis (Paper IV) the throat and fecal microbiota was analyzed from patients that had received a commonly used treatment regimen consisting of clarithromycin and metronidazole in combination with omeprazole, for *H. pylori* (de Boer and Tytgat 2000, Dunn et al 1997, Graham 2000). Treatment with these antibiotics can lead to antibiotic resistance development among *H. pylori* strains (Dunn et al 1997). However, this treatment regimen has also been shown to lead to resistance among members of the indigenous microbiota (Adamsson et al 1999, Jakobsson et al 2007, Jönsson et al 2005, Sjölund et al 2003, Sjölund et al 2005). One mechanism for macrolide (e.g. clarithromycin) resistance is via erythromycin resistance methylases encoded by *erm* genes. These genes have been found in different genera with *erm*(B) having the largest host range (Roberts 2008). The *erm*(B) gene is normally found on transposons located in the chromosome or on plasmids (Roberts 2008) and encodes a ribosomal methylase that methylates the 23S ribosomal RNA and thereby prevents the antibiotic from binding (Portillo et al 2000).

By using culture-based approaches the antibiotics studied in this thesis (clarithromycin and metronidazole) have been shown to result in short-term ecological effects on the indigenous gut microbiota (Adamsson et al 1999, Buhling et al 2001, Tanaka et al 2005). Adamsson et al. found that the numbers of culturable *Bifidobacterium*, *Clostridium*, and *Bacteroides* spp. significantly decreased in feces after treatment, while the numbers of enterococci significantly increased one week after treatment (Adamsson et al 1999). They also detected a persistent decrease of *Bifidobacterium* sp. and *Bacteroides* sp. 4 weeks after treatment. In Paper IV the antibiotic treatment impacted the indigenous microbiota differently in the individual treated subjects, probably due to the known unique bacterial community compositions in different individuals (Costello et al 2009, Donskey et al 2003, Eckburg et al 2005, Tannock et al 2000, Zoetendal et al 1998). However, there are some differences in the target bacteria for the two antibiotics. Metronidazole is known to be active against anaerobic bacteria and both antibiotics are active against the genus *Streptococcus*. It is also important to highlight that macrolide excretion is primarily accomplished through
bile, approximately 20–30% is excreted through the urine and the rest is excreted through feces. For metronidazole 10–15% is excreted into feces. At the phylum level it was found that Actinobacteria were strongly reduced, especially in feces, following antibiotic treatment, probably because clarithromycin is known to target this group (Williams et al 1992), while they are naturally resistant to metronidazole (Paper IV) (http://www.srga.org/). Increasing evidence has linked some members of the Actinobacteria to human health (Kassinen et al 2007, Moore and Moore 1995). For example, an increased abundance of Actinobacteria has recently been associated with obesity (Turnbaugh et al 2009). The antibiotic treatment affected some genera negatively, for example Ac tinomyces and Prevotella in throat samples and Bif idobacterium and Collinsella in fecal samples (Paper IV). Other genera were positively affected by the antibiotics, for example the genus Neisseria in throat samples and the genus Enterococcus in fecal samples. The effects of treatment that were observed on the bacterial communities in the throat and feces might be due to either clarithromycin or metronidazole or both in combination.

In Paper IV as well as in other recent studies it has been shown that some antibiotics have generated long-term impacts on the microbiota (Dethlefsen et al 2008, Jakobsson et al 2007, Jernberg et al 2007, Jönsson et al 2005, Lindgren et al 2009, Löfmark et al 2006, Nyberg et al 2007, Sjölund et al 2003, Sjölund et al 2005). Long-term impacts (six months) were seen in the gut microbiota after treatment with ciprofloxacin (Dethlefsen et al 2008). In this study, some taxa failed to recover to pretreatment levels for periods up to 6 months although the majority of the gut microbiota returned to pretreatment levels after four weeks. In Paper IV it was found that although the diversity of the microbiota subsequently recovered to resemble the pretreatment states, the microbiota remained perturbed in some cases for up to four years post treatment. A quantitative real-time PCR approach was employed to measure the abundance of the ermB gene in the samples over time (Paper IV). This approach has successfully been used in a previous study (Jernberg et al 2007). For relative quantification we used the 16S rRNA gene as a control gene. A study by Nadkarni et al., showed that the bacterial load was similar when an artificial mixture was analyzed by PCR and by colony counting, despite different species harboring different copy numbers of the 16S rRNA gene (Nadkarni et al 2002). A marked increase of erm(B) gene levels in the fecal samples was observed that persisted up to four years after treatment (Paper IV).

