

# Microbiota of the alimentary tract of children

-implications for allergy and inflammatory bowel disease

Fei Sjöberg

孙红飞

Department of Infectious Medicine, Clinical Bacteriology  
Institute of Biomedicine,  
Sahlgrenska Academy at University of Gothenburg



UNIVERSITY OF GOTHENBURG

Gothenburg 2014

Cover illustration: Chinese calligraphy by my mother Yun Wang

Microbiota of the alimentary tract of children  
© Fei Sjöberg 2014  
Fei.sjoberg@microbio.gu.se

ISBN 978-91-628-9090-2

Printed in Gothenburg, Sweden 2014  
Kompendiet. Aidla Trading AB

To my family

給我的家人



# Microbiota of the alimentary tract of children

-implications for allergy and inflammatory bowel disease

Fei Sjöberg

Department of Infectious Medicine, Clinical Bacteriology, Institute of Biomedicine,  
Sahlgrenska Academy at University of Gothenburg  
Göteborg, Sweden

## ABSTRACT

Allergy, which is the most common chronic disease in Swedish children and adolescents, is associated with a high standard of living and Western lifestyle. According to the hygiene hypothesis, allergy is due to inadequate stimulation of the immune system by microbes during early childhood, leading to failed maturation of the immune system. The incidences of inflammatory bowel diseases, i.e., ulcerative colitis and Crohn's disease, have also increased dramatically in Western countries over the last few decades, and currently, these diseases are often diagnosed already in childhood. Epidemiological evidence suggests links between alterations to the intestinal microbiota and these diseases. Thus, studies of the composition of the bacterial microbiota in infants and young children are of relevance for the pathogenesis of both allergic diseases and inflammatory bowel diseases.

In this thesis, quantitative culture and DNA-based methods are compared for their abilities to characterize the gut microbiota in infants. Terminal-Restriction Fragment Length Polymorphism (T-RFLP) is based on differences in the 16S rRNA gene sequences between bacteria, as revealed by differences in fragment sizes after restriction enzyme digestion. A database was constructed to identify bacteria based on their fragment sizes. Multi-parallel sequencing of the 16S rRNA gene by pyrosequencing and T-RFLP were compared for sensitivity with quantitative culture of infant fecal samples. Bacterial genera that were present at  $>10^6$  colony

forming units/g feces, as determined by culture, were generally readily detected by DNA-based methods, with the exception of bifidobacteria, which generated only one sequence read per  $10^8$  viable bacteria. Clinically and immunologically relevant facultative bacteria, e.g., staphylococci, were often missed by the DNA-based methods due to having low counts in the fecal samples. The studies presented in this thesis indicate that cultivation and molecular-based assays are complementary in generating an overall picture of the complex gut microbiota.

In the ALLERGYFLORA cohort study, T-RFLP was used to analyze the salivary microbiota in 4-month-old infants whose parents had the habit of “cleaning” their pacifier by sucking on it, and of control children whose parents did not have this habit. Sharing of the pacifier between parent and infant was associated with reduced risk of the child developing an allergy and altered salivary microbiota in the child. We hypothesize that the oral bacteria transmitted from the parents stimulate the child's immune system in such a way that allergy development is avoided.

Samples of the duodenal fluids of children with newly diagnosed and untreated inflammatory bowel diseases (ulcerative colitis and Crohn's disease) and controls (having functional bowel disorders without signs of intestinal inflammation) were analyzed by culture and pyrosequencing. The microbiota of children with ulcerative colitis displayed lower bacterial diversity than that of the control children, and certain bacterial groups were less abundant in the former group.

Taken together, the studies presented in this thesis suggest that the compositions of the commensal microbiota in the oral cavity and small intestine affect the risk of developing immunoregulatory diseases, such as allergies and inflammatory bowel diseases.

**Keywords:** microbiota, alimentary tract, children, allergy, duodenum, inflammatory bowel disease, culture, T-RFLP, pyrosequencing

**ISBN:** 978-91-628-9090-2

## SAMMANFATTNING PÅ SVENSKA

Allergi är den vanligaste kroniska sjukdomen hos barn och ungdomar i Sverige och är förknippad med hög boendestandard, västerländsk livsstil och god hygien. Enligt hygienhypotesen beror allergi på att immunsystemet inte utsätts för adekvat stimulering av mikrober under uppväxten och därför inte mognar på ett korrekt sätt. Även inflammatorisk tarmsjukdom är vanligare i västländer än i fattiga länder med dålig hygien. Sjukdomen har ökat kraftigt och debuterar allt oftare under tidig barndom, vilket kan tyda på att bristande mikrobiell stimulering är en riskfaktor även för denna sjukdom. Västerländska spädbarn koloniserar senare med vanliga tarmbakterier än barn i fattiga länder och vår hypotes är att barnets bakterieflora kan tänkas påverka både risken för allergi och inflammatorisk tarmsjukdom senare i livet.

DNA-baserade metoder såsom pyrosekvensering och Terminal-Restriction Fragment Length Polymorphism (T-RFLP) används ofta för att karakterisera komplexa bakteriella ekosystem men studier av metodernas känslighet saknas. I avhandlingen har odling- och DNA-baserade metoder jämförts med avseende på förmågan att detektera och identifiera olika grupper av tarmbakterier. En databas utvecklades för att kunna identifiera bakterier utifrån fragmentens storlek i T-RFLP-analys. Resultatet visade att bakterier med odlingstal högre än  $10^6$ /g feces, i allmänhet kunde detekteras med DNA-baserade metoder, medan fakultativa bakterier som fanns i lågt antal ofta missades. Både odlings- och DNA-baserade analyser har sina begränsningar och båda kan behövas för att få en helhetsbild av den komplexa floran.

I Florastudien, en prospektiv kohortstudie, användes T-RFLP för att karaktärisera munflora hos spädbarn vars föräldrar sög på deras napp för att "rengöra" den, och från spädbarn vars föräldrar inte hade denna vana. Den mikrobiella sammansättningen i saliven skilde sig mellan spädbarnen i de båda grupperna vid 4 månaders ålder och barn vars föräldrar sög på deras napp hade lägre risk att senare utveckla allergi. Vår hypotes är att bakterier från

föräldrarnas munflora överförs till barnet och att dessa bakterier stimulerar barnets immunsystem så att allergiutveckling undviks.

Vi har också kartlagt tunntarmens bakterieflora med odling och pyrosekvensering hos barn med nydebuterad och obehandlad inflammatorisk tarmsjukdom. Resultatet visade att tunntarmsfloran hos patienter med den inflammatoriska tarmsjukdomen ulcerös kolit uppvisade en lägre bakteriell mångfald jämfört med floran hos kontrollbarn, alltså barn med funktionella tarmbesvär och normal tarmslemhinna.

Våra studier tyder på att sammansättningen av bakterieflora i mun och tunntarm kan påverka risken att utveckla immunregleringsjukdomar som allergi och inflammatorisk tarmsjukdom.

---

## LIST OF PAPERS

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

- I. Hesselmar B., Sjöberg F., Saalman R., Åberg N., Adlerberth I., Wold AE. **Pacifier Cleaning Practices and Risk of Allergy Development.** *Pediatrics* 2013; 131: 1-9
  
- II. Sjöberg F., Nowrouzian F., Rangel I., Hannoun C., Moore E., Adlerberth I., Wold AE. **Comparison between terminal-restriction fragment length polymorphism (T-RFLP) and quantitative culture for analysis of infants' gut microbiota.** *J Microbiol Methods* 2013; 94: 37-46
  
- III. Sjöberg F., Barkman C., Östman S., Nookaew I., Adlerberth I., Saalman R., Wold AE. **Altered composition of duodenal microbiota of children with newly diagnosed inflammatory bowel disease.** In manuscript.
  
- IV. Sjöberg F., Nookaew I., Adlerberth I., Wold AE. **Comparative analysis of infants' gut microbiota by next generation sequencing and quantitative culture.** In manuscript.

*Papers I and II are reprinted with permission from the American Academy of Pediatrics and Elsevier Limited, respectively.*

## TABLE OF CONTENTS

ABBREVIATIONS .....	4
INTRODUCTION .....	6
Bacterial classification-taxonomy .....	8
Microbiota composition in the alimentary tract .....	11
Oral cavity .....	11
Stomach .....	11
Small intestine .....	12
Large intestine .....	12
Microbial colonization of newborns .....	15
Methods for studying the microbiota .....	20
Culture .....	21
DNA-based approaches .....	26
Culture versus DNA-based method .....	39
Influence on the host of the microbiota .....	41
Infection .....	41
Microbiota interactions with the immune system .....	41
Gut microbiota and host metabolism .....	45
Microbiota and Allergy .....	46
Allergy – “The modern plague” .....	46
The hygiene hypothesis .....	49
Neonatal microbial exposure and the risk of allergy development .....	50
Oral flora and allergy development .....	51
Microbiota and inflammatory bowel diseases .....	52
AIM .....	55
MATERIAL AND METHODS .....	56
General overview of the four papers .....	56

---

Children and Samples .....	58
The AllergyFlora birth cohort (Papers I, II and IV).....	58
Inflammatory bowel disease study (Paper III).....	60
Methods .....	61
Quantitative culture (Papers II, III, and IV).....	61
DNA-based methods (Papers I-IV).....	63
Statistical analysis .....	66
RESULTS AND COMMENTS.....	67
Pacifier cleaning practices and allergy development in infants...	67
Development of a database for the identification of T-RFs.....	74
Infant microbiota development .....	76
Methodologic considerations-Advantages and drawbacks .....	82
Detection limits of the three methods.....	82
Comparison of the performances of pyrosequencing for analyzing small and large bowel microbiota.....	89
Duodenal microbiota and IBD.....	92
DISCUSSION.....	93
ACKNOWLEDGEMENTS .....	106
REFERENCES.....	108