Highly macrolide-resistant enterococci have previously been reported to be selected by treatment with clarithromycin and metronidazole as evidenced by the increase in erm(B) levels in cultured enterococci after treatment (Sjölund et al 2003). One resistant clone persisted for three years without any further antibiotic pressure (Sjölund et al 2003). In a study by Jernberg et al., clindamycin treatment resulted in a lower Bacteroides diversity in fecal samples, an enrichment of resistant Bacteroides clones, specifically B. thetaiotamicron, and an increase in resistance erm-genes up to two years after treatment (Jernberg et al 2007). The increase and persistence of erm(B) can either be explained by clonal expansion of stable, resistant isolates originally present in the intestinl microbiota pretreatment, or by erm(B) acquisition by new populations via horizontal gene transfer. Transfer of the erm(B) gene has been shown between
different Gram positive strains, such as *Enterococcus* strains in the gut (Lester et al 2004) and *Streptococcus* strains in the throat (Nys et al 2005), but also for Gram negative strains such as between different *Bacteroides* strains (Gupta et al 2003, Shoemaker et al 2001). Epidemic spread of *erm*(B) has been shown to be the cause of increased macrolide resistance in *Streptococcus pyogenes* (Cresti et al 2002). Our results suggest a link between *erm*(B) gene levels and the increase in enterococci after treatment, indicating that the enterococci might be the bearers of antibiotic resistance (Paper IV).

Selection of resistant strains among the normal microbiota found in the throat has also been found. Clarithromycin has been shown to also select for macrolide-resistant bacteria in the throat microbiota (Jönsson et al 2005). In Paper IV, *erm*(B) was detected in throat samples from both patients and controls, although not in a high abundance in the controls.

The human indigenous microbiota could thereby potentially serve as a reservoir of resistance genes and contribute towards antibiotic resistance development for example, by transferring resistance genes to other species (Courvalin 1994, Sullivan et al 2001). Different factors have impact on the spread of antibiotic resistance, with the most important factor being antibiotic usage (Jernberg et al 2010). Other important factors include the relative fitness of the resistant strain and the ability of the resistant strain to survive and expand (Jernberg et al 2010). Normally a resistant strain incurs a fitness cost for carrying the resistant mutation and would be outcompeted by the sensitive bacteria when the antibiotic is removed (Andersson and Hughes 2010). However, by acquiring compensatory mutations, the resistant clone may compensate for the resistance becoming as fit as the sensitive part and thus still exist in the indigenous microbiota in the absence of antibiotics.
CONCLUDING REMARKS

Knowledge regarding the composition of GI-tract microbiota has increased through the use of high-throughput technologies like 454-pyrosequencing, especially for the intestinal microbiota. However, the current view is still far from complete regarding other body sites as well as for the intestinal microbiota, especially in relation to disease outcome. By further development of sequencing methods, including longer sequence lengths, we will be able to extract more information regarding the microbiota and its interaction with the host. In this thesis different molecular approaches were used to monitor the dynamics of the total microbiota in different patient cohorts. We also correlated antibiotic resistance genes and components of the immune system to the microbiota. The major conclusions from this thesis are:

In paper I, the 454-pyrosequencing technique was developed for amplicon sequencing and it was found to be a suitable method for analysis of the GI-tract microbiota. Thus, the use of high-throughput sequencing has revolutionized the field of clinical microbiology and increased our knowledge about the composition of the indigenous microbiota. However, this type of approach still demands considerable computing resources and bioinformatics skills to deal with the data. Therefore, depending on the scientific question raised other more traditional methods are still viable choices.

In paper II, a difference in intestinal microbiota composition was found between vaginally delivered infants compared to infants born by caesarean section, especially regarding the Bacteroidetes phylum, with a lower prevalence in the CS delivered infants. The Bacteroidetes phylum is mainly transmitted from the infant’s own mother and the presence of the genus Bacteroides was also associated with a higher abundance of Th1-chemokines at one and three months of age. This suggests that early differences in the gut microbiota may shape later immune responsiveness.

In paper III, a lower diversity in the intestinal microbiota was found at one month of age in infants that later developed allergic disease. The rise in allergic disease is thought to be a cause of an altered microbial exposure, especially during infancy. It was not possible to clarify whether a low total diversity of the gut microbiota in early childhood is more important than altered prevalence of particular bacterial species in allergy development.

In paper IV, a commonly used one-week antimicrobial treatment regimen resulted in marked ecological disturbances in the throat and gut microbiota with potential long-term consequences. By using 454-pyrosequencing in combination with T-RFLP and real-time PCR the ecological disturbances due to antibiotic treatment were thoroughly monitored. The findings underpin the importance of restrictive and proper use of antibiotics in order to prevent long-term ecological disturbances of the indigenous microbiota. In hospital environments, a decreased colonization resistance and persistence of antibiotic resistance
could potentially lead to an overgrowth and spread of multi-resistant potential pathogenic bacteria and thereby increase the risk of treatment failure.