## **ABBREVIATIONS**

IBD	Inflammatory Bowel Disease
CFU	Colony Forming Unit
OPLS	Orthogonal Partial Least Squares
O2PLS- DA	Orthogonal 2 Partial Least Squares- Discriminant Analysis
OTU	Operational Taxonomic Unit
T-RFs	Terminal-Restriction Fragments
T-RFLP	Terminal-Restriction Fragment Length Polymorphism

---

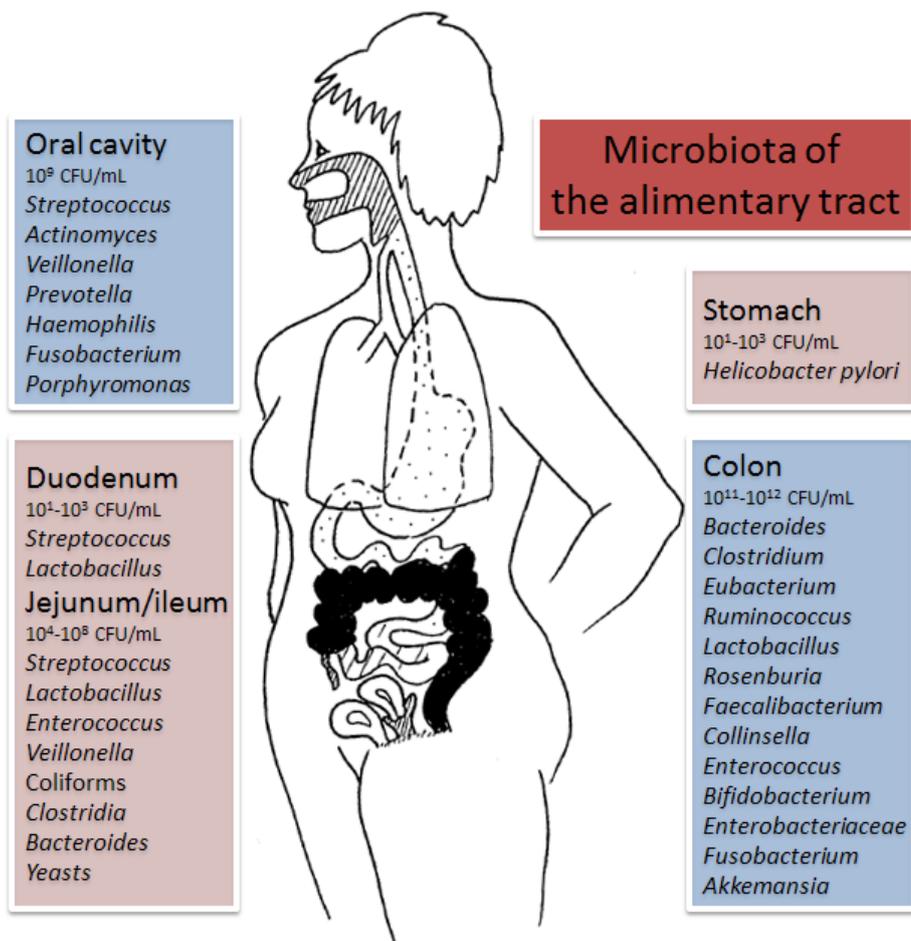
## INTRODUCTION

The alimentary tract is an approximately 8-m long tubular passage that extends from the mouth to the anus, through which food passes and becomes digested. The alimentary tract is colonized by diverse microbial communities, which are collectively known as the microbiota.

The microbiota of the human gastrointestinal tract is composed of hundreds of species and up to  $10^{14}$  bacterial cells. This microbial society contains potential pathogens and the mucosal immune system faces a challenging task in protecting the host from invasive pathogens, while, at the same time, tolerating harmless dietary antigens and not reacting excessively to the bulk of commensal bacteria and their inflammatory products.

There is growing interest in the physiologic interactions between the microbiota and the host. It has been suggested that the human microbiota contributes to health and disease through providing nutrients, stimulating the immune system, and exerting colonization resistance against pathogens. Most of the existing body of knowledge regarding microbiota composition comes from culture-based studies, although massive nucleic acid sequencing studies of the microbiota at various bodily sites are underway, e.g., in the Human Microbiome Project. Few studies have compared the performance of culture-based and sequence-based methodologies for characterization of the commensal microbiota.

The figure exemplifies some of the major bacterial groups colonizing different parts of the alimentary tract.



Female figure by Agnes Wold

## Bacterial classification-taxonomy

Living organisms are classified into a common taxonomic system, the basic level of which is the species. Each newly discovered species must be assigned to a genus in a binary nomenclature established by Carl von Linné. A genus is a member of successively higher ranks: subfamily, family, suborder, order, subclass, class, phylum (or division), and domain (or empire), and a specific suffix is added to a taxon between genus and class (Table 1). A taxon (plural, taxa) is a general definition in microbiology; a taxonomic category or group, such as a phylum, order, family, genus, or species.

Table 1. Bacterial taxonomy

Taxonomic rank	LPSN <sup>#</sup>	Suffix	Example
Subspecies	416	No	<i>P. freudenreichii</i> subsp. <i>freudenreichii</i> <i>P. freudenreichii</i> subsp. <i>shermanii</i>
Species	10 599	No	<i>Propionibacterium freudenreichii</i>
Genus	2001	No	<i>Propionibacterium</i>
Family	271	-aceae	<i>Propionibacteriaceae</i>
Suborder	21	-ineae	<i>Propionibacterineae</i>
Order	126	-ales	<i>Actinomycetales</i>
Subclass	6	-ida*	<i>Actinobacteridae</i>
Class	97	-ia*	<i>Actinobacteria</i>
Phylum	30	No	Actinobacteria

<sup>#</sup>Total validly published since the Approved Lists of Bacterial Names up to August. 2013 [1]. \*Proposed suffix

Bacterial nomenclature (a system of names) is covered by the rules of the Bacteriological Code [2]. Several classification systems exist, including those proposed by Pace [3], Ludwig et al. [4], Hugenholtz [5], and Bergey's Manual of Systematic Bacteriology [6], the latter being the most widely accepted authority on this topic among microbiologists.

---

According to Bergey's Manual, the bacteria domain is divided into 30 phyla, and as of November 2013, 15,974 taxa have been identified; these are listed in the List of Prokaryotic Names with Standing in Nomenclature (LPSN) [1]. LPSN is a database that lists the names of prokaryotes (Bacteria and Archaea) that have been validly published in the International Journal of Systematic and Evolutionary Microbiology, under the Rules of the International Code of Nomenclature of Bacteria.

### **The bacterial species**

A species is defined as “a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity in many independent characteristics, and is diagnosable by a discriminative phenotypic property” [7]. However, for bacteria, this definition is ambiguous, which is known as the ‘species problem’ [8]. The original taxonomic schemes were based on isolation by culture and classification according to macroscopic and microscopic appearances, metabolic behavior (mostly the ability to ferment various carbohydrates), and the presence of surface antigens that were reactive to specific antibodies produced for this purpose (serology). Since DNA methods started to be applied for species definition, separate species have been found to be genetically very similar. Thus, *Escherichia coli* and the four species of the genus *Shigella*, which were previously categorized based on morphology, serotyping, and biochemical tests [9], are now assigned to a single species based on DNA relatedness. However, they are still considered to be separate species, as distinguishing *E. coli* from the highly pathogenic *Shigella* is important in the clinical setting.

Today, a novel species is typically identified through a combination of traditional tests, mainly biochemical analyses and lipid profiling, as well as on the basis of DNA relatedness. The naming of a new species requires that its similarity to any other known species is <98.5% for the 16S rRNA gene and <70% for DNA-DNA hybridization of the entire genomes [10] [11]. Recently, based on the available bacterial whole-genome sequences, Kim *et al* [12] showed that the average nucleotide identity (ANI) threshold range of 95%–96% has taxonomic potential for species differentiation.

This ANI range corresponds to 16S rRNA gene sequence similarity of 98.65%. ANI could serve as a substitute for the labor-intensive DNA–DNA hybridization technique.

➤ *Subspecies and strains*

Some species are subdivided into subspecies (abbreviated as ‘subsp.’). There are no clear rules for this; it depends on practical interest in division below the species level. *Propionibacterium freudenreichii* is a Gram-positive bacterium that plays an important role in the fermentation of Swiss Emmental cheese. Table 2 shows the example of the two subspecies *P. freudenreichii* subsp. *freudenreichii* and *P. freudenreichii* subsp. *Shermanii*, which are differentiated on the basis of nitrate reduction and lactose fermentation [13], factors of importance in cheese production.

Table 2. *Propionibacterium freudenreichii* subspecies differentiation.

<i>P. freudenreichii</i> subsp.	NO <sub>3</sub> reduction	Lactose fermentation
<i>freudenreichii</i>	+	-
<i>shermanii</i>	-	+

Strain refers to a colony of genetically uniform microorganisms that derive from a common ancestor, e.g., a clone growing out in culture or a clone of *E. coli* that inhabits the large intestine of an individual for a period of weeks or months. Strains within a species may be separated by serotyping, typing of enzyme electrophoretic mobility (due to point mutations), or characterization of the outer membrane protein profiles (for Gram-negative bacteria). A clinically important application of strain/clone typing is the tracing of the source of a disease outbreak.

---

## Microbiota composition in the alimentary tract

### Oral cavity

The oral cavity is a major gateway for bacteria to enter the human body and provides a number of bacterial colonization habitats, e.g., the surfaces of the teeth, tongue, and tonsils, and the gingival crevices. Biofilms, which are composed of multilayered microcolonies, may form on several of these oral surfaces [14]. The saliva is rich in bacteria, usually containing around  $10^9$  colony-forming units (CFU)/mL.

The oral microbiota is estimated to contain more than 600 species, distributed in 13 phyla. The six major phyla contain 96% of the taxa, i.e., Firmicutes (37%), Bacteroidetes (17%), Proteobacteria (17%), Actinobacteria (12%), Spirochaetes (8%), and Fusobacteria (5%) [15]. Notably, the majority (65%) of the bacterial taxa in the oral cavity have been successfully cultured.

At the genus level, *Streptococcus*, *Actinomyces*, *Veillonella*, *Prevotella*, *Haemophilis*, *Fusobacterium*, and *Porphyromonas* are frequently detected in the oral cavity [16-18].

### Stomach

The stomach has a sparse flora due to the low pH of the lumen. Using culture, *Lactobacillus*, *Streptococcus*, *Clostridia*, *Veillonella*, coliforms, and yeasts are found at population levels that vary according to the diet and geographic location of the person [19]. The composition of the gastric microbiota varies considerably between individuals, as revealed by both culture-based and DNA-based methods. Indeed, the microbes isolated from gastric contents may be transients, either having passed down from the nasopharynx or oral cavity, or having being introduced in ingested food. *Helicobacter pylori* is a true long-term colonizer of the stomach. It is more readily detected by DNA-based methods than by culturing [19, 20].

## **Small intestine**

The small intestine is divided into the duodenum, jejunum, and ileum.

### **Duodenum**

The stomach acid kills most of the ingested bacteria, and the excretion of digestive enzymes and bile acids into the duodenal lumen also exerts a strong selective pressure on the numbers and types of transiting bacteria. Bacterial counts are therefore generally low in samples taken from the duodenum, typically  $10^2$ – $10^3$  CFU/mL of contents. Only a few studies have examined the composition of the duodenal microbiota. Two duodenal biopsy studies have shown that the genera *Streptococcus* and *Neisseria* (both belonging to the Firmicutes phylum) predominate, although *Veillonella*, *Gemella*, *Clostridium* cluster XIVa, and *Haemophilus* are also detected. In contrast to the large intestinal microbiota, members of the Bacteroidetes and Actinobacteria phyla are scarce in the duodenal microbiota. When present, they are represented by the genera *Prevotella*, *Bacteroides*, *Actinomyces*, and *Rothia* [21] [22].

### **Jejunum and ileum**

The numbers of bacteria increase along the small intestine, from  $10^{2-4}$  bacteria/mL in the proximal jejunum to  $10^6$ – $10^7$  bacteria/mL in the terminal ileum [23]. The milieu is rather rich in oxygen. Culture-based studies have indicated that *Lactobacillus*, *Streptococcus*, *Clostridia*, *Veillonella*, *Bacteroides*, and coliforms are the dominant bacterial groups, and that yeasts, mainly *Candida*, are also present [19].

## **Large intestine**

In the colon, the bacterial density is very high, often exceeding  $10^{11}$  CFU/g colonic content. As there is approximately 1 kg of colonic contents, an adult individual harbors  $10^{14}$  bacteria in the colon [24]. Most studies of the large intestinal microbiota have been based on analyses of fecal samples because of their ready availability. However, differences in community composition between stool and mucosal samples have been reported [25]. The

---

colonic microbiota of an individual is rather stable over time, although inter-individual variations can be substantial [26]. Geographic region, diet, and age play significant roles in shaping the composition of the colonic flora [27] [28].

Strictly anaerobic bacteria predominate in the large intestine due to the low level of oxygen in the lumen. The number of species harbored by a given individual is a subject for debate. Classical culture-based studies performed in the 1970's estimated that approximately 400 species could be present in an individual [29]. In one PCR-based Next generation sequencing study, the V4 region of the 16S rRNA gene was sequenced after PCR amplification. Numerous reads ( $1.8 \pm 0.6$  million/sample), were obtained from 531 individuals from the Amazonas of Venezuela, rural Malawi, and US metropolitan areas. The mean number of operational taxonomic units (OTUs)/sample for each area were 1600, 1400, and 1200, respectively [27]. The shot-gun metagenomics sequence study of the Human Intestinal Tract (MetaHIT) project, which involved the examination of 124 European subjects, showed that each individual harbored at least 160 bacterial species [30]. The MetaHIT study concluded that "given the large overlap between microbial sequences in this and previous studies, we suggest that the number of abundant intestinal bacterial species may be not much higher than that observed in our cohort", i.e., a few hundred, which is clearly much lower than the number revealed by the study using 16S rRNA sequencing. Using a mock community, shotgun metagenomics sequencing showed more accurate species determination than PCR-based 16S rRNA gene next generation sequencing [31].

DNA-based studies show that up to 80% of the microbial sequences identified in fecal samples are novel and represent uncultivated bacteria [25, 32]. At the phylum level, the fecal microbiota is dominated by Bacteroidetes and Firmicutes, which comprise 60%–80% of the fecal bacteria. Actinobacteria, Proteobacteria, Verrucomicrobia, and Fusobacteria are also found, albeit in lower numbers (Table 3) [33, 34]. Table 3 shows the proportions of bacteria in the colonic microbiota, as demonstrated by both DNA-based technique [33].

Table 3. Colonic microbiota in composition

Taxon	Share of the microbiota (%)	Genus/Species
Firmicutes	40–65	
Clostridial cluster		
IV	25	<i>Clostridium leptum</i>
		<i>Faecalibacterium prausnitzii</i>
		<i>Ruminococcus</i>
IX	7	<i>Megasphaera</i>
		<i>Veillonella</i>
		<i>Selenomonas</i>
		<i>Megamonas</i>
XIVa	25	<i>Clostridium</i>
		<i>Anaerostipes</i>
		<i>Ruminococcus</i>
		<i>Roseburia</i>
		<i>Eubacterium</i>
		<i>Coprococcus</i>
XI		<i>Clostridium bartlettii</i>
XV		<i>Anaerofustis stercorihominis</i>
<i>Lactobacillales</i>		<i>Lactobacillus</i>
Bacteroidetes	25	
		<i>Bacteroides</i>
Actinobacteria	3–5	
		<i>Bifidobacterium</i>
		<i>Collinsella</i>
Proteobacteria	0–3	
		<i>E. coli</i>
Verrucomicrobia	1–2	
		<i>Akkermansia mucinophila</i>
		<i>Victivallis vadensis</i>
Archaea		<i>Methanobrevibacter smithii</i>
		<i>Methanosphaera stadtmanae</i>

Data adapted from Duncan *et al.*, Appl. Microbiol, 2007 [33]

---

Metagenomic studies have shown that 99% of the phylogenetically assigned genes of the large intestinal microbiota are of bacterial origin, although Archaea (e.g., *Methanobrevibacter*), Eukaryotes (e.g., yeasts), and viruses (e.g., bacteriophages) are also present [30] [35].

### **Microbial colonization of newborns**

Microbial colonization of the neonate begins at birth as the fetus leaves the normally sterile environment of the womb. The establishment of the microbiota during infancy proceeds in a sequential manner over several years until hundreds of different bacterial species are established in both the oral cavity and the gut. The sources of the colonizing bacteria are: parents, siblings, pets, foods, and the air. Given the different exposure conditions, the composition of the microbiota may vary with delivery mode, feeding regime, hygiene measures, and living standard, as well as the extent of social contacts.

#### **Oral cavity colonization**

Relatively little is known regarding the establishment of the oral microbiota in infants. There is a paucity of longitudinal studies, and most of the studies that have been conducted have been culture-based. More recent studies employing DNA-based methodologies have focused on oral pathogens, such as colonization with *Streptococcus mutans* in relation to the risk of developing dental caries.

Teeth appear at a few months of age in infants. Before eruption of the teeth, various streptococci act as the first colonizers, e.g., *S. salivarius*, *S. mitis*, and *S. oralis* [36]. For example, in one study, streptococci were detected 8 hours after birth, while *Staphylococcus* and *Neisseria* were isolated later on during the first day, and anaerobic species of *Veillonella* and *Bifidobacterium* appeared from the second day [37]. All these bacteria were consistently present throughout the first year of life, although the bacterial counts of streptococci decreased over time [38]. Another study showed that the most frequently isolated early anaerobic colonizers of the oral cavity were *Prevotella melaninogenia*,

*Fusobacterium nucleatum*, and non-pigmented *Prevotella* spp. [39].

After emergence of the teeth, the bacterial diversity in the oral cavity increases as the numbers of retention sites and potential niches increase. It is unclear at what age the oral microbiota becomes adult-like. *Bacteroides melaninogenicus* was found in <40% of five-year-old children, but in essentially all 13–16-year-old adolescents [40].

In the oral cavity, bacteria may adhere to each other, forming a biofilm on available surfaces, within which they live in a symbiotic partnership [36]. Fastidious anaerobic bacteria growth even in edentulous mouths can be explained by the formation of biofilms. *Fusobacterium nucleatum* may have a central role in the maturation of oral biofilm communities [41]

### **Gut colonization**

The neonatal gut is initially an aerobic environment, and facultative anaerobic bacteria are usually the first colonizers. Initially, many of the bacteria are transient colonizers [42]. The gut colonization fluctuates markedly from hour to hour in the neonatal period [43]. In the early period of life, the gut microbial composition varies significantly from baby to baby [42, 44, 45].

As shown in twin and kinship studies, common environmental exposures play greater roles in shaping gut microbial ecology than does genetic relatedness [27]. Vaginally delivered infants are colonized first by the cervical and fecal flora of the mother [46] [47]. Babies delivered by Cesarean section are exposed initially to bacteria from the hospital environment and healthcare workers [47, 48]. Cesarean section-delivered infants less often harbor typical bacteria of fecal origin, such as *Escherichia coli*, *Bacteroides*, and *Bifidobacterium* species, while other bacteria encountered in the environment are more frequent, e.g., *Klebsiella* and *Clostridium* species [44, 45, 49]. One study has reported a lower level of microbial diversity in the gut microbiota of Cesarean section-delivered babies, as compared with that in babies delivered vaginally [42].