Thus, the results from this thesis have increased our knowledge of complex human microbial communities during infancy and over time in adults. The results from this thesis also provide new knowledge regarding the interaction between the immune system and specific bacteria. Improving our understanding of the microbiota residing in the human GI-tract and its interaction with the immune system is important to better understand how the immune system develops and also to better understand disease outcome.
ACKNOWLEDGEMENTS

Lars Engstrand, my main supervisor. Thanks for the opportunity to work in this exciting research field. By always listening to me, supporting me in my decisions, and being generous and enthusiastic, you have helped me to develop both as a scientist and as a person. By also providing with social activities such as kick-offs and ski-trips outside work you have created a nice working atmosphere and have helped the research group to flourish.

Janet Jansson, my co-supervisor. Thanks for the opportunity to do part of my thesis work together with Cia at Södertörn University. Discussing science with you has been a privilege. Your hard-working and enthusiastic attitude towards science has encouraged me in my work. Thanks for your great involvement in the antibiotic manuscript and for pushing it to become published.

Cecilia Jernberg, my co-supervisor. Thanks for always being there for me and always listening to me regardless of the topic. It was a privilege to start my PhD studies working with you. I have learned a lot from you. You have influenced me in many ways, especially by sharing your deep knowledge of the microbiota and molecular research fields along with life in general.

I would like to thank all co-authors for their contributions: Anders Andersson, Mathilda Lindberg, Thomas Abrahamsson, Maria Jenmalm, Bengt Björkstén, Christopher Quince, Fredrik Bäckhed, and Pål Nyrén.

Anders Andersson, my bioinformatics supervisor! By providing your extensive knowledge of bioinformatic analysis and your ideas and creative way of questioning things the day-to-day work has become so much more fun and interesting. This has really contributed to the progress of the manuscripts. I have appreciated working with you and that you have been so helpful all the time.

Mathilda Lindberg, my fellow PhD-student. Thanks for your support and for always being kind and helpful, and for being my constant conference travel mate. Optimization our 16S rRNA gene analysis has been our major discussion topic over the years and we could definitely conclude that we have left nothing out.

Thomas Abrahamsson, Maria Jenmalm, and Bengt Björkstén, thanks for nice collaborations and great scientific discussions. It has been a long, but very nice, collaboration!

Christopher Quince, thanks for nice scientific discussions and help with the sequence analysis.
I would like to thank all present and past members of Lars Engstrand’s research group for being the best colleagues one could have, especially Sönke Andres, Annika Fahlén, Maria Nygård, Cecilia Svensson, Elin Lundin, Anna Skoglund, Britta Björkholm, Helene Kling Bäckhed, Wilhelm Paulander, Zongli Zheng, Sandra Rodin, Christina Persson, Maria Sjölund Karlsson, Annelie Lundin, Mårten Kivi, Christina Nilsson, Marianne Ljungström, Kristina Schönmeyer Kempe, Lena Eriksson, Britt-Marie Hoffmann, Karin Wreiber, Martin Storm, Heather-Marie Schmidt, Hazel Mitchell, Valtteri Virta, Philippe Lehours, Sophie Maisnier-Patin, Alexandra Pennhag, Katrin Pütsep, and Alma Brolund.

I especially would like to thank: Lena, Kristina, and Marianne for always being so nice and helpful in the lab. Britt-Marie for invaluable help and for providing with a fun and enjoyable working atmosphere. Wilhelm for always being so helpful, supportive and for providing annoying jokes in the lab. Sandra, Anna, and Helene for “belly times.” Sönke for support and fun times during our time as PhD-students and for continuing running the success “Tjorrej”. Annika for being a great exam worker.

I also would like to thank the people at CCUG and the Department of Clinical Bacteriology at Göteborg University, especially Edward, Liselott, Sofia, Maria, Elisabeth, Kent, Hesho, Anders, and Nahid, for providing me with a writing place when writing my thesis and also for nice coffee and lunch breaks.

I also would like to thank all other people that I have gotten the opportunity to get to know during my time as a graduate student, and who have contributed to a nice working atmosphere, scientific discussions and enjoyable coffee breaks. These people include past and present doctorate students, researchers, and other staff at SMI, MTC, and Södertörns högskola.


Mest av allt skulle jag vilja tacka min Joel och våra underbara barn Samuel och Liv för kärleken, livet, kasset och lugnet. Ni är min lycka!

Hedvig

This work was supported by grants from the Ekhaga foundation, the Söderbergs foundation, the Swedish Research Council, the Swedish Institute for Communicable Disease Control, and the Karolinska Institute.
REFERENCES


37


Johansson ME, Larsson JM, Hansson GC (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4659-4665.


Kalliomäki M, Kirjavainen PV, Eerola E, Kero P, Salminen SJ, Isolauri E (2001). Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing. *Journal of Allergy and Clinical Immunology* 107: 129-134.