---

Milk formula-fed infants have a more diverse microbial community than breast-fed infants [44]. While this might reflect greater exposure to bacteria, it is equally likely that breast milk exerts a strong selective pressure on the gut bacteria population and that only a few bacterial groups are able to withstand this selection pressure.

Figure 1 shows the gut microbial colonization patterns of 300 children from three European birth cohorts, followed from 3 days of age to 1 year of age using culture-based analyses of fecal samples [49]. The left panel of Figure 1 shows the cumulative colonization frequency, i.e., the proportion of infants who harbor or have harbored a certain bacterial genus, while the right panel shows the population levels of the same genera in infants who harbor the genus in question. It is clear that coagulase-negative staphylococci (CoNS) and enterococci are the first colonizers among the facultative bacteria, and that the frequencies of colonization with *E. coli* and *S. aureus* are approximately equal. Colonization by *Klebsiella* and other members of the *Enterobacteriaceae* family is also relatively common in infancy (panel A). Regarding population levels, it is evident that the population levels of *E. coli* are consistently retained over the first year of life (panel B). Other members of the *Enterobacteriaceae* family and enterococci are present in lower level than *E. coli* at one year of age. In sharp contrast, the population levels of staphylococci, both coagulase-negative staphylococci and *S. aureus*, decrease dramatically over the first year of life, and in 1-year-old children, their population levels average  $10^4$ – $10^5$  CFU/g feces. The early microbiota contains few species and the competition for nutrients and space is limited, with the consequence that most bacteria grow to quite high population levels. In parallel with the increase in the complexity of the microbiota during the first year of life, the population levels of bacteria that cannot survive the competition decrease. Staphylococci, which are typical skin colonizers, are examples of bacteria that are not well suited to persisting in a microbiota that has reached full complexity.

Panels C and D in Figure 1 depict the colonization patterns of some groups of strictly anaerobic bacteria. Bifidobacteria are the first

colonizers, followed by a variety of *Clostridium* spp. [49]. *Bacteroides*, which is a Gram-negative commensal gut bacterium, colonizes only a minority of these infants during the first months, although its colonization frequency increases over the entire first year of life. Lactobacilli are not a major colonizer of the infant microbiota, although they become more prevalent towards the end of the first year of life (panel C).

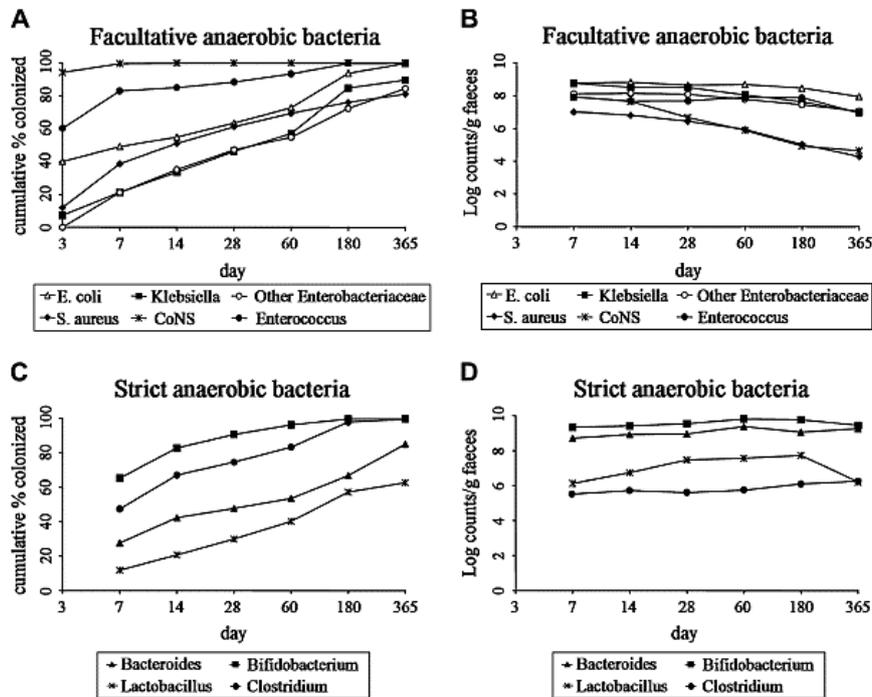


Figure 1. Intestinal colonization pattern during the first year of life. The results are presented as the proportion of infants ever colonized by each time point and the mean log<sub>10</sub> count for colonized infants only at each time point for facultative anaerobic (A and B) and strict anaerobic bacteria (C and D). CoNS, Coagulase-negative staphylococci. Reproduced with permission from Adlerberth et al. Data shown are based on the results of cultivation of fecal samples from 300 European infants [49]

Regarding population levels, infants who are colonized by bifidobacteria or *Bacteroides* harbor consistently high levels of these bacteria, i.e., >10<sup>9</sup> CFU/g feces (panel D). The lactobacillus population levels in colonized infants are generally much lower

---

(panel D). Clostridia are detected and enumerated by alcohol-treatment of the samples, which kills living bacteria but leaves spores intact. Thus, the population levels shown in panel D represent clostridial spore counts in the feces of culture-positive infants.

***Conversion to an adult-type microbiota***

The complexity of the microbiota increases as more and more anaerobic species become established in the infant gut [50] [27]. From the time of weaning, diversification of the diet may promote replacement of pioneer gut colonizers with members of the microbiota that will persist into adulthood [51]. At 1 year of age, although still distinct, the microbiota starts to converge toward that of an adult [42]. However, it is not until 2–3 years of age that the young child’s gut microbiota resembles the adult flora [27, 52].

With decreasing oxygen tension, the numbers of facultative bacteria decline. Thus, the ratio of facultatives to anaerobes changes from roughly 1:1 in newborn infants to 1:500 in adults [53].

A meta-analysis that combined NGS data on infant and adult gut microbiota from three separate studies revealed an age gradient, with a transition from *Enterococcaceae*, *Enterobacteraceae*, *Streptococcaceae*, *Lactobacillaceae*, *Clostridiaceae*, and *Bifidobacteraceae* at an early age to communities in adults that were enriched for *Lachnospiraceae*, *Ruminococcaceae*, *Bacteroidaceae*, *Prevotellaceae* and others [54].

## Methods for studying the microbiota

Much of our knowledge regarding the human microbiota has been derived from culture-based studies. Cultured bacteria are subcultured for purity and may be identified to the family, genus, species or even strain level. Various methods are available for the identification of bacterial isolates. The culture-based and molecular techniques that are commonly used for microbial identification are listed in Table 4. All of these methods require pure culture isolates as starting point.

*Table 4. Culture-dependent taxonomic resolution levels based on current bacterial identification techniques.*

Method	Taxonomic resolution level			
	Family	Genus	Species	Strain
<b>Non-DNA-based</b>				
Biotyping	X	X	X	(X)
Fatty acids (VFA/FAME) <sup>a</sup>	X	X	X	
MALDI-TOF <sup>b</sup>	X	X	X	
<b>DNA-based</b>				
<b>Sequencing non-dependent</b>				
PCR (specific-primer)	X	X	X	
DNA-DNA hybridization			X	
PFGE <sup>c</sup>				X
MLVA <sup>d</sup>				X
<b>Sequencing-dependent</b>				
16S rRNA sequencing	X	X	(X)	
Whole-genome sequencing	X	X	X	X
Multilocus sequence typing				X

(X) denotes occasionally achievable

<sup>a</sup>Volatile Fatty Acid/Fatty Acid Methyl Esters

<sup>b</sup>Matrix-Assisted Laser Desorption-Ionization-Time-Of-Flight

<sup>c</sup>Pulsed-Field Gel Electrophoresis

<sup>d</sup>Multiple-Locus Variable-number Tandem Repeat

In recent years, molecular techniques that permit the identification of previously non-culturable anaerobes have been developed, and these are now widely used to describe microbial ecosystems (Table 5).

*Table 5. Non-culture-dependent methods used to study bacterial communities*

Method	Taxonomic resolution level		
	Family	Genus	Species
<b>Sequencing-non-dependent</b>			
PCR (specific-primer)	X	X	X
Microarray	X	X	(X)
T-RFLP, D/TGGE*	X	X	(X)
<b>Sequencing-dependent</b>			
Cloning and sequencing	X	X	(X)
NGS (PCR-based) <sup>a</sup>	X	X	(X)
NGS (DNA-based)	X	X	(X)
Single-cell sequencing	X	X	X

(X)denotes occasionally achievable. <sup>a</sup>Next generation sequencing

\*Terminal Restriction Fragment Length Polymorphism

\*Denaturing/Temperature Gradient Gel Electrophoresis

## **Culture**

Microbial culture is essential for the isolation of a pure culture of a bacterium. Pure cultures are required to assess antibiotic susceptibility patterns. Bacterial isolates in pure cultures may also be subtyped to the strain level, which allows the spread of clones to be monitored in epidemiologic studies. Methods, which include serotyping, PFGE (Pulsed-Field Gel Electrophoresis), MLVA (Multiple-locus variable-number Tandem-repeat), and MLST (Multilocus sequence typing), can be used to distinguish strains within a particular species, although all these methods require pure culture isolates as the starting point [55].

### Culture conditions

Bacteria can be classified based on dependence on oxygen into: aerobic, anaerobic, and facultative anaerobic. Thus, a milieu that is rich or poor in oxygen can be applied as a selective force to enable the selection of facultative or strictly anaerobic bacteria.

Aerobic bacteria derive energy from oxidative processes, whereas obligate anaerobic bacteria use fermentation or anaerobic respiration as their energy source. Obligate anaerobic bacteria are also vulnerable to oxygen exposure, as they die in air. Obligate anaerobes show extensive differences with respect to oxygen sensitivity. For example, *Roseburia intestinalis* survives for less than 2 minutes when exposed to air on an agar surface, while other anaerobic bacteria, here exemplified by *Anaerostipes caccae*, survives quite well in ambient air for some time (Figure 2).

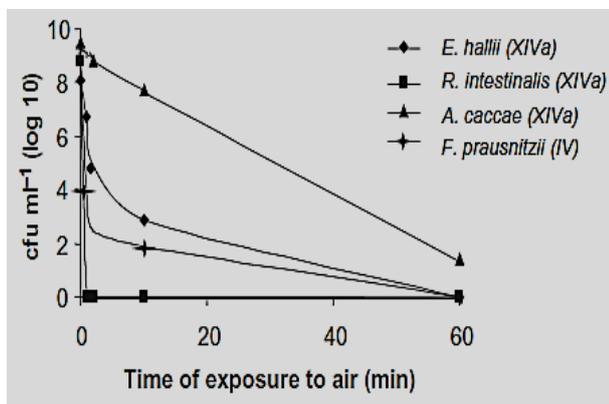


Figure 2. Aerotolerance of some Clostridium-related Firmicute bacteria from the human gut. Reproduced with permission from John Wiley and Sons. *Environmental Microbiology*, 2007 [56]

Facultative anaerobic bacteria can survive under anaerobic conditions but they grow better in the presence of oxygen. Culturing in air can be applied as a selective pressure for facultative bacteria that are present in relatively low numbers among far higher numbers of strict anaerobes.

The nutritional requirements vary between bacteria. Non-selective media meet the nutrient requirements of many different bacteria, an example being Brain-Heart Infusion agar/broth. Selective media are used to isolate specific bacteria from a mixed population. They

---

permit growth only of the targeted microbe or a few types of microbes. For example, MacConkey agar contains bile salts, which permit the growth of Gram-negative bacteria of the types found in the intestinal flora, i.e., the *Enterobacteriaceae* family, while inhibiting the growth of most Gram-positive bacteria. In Scandinavia, Drigalski agar is used for the same purpose. It is slightly less selective than MacConkey agar, permitting scant growth of Gram-positive enterococci. Staphylococci are able to grow in the presence of 10% NaCl, and media that are selective for staphylococci are based on this principle. As all other gut-colonizing bacteria die in this heavy salt environment, staphylococci that are present at very low concentrations, e.g.,  $10^3$  CFU/g, can be isolated from a mixed population of  $10^{11}$  fecal bacteria.

The number of bacteria is determined by quantitative culturing. Serial dilutions of a sample are plated onto solid medium. The dilution that yields 10–100 colonies is selected. The colonies are enumerated and multiplied by the dilution factor, yielding the bacterial counts as colony-forming units (CFUs).

Gram-staining was developed by the Danish pathologist Hans-Christian Gram [57]. First, crystal violet is added to bacteria smeared on a glass slide. After fixation with iodide solution, the slide is destained with ethanol or acetone and then counter-stained with saffranin. Gram-positive bacteria stain blue, as their tightly meshed and thick cell wall retains complexes of cytoplasmic proteins and crystal violet. The thin cell wall of Gram-negative bacteria cannot retain the crystal violet-protein complexes, which leak out of the cell, so that it loses the color. Gram-negative bacteria are stained red by the saffranin applied at the end of the staining process.

### **Isolate based identification methods**

A trained microbiologist can, at times, identify a bacterial species solely based upon colony morphology on selective or non-selective media and smell, e.g., *Pseudomonas aeruginosa*, *E. coli*, and *S. aureus*. Gram-staining followed by observation under the microscope adds information regarding the presence of cocci (round bacteria), bacilli (rod-shaped bacteria), and endospores

(bacilli, clostridia). In a clinical laboratory, many of the clinically relevant bacteria may be relatively easily identified in this way.

➤ *Biochemical reactions*

Biochemical tests are commonly used to identify bacteria. The presence or absence of certain bacterial enzymes may be highly informative. Catalase converts hydrogen peroxide to water and oxygen, and all staphylococci have this enzyme. Coagulase converts fibrinogen to fibrin and a clot is formed. *Staphylococcus aureus* is coagulase-positive, while many other staphylococcal species, such as *S. epidermidis* and *S. haemolyticus*, are coagulase-negative and are grouped together as the coagulase-negative staphylococci. Coagulase-positive *S. aureus* is a pathogen, in that it causes sepsis, osteomyelitis, endocarditis, pneumonia, and local skin abscesses. In addition, its production of toxins, especially “superantigens”, may cause food poisoning and shock. In contrast, the coagulase-negative staphylococci seldom cause disease, except in immunocompromised individuals and newborn infants. Low-grade and long-term infections of prosthetic devices are, however, caused by coagulase-negative staphylococci.

Biotyping, which entails enrichment of a pure culture of the bacteria to ferment different types of carbohydrates, was, prior to the advent of DNA-based methods and matrix-assisted laser desorption ionization – time of flight (MALDI-TOF), the major way to achieve bacterial speciation. Commercial biotyping kits are available, such as API 20E for the *Enterobacteriaceae* family and Rapid ID 32A for anaerobic bacteria, e.g., *Bacteroides* and *Clostridium*.

➤ *Fatty acid characterization*

There are two types of fatty acid analysis methods: Fatty Acid Methyl Esters (FAME); and Volatile Fatty Acids (VFA). FAME analyzes long (>12) fatty acids that are part of the bacterial cell wall, whereas VFA analyzes the short-chain fatty acids (SCFA), which are produced by anaerobic bacteria through the fermentation of carbohydrates in the culture medium. SCFA include acetic acid, lactic acid, propionic acid, and butyric acid. Different groups of bacteria produce different proportions of fatty

---

acids, which are separated and detected by gas chromatography, with the resulting pattern indicating the identity of the bacterium [58]. For example, bifidobacteria mainly produce acetate, while some clostridial species are butyrate producers.

➤ *MALDI-TOF*

Since 2005, matrix-assisted laser desorption ionization – time of flight (MALDI-TOF) has become the standard method for rapid identification of bacteria in the clinical laboratory [59]. Species identification by MALDI-TOF is more rapid, accurate, and inexpensive than other procedures based on molecular or biochemical tests. A colony of the microbe in question is smeared directly onto a metal surface, which is then bombarded by a laser, and the macromolecules spread from the colony, i.e., nucleic acids, proteins, and carbohydrates, are analyzed by mass spectroscopy. The spectrum generated by a specific bacterium is analyzed using dedicated software, and identification is achieved by comparison with stored profiles.

## **DNA-based approaches**

During the last decades, many DNA-based methods have been developed to classify bacteria. These methodologies are generally referred to as “molecular” methods. They all depend on the detection of differences in DNA sequences between bacteria. They can be basically divided into sequencing-dependent and sequencing-non-dependent methods.

### **Sequencing-non-dependent**

#### ➤ *Polymerase chain reaction*

Polymerase chain reaction (PCR) was developed in 1983 by Kary Mullis [60] and is now a widely used technique in medical and research laboratories. Species-specific PCR can be used to identify bacteria down to species level. The primers used comprise synthetic small fragments of DNA that hybridize to the targeted DNA sequence of interest. During the PCR, the target DNA sequence is amplified many times, enabling its detection based on size or by sequencing.

#### ➤ *16S rRNA gene*

In the 1980s, Carl Woese described an rRNA-based phylogenetic taxonomy for the classification of bacteria [61]. The 16S rRNA gene contains highly conserved regions in which the DNA sequences are very similar between different bacteria. Interspersed among these conserved regions are hypervariable regions, known as Variable (V) regions, the sequences of which differ among bacterial taxa. Conserved regions are suitable for universal PCR primer construction, while the variable regions can be used for taxon identification. However, the conserved regions are not absolutely identical between all bacteria. Therefore, primers are always more or less selective, favoring certain groups and disfavoring others [62]. Another problem is that, depending on the size of the sequenced fragment, it may not be possible to identify a bacterial taxon, since there are too few sites within that region that unequivocally define specific taxa.

Figure 3 shows a representation of the entire 1500-bp 16S rRNA gene, with its nine V regions, V1-V9, which range in length from 50

to 100 nucleotides [63]. The figure shows the chance of correctly classifying a bacterium to the phylum, class, order, family or genus level, based on the utilization of a 100-nt that encompasses different parts of the 16S rRNA gene. It is evident that the most potent discrimination is achieved using a primer that encompasses either the V2 region or the V4 region. However, it is also clear that classification to the genus level is near impossible using a 100-nt sequence, as the maximal accuracy that can be achieved is 80%. Using such relatively short sequences, bacteria can be reliably classified only to the phylum or class level (Figure 3).

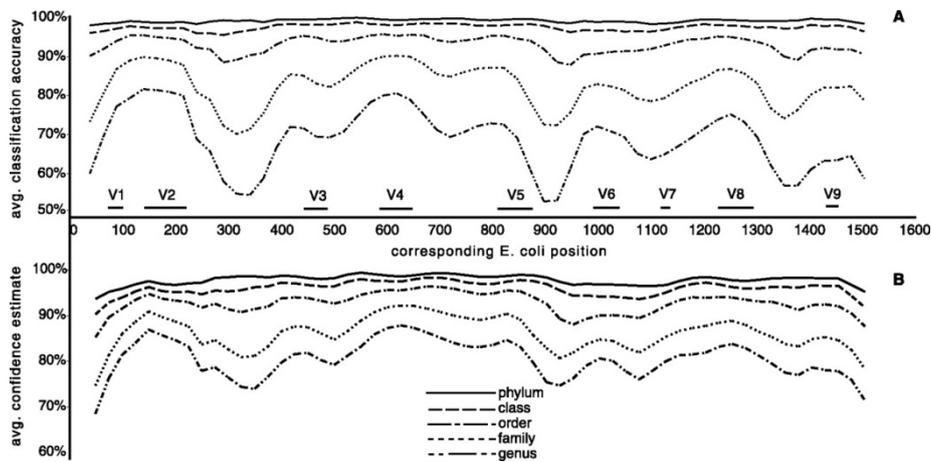


Figure 3. (A) Classification accuracy rate for the Bergery corpus with sequence segments of 100 bases, moving 25 bases a time. The average classification accuracy rate at the genus level was 70% over all 100-base regions. (B) Average bootstrap confidence. Reproduced with the permission of the American Society for Microbiology. Wang et al. *Appl Environ Microbiol*, 2007.

Figure 4 shows the chance of correctly assigning a genus, family, order, class or phylum to a 16S rRNA gene segment of different length. As shown in the figure, with 400-bp sequences of the 16S rRNA gene, there is an 88.7% chance of correctly identifying the genus of the bacterium in question, a 94.6% chance of assigning it to the correct family, and a >99% chance of correctly determining its class or phylum.

Despite this general relationship, some species cannot be resolved to the species level, even by sequencing the full-length 16S DNA. As this is true for streptococcal species, highly pathogenic pneumococci cannot be distinguished from a wide range of other streptococcal species, collectively referred to as the  $\alpha$ -streptococci or “viridans” streptococci. These streptococci are numerous in the oral cavity flora and upper small intestinal flora. The *Enterobacteriaceae* family (*E. coli*, *Klebsiella* etc.) is also notoriously difficult to separate solely based on 16S rRNA analysis, due to the close genetic relationships. *Enterobacteriaceae*, particularly *E. coli*, are important members of the gut flora.

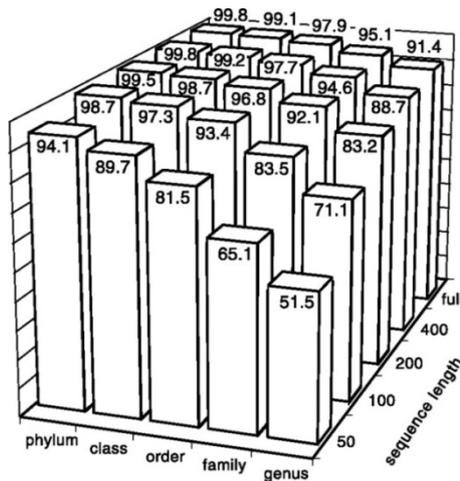


Figure 4. Overall classification accuracy by query size. Numbers are percentages of tests correctly classified. Figure adopted with the permission from the American Society for Microbiology Wang et al. *Appl Environ Microbiol*, 2007.

Another complication is that bacteria may have several copies of 16S rRNA genes. Although 40% of bacteria have one or two copies, up to seven copies are common [64].

Within a single bacterium, the different copies of the 16S rRNA genes may differ in their DNA sequences, which may hamper the sequence annotation [65].

Another factor that influences the detection limit of 16S rRNA gene-based analysis is the efficacy of DNA extraction. DNA is more readily released from Gram-negative bacteria, such as *Bacteroides* and *Enterobacteriaceae*, while some Gram-positive bacteria, such

---

as *Enterococcus* and *Staphylococcus*, have very thick and sturdy cell walls, so it may be difficult to access the DNA of these bacteria.

A final concern is “universal” PCR primer mismatch [66]. In particular, Gram-positive bacteria that have a high GC content (e.g., Actinobacteria) are notoriously difficult to analyze with commonly used PCR primers [67].

➤ *Fingerprinting methods*

DNA fingerprinting is a technique that is used to identify individual microbes in a community based on their respective DNA profiles. Terminal Restriction Fragment Length Polymorphism (T-RFLP), Denaturing Gradient Gel Electrophoresis (DGGE), and Temperature Gradient Gel Electrophoresis (TGGE) are DNA fingerprinting methods. They generate a pattern of fragments that are characteristic of a strain or an entire bacterial community. They are based on differences in the DNA sequence of a specific gene, often the 16S rRNA gene, which are revealed by electrophoretic separation. Whereas fingerprinting techniques provide less information than current sequencing methods, they are orders of magnitude cheaper and faster to perform. A significant disadvantage of fingerprinting methods is that taxon identification is not straightforward and requires further processing, e.g., the excision of DNA bands from the gel and subsequent identification of the DNA fragments by sequencing.

DGGE and TGGE are based on the GC-contents of the PCR-amplified sequences of a microbial community. The amount of energy needed to separate the double-stranded DNA varies depending on the GC content, with a single strain producing one band and a community giving a series of bands in the gel. Downstream identification requires band excision and sequencing.

T-RFLP was developed by Liu et al. in 1997 [68] and is widely used. T-RFLP is based on differences in the 16S rRNA gene sequence between bacteria (Figure 5) After amplification of the 16S gene by PCR and cleavage of the products with restriction enzymes, each bacterial taxon (genus or species) generates a fragment of 16S rRNA gene of a certain length, while a complex bacterial community

generates a series of DNA fragments of different sizes. The terminal fragments are detected by the generated fluorescence, as the primer is labeled with a fluorochrome. A series of non-labeled fragments is also generated upon digestion, but only one of the fragments (the terminal one) is detected per bacterium (Figure 5).

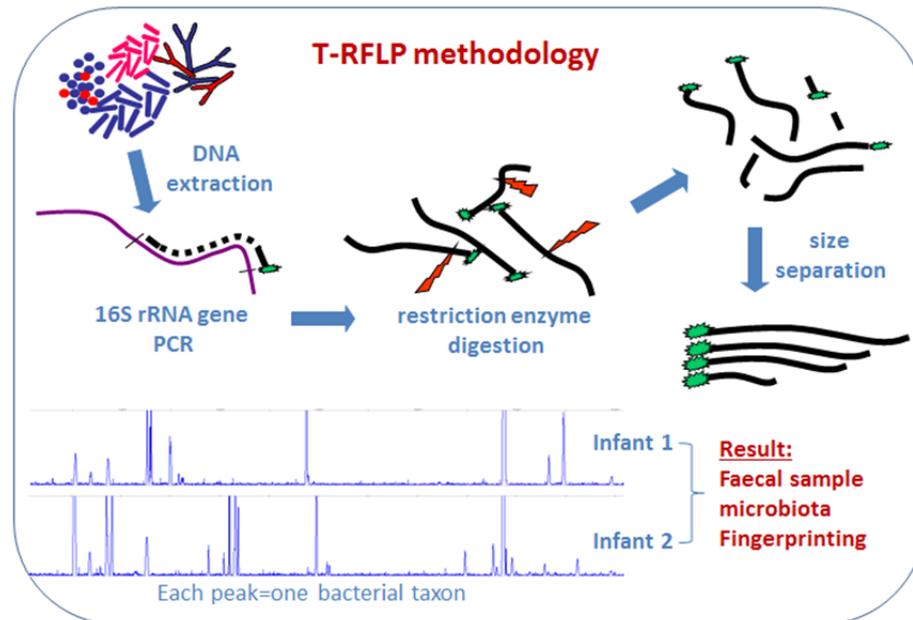


Figure 5. Schematic of the T-RFLP methodology and the fingerprints obtained for the faecal samples from two infants

As is the case with other fingerprinting methods, direct information regarding the taxonomic identity of the species is not generated. In the case of T-RFLP, databases can be produced that "translate" fragment size to genus/species [69]. In theory, it is possible to predict the size of the terminal fragment based on knowledge of the sequence of the 16S rRNA gene for the species in question, combined with knowledge of the cleavage sites of the restriction enzymes used, in a so-called '*in silico* analysis'. However, this does not work well in practice, due to technical problems with the method, such as the nonlinear relationship between the fragment size and the migration time of the fragment [48, 70]. Furthermore, the fluorescent dye label contributes to the overall molecular mass of the fragment. The molecular mass

---

standard is usually labeled with another fluorochrome, and in many cases, the manufacturer does not reveal the size of their fluorochrome. As a result, while *in silico* analysis can aid in identification, it is insufficient on its own.

➤ *Microarray*

Microarray analysis, which is an approach that is intermediate between fingerprinting and next-generation sequencing, allows rapid analysis of multiple samples, as do fingerprinting methods, but it also allows identification. Microarray analysis is based on short DNA probes that are attached to a solid surface; the probes usually consist of parts of the 16S rRNA gene that are specific for different bacterial taxa. Thousands of probes can be synthesized, and the bacterial DNA in the sample can hybridize to the species-specific oligonucleotide probes. Hybridization is detected by fluorescence or chemiluminescence. The human intestinal tract chip (HITChip) [71] and phyloChip [72] are two microarrays that contain probes that hybridize with the sequences of common members of the human gut microbiota. A disadvantage of the method is that the microarray can only identify already known bacteria. A more serious problem is that sequence homologies among different bacteria may result in cross-hybridization and incorrect classification.

### **Sequencing dependent DNA-based methods**

➤ *Cloning and sequencing*

The first method of sequencing to identify individual bacteria within a community, e.g., in a fecal sample, was the cloning of a PCR-amplified fragment of a bacterial gene (usually the 16S rRNA gene) into a vector carried by *E. coli*, followed by Sanger sequencing analysis of the cloned PCR fragment. This method is very labor-intensive and at most some hundreds to a thousand DNA molecules in a sample can be reasonably sequenced. On the one hand, only the most abundant bacteria in the community can be detected and minor populations are often missed. However, long fragments, sometimes the entire 16S rRNA gene of 1500-nt, can be cloned and sequenced, providing good taxonomic accuracy.

➤ *Next generation sequencing*

Since 2005, various next-generation sequencing (NGS) techniques have been developed and applied in studies of the microbiota. NGS is based on parallel sequencing of high numbers of DNA molecules. Parallel sequencing technologies overcome the limitations of Sanger sequencing (referred to as 'first-generation sequencing'). NGS enables rapid analysis of large-scale sequences in microbiota studies, e.g., the human microbial project (HMP), and at different body sites, such as the digestive tract, mouth, skin, nose, and female urogenital tract [73].

There are several NGS techniques available. The most frequently used techniques for studies of microbiota involve the Roche/454 and Illumina/Solexa sequencers. In general, the Roche sequencer generates long read lengths ( $\approx 700$  bases) but has lower data output capacity ( $\approx 700$  Mb/run), whereas the Illumina sequencer produces shorter but paired-end reads ( $\approx 100$ – $300$  bases) and has a greater output ( $\approx 15$ – $600$  Gb/run).

NGS microbiota studies can be further divided into PCR-based NGS and DNA sequence-based metagenomic NGS (Figure 6). In PCR-based NGS, a specified region of the 16S rRNA gene of the studied community is amplified, and the amplicons are sequenced. In metagenomic NGS, all the DNA present in a sample (including DNA of human, viral and Archaeal origin) is recovered directly and sequenced, without prior PCR amplification. Thus, the differences detected in many parts of the genome between bacterial taxa contribute information and also PCR amplification biases are avoided, which increases the ability to analyze faithfully the genetic variations within a community. However, the quality and reliability of the sequence assembly remain critical, since the full DNA sequences are based on the assembly of shorter DNA sequences. This assembly is based on either known full-genome sequences of bacteria (mapping assembly) or, for bacteria that have not been cultured and therefore have not generated whole-genome data, algorithms (so-called '*de novo* assembly') [74]. Naturally, such a construction of genes and genomes faces many potential pitfalls. Metagenomic analysis may be further divided into functional metagenomics, in which metabolic pathways are

identified, and sequence-based metagenomics, which addresses population structure and genetic relatedness.

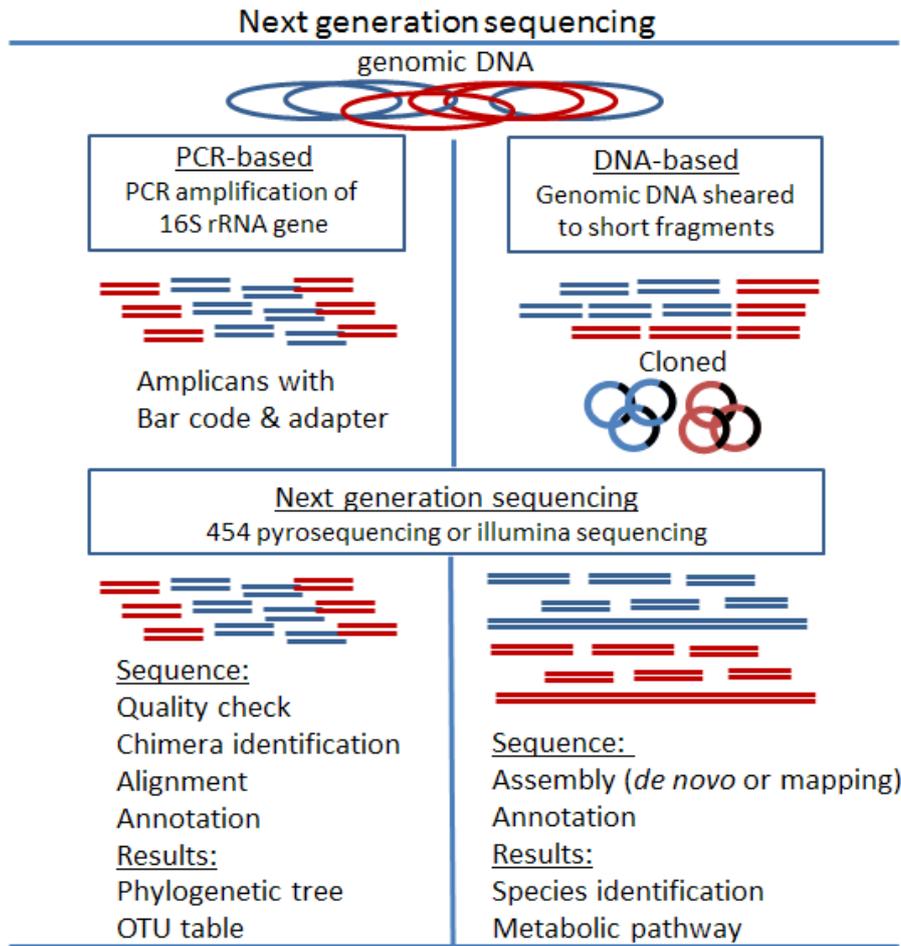


Figure 6. Schematic of the next generation sequencing methodology

### ***Single-cell sequencing***

Single-cell sequencing is an emerging technique. Individual cells can be isolated from microbial communities using fluorescence-activated cell sorting (FACS) or microfluidic chips. After cell lysis,

the genomic DNA is amplified by the multiple displacement amplification (MDA) technique [75]. MDA is a non-PCR method; it starts with the annealing of random hexamer primers to the denatured DNA, which is followed by strand-displacement synthesis at constant temperature. This generates DNA products of high molecular weight (7–10 kb in length) with lower error rates than traditional Taq polymerase-based PCR amplification. *Porphyromonas gingivalis* that was recovered from a biofilm in a hospital sink was successfully analyzed using MDA combined with NGS techniques [76]. Single-cell sequencing is a promising method that may overcome the limitations of other NGS methods due to its abilities to detect strain variations and to obtain sequences from taxa that are present in low abundance in a community.

---

### ***Advantages and drawbacks of PCR-based NGS and metagenomics***

Although NGS analysis of 16S rRNA gene amplicons is a powerful tool, there are many technological challenges involved in the analysis of the resulting massive dataset. In principle, all the data from 16S-based analyses of the human gut microbiota should be comparable. However, meta-analyses of 14 different studies of the Western adult fecal microbiota using PCR-based NGS of the 16S rRNA gene have shown that “the samples generally clustered by study, and the 16S rRNA target region, DNA extraction technique, and sequencing platform produced systematic biases in observed diversity that could obscure biologically meaningful compositional differences” [54]. The thresholds are used in the assignment of the sequence to a new “operational taxonomic unit” are also factors that may generate differences in the observed results [77]. As the cut-off levels for DNA identity that are used to define a new species are based on the sequence of full-length 16S rRNA, there is no unambiguous definition as to when a short NGS sequence represents a “new” taxon, when compared to a similar, albeit non-identical, short NGS sequence [78]. PCR has an inherent imperfection, misincorporation of bases by the DNA polymerase enzymes is roughly  $1.2 \times 10^{-6}$ /PCR cycle [79]. Hence, mistakes early in the amplification cycle can multiply exponentially. Incorporation of nucleotide base during a sequencing reaction produces a light reaction, homopolymers (single-base repeats) could be miss recorded since the light intensities do not necessarily increase linearly with the number of consecutive bases [80].

Which sequence alignment program is chosen and how sequence gaps are treated, whether variable positions are removed (a process termed “masking”) [78] (Figure 7). Effects of the different gap treatments influence genetic distance which was also strongly affected by the variation in sequence length and region of the 16S rRNA genes. All these factors will influence the subsequent diversity estimation as phylogenetic distance was calculated by summing the branch length for the entire tree [78].

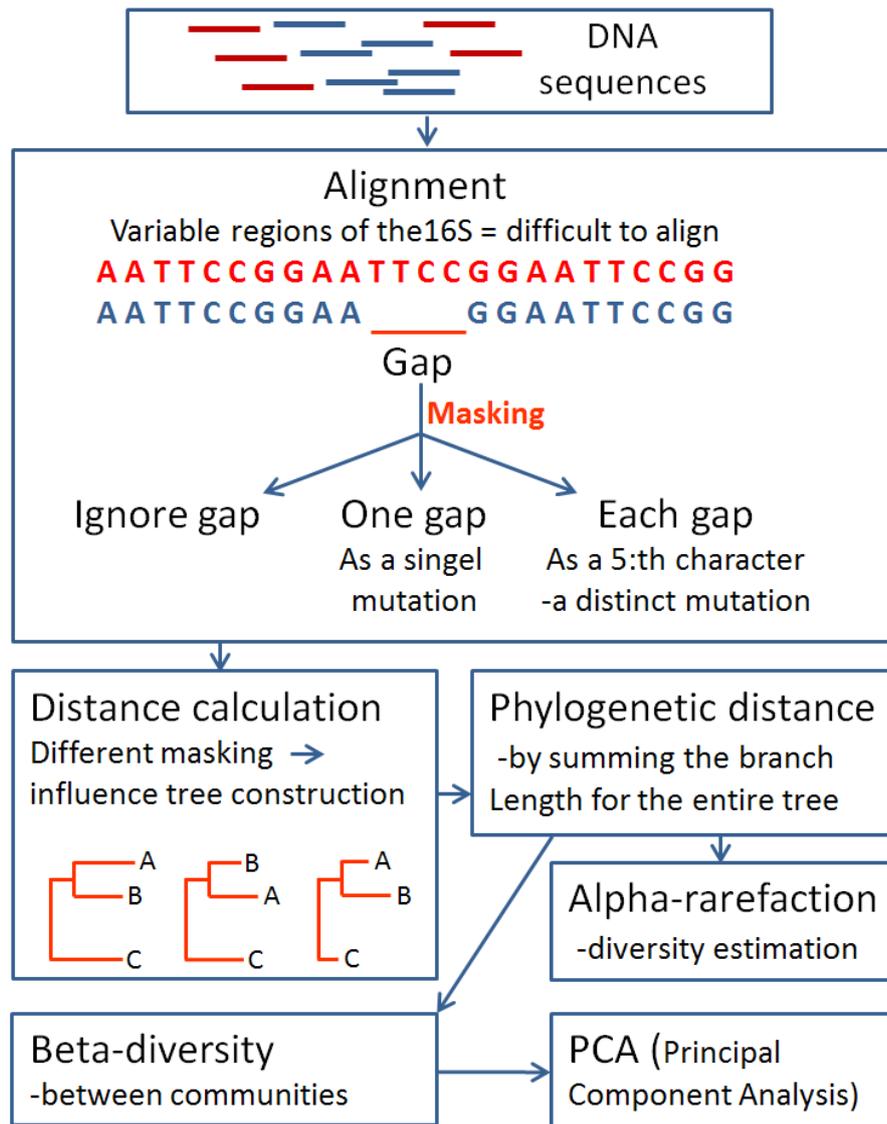
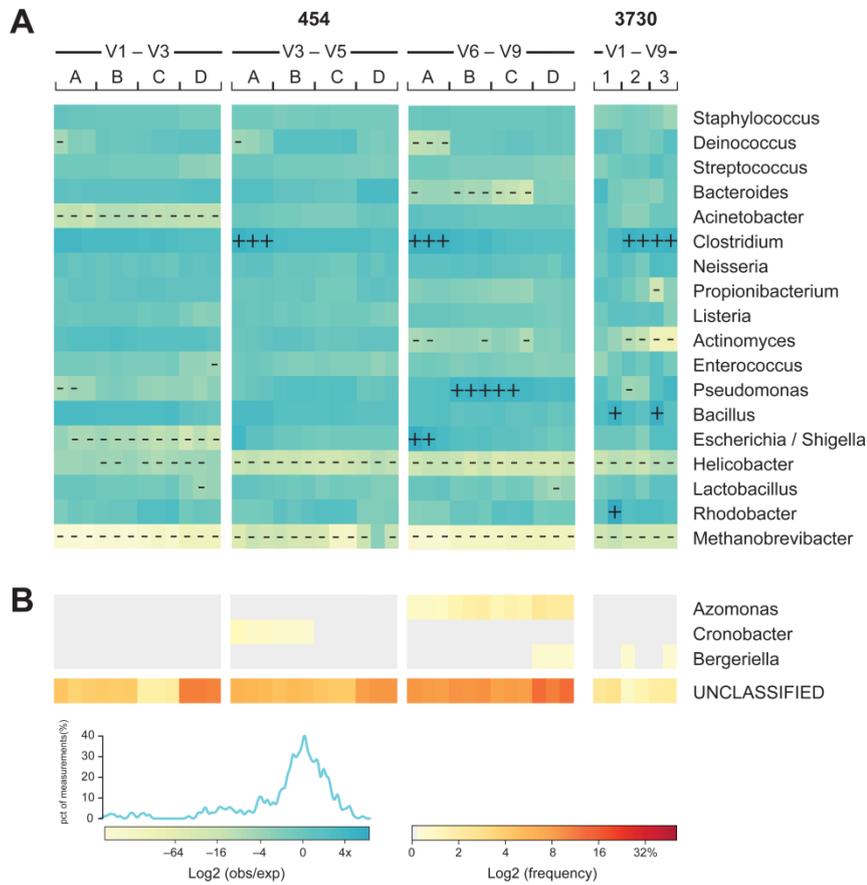


Figure 7. Schematic of effects of alignment quality, distance calculation and sequence masking on the sample's richness or phylogenetic diversity.

The Human Microbial project is a systematic project to identify and address the methodologic problems associated with PCR-based NGS of the 16S rRNA gene. Four different leading NGS centers were assigned a community of 21 species, i.e., “the Mock” (Figure 8). Using different sequencing platforms and targeted 16S

regions, sequence artifacts could be identified that resulted in the misclassification of bacteria, generating false measures of species diversity and creating “new” taxa that were actually not included in the Mock *i.e.*, *Azomonas*, *Cronobacter* and *Bergeriella* [66], but that were instead the consequences of mistakes in the PCR, sequencing, and alignment processes.



**Figure 8. Mock community-based accuracy of community representation compared across technology and 16S window.** The MC was sequenced by different centers on both 3730 and 454 platforms. Each sequencing trial is represented as a column. For 3730 sequencing of the V1-V9 window, amplicons derived from a common amplification protocol were sequenced with short capillaries (1), long capillaries (2), and three reads per clone (3). 454 sequencing was performed by four centers (A, B, C, and D) with three 16S windows (V1-V3, V3-V5, and V6-V9). (A) The observed genus level frequency data over expected genus-level frequency ratio for each of the MC members is shown as a heatmap using a binary logarithm scale. The expected frequency ratio is based on the whole genome coverages inferred from mapped Illumina WGS reads to the MC reference genome sequences. Genera with observed frequencies differing more than four-fold from expected are marked with +

or – for over- or under-representation, respectively. (B) The fraction of misclassified (0.1% of the total combined data set) and unclassified (4.6% of the total combined data set) sequences displayed as a frequency heatmap. The frequency values are depicted as a binary logarithm scale [66].

➤ *Chimera formation during PCR*

Falsely identified novel taxa could be due to methodological problems such as formation of chimeras, introduced variations in the DNA sequence during PCR multiplication (Figure 9). Chimeras were found to reproducibly form among independent amplifications and contributed to false perceptions of sample diversity and the false identification of novel taxa, with less-abundant species exhibiting chimera rates exceeding 70% [31].

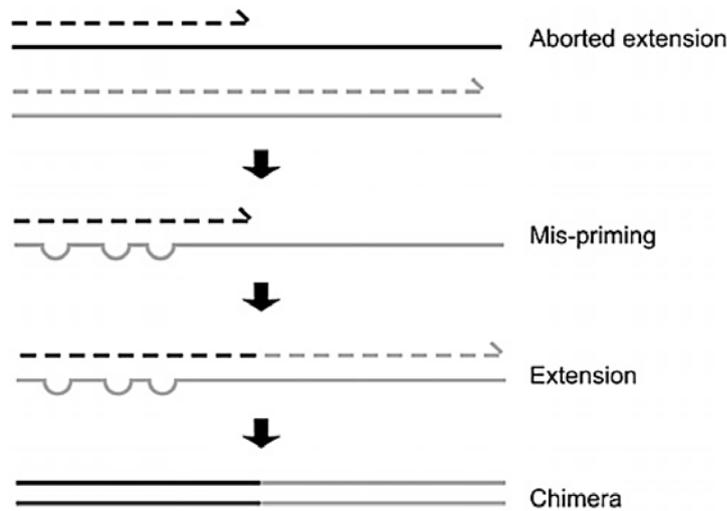


Figure 9. Formation of chimeric sequences during PCR. An aborted extension product from an earlier cycle of PCR can function as a primer in a subsequent PCR cycle. If this aborted extension product anneals to and primes DNA synthesis from an improper template, a chimeric molecule is formed [31]. Haas et al. *Genome Res.* 2011, Cold spring Harbor laboratory Press.

---

## Culture versus DNA-based method

Bacteria that can be cultured in the laboratory represent only a small fraction of the bacteria that exist in nature. Nonetheless, it is only through the isolation of individual bacterial species in pure culture that characterization of physiologic features and a full assessment of virulence potential, i.e., antibiotic resistance and pathogenicity, can be performed.

Molecular techniques permit the identification of bacteria that have hitherto not been cultured, mainly very strict anaerobes. However, many strictly anaerobic bacteria, such as *Roseburia*, *Faecalibacterium*, and *Eubacterium*, are readily cultured [33], so the failure to culture certain gut bacteria may reflect a lack of effort rather than an inherent unculturability [56]. Bomar *et al.* [81] characterized the bacterial genome to identify growth factors and to design media for cultivating previously “unculturable” organisms. A possible reason for a failure to culture a bacterium is its lack of specific metabolic pathways. For example, *Tropheryma whipplei* lacks the genes for nine pathways of amino acid biosynthesis; genome sequencing was used to identify the missing pathways and to design a culture medium that supported the growth of *T. whipplei* [82].

There is a general belief that DNA-based methods, especially NGS, have the ability to reveal “all” the bacteria in the gut microbiota. However, every method has its limitations. It is important to point out that in a complex bacterial community, Bacteria with small population size will be missed. The table below exemplifies the number of reads that will be generated if 10,000 sequences are obtained from a faecal sample in which a *Bacteroides* strain is present at  $10^{10}$  CFU/g, a *Bifidobacterium* strain is present at  $10^9$  CFU/g, etc (Table 6). Species with population sized of  $<10^6$  will be missed e.g. *Enterococcus*

Table 6. Theoretical detection limit of sequencing of a complex microbiota.

**10,000 DNA sequence reads from a complex microbiota (human colon)**

	CFU/g	No. of reads
<i>Bacteroides</i>	$10^{10}$	9000
<i>Bifidobacterium</i> spp	$10^9$	900
<i>Clostridium</i> spp	$10^8$	90
<i>E. coli</i>	$10^7$	9
<i>Staphylococcus</i>	$10^6$	<b>1 (0.9)</b>
		10,000
<i>Enterococcus</i>	$10^5$	<b>0 (0.09)</b>

---

## **Influence on the host of the microbiota**

### **Infection**

The gut microbiota serves as a reservoir for pathogenic bacteria. *E. coli*, *Klebsiella*, and other members of the *Enterobacteriaceae* family are the major causes of urinary tract infections and are often the causes of infant sepsis. [83]. Commensal bacteria seldom cause disease at the site of colonization, although there are exceptions to this rule. While *Clostridium difficile* is carried by most people without producing any symptoms, if competing bacteria are suppressed by treatment with broad-spectrum antibiotics, *C. difficile* may expand in numbers and its toxins may trigger diarrhea and mucosal inflammation [84]. The gut microbiota may serve as a reservoir for antibiotic-resistant strains, and also as a milieu in which resistance genes can be transferred between strains [83].

### **Microbiota interactions with the immune system**

#### **Basic immunology concepts**

The fundamental role of the immune system is to eliminate or neutralize microorganisms. Innate immunity and adaptive immunity are the two arms of the human immune system.

Innate immunity recognizes conserved microbial structures and becomes activated within minutes of infection. The innate immune system is therefore crucial for the elimination of invading microorganisms. It comprises cells (neutrophils, eosinophils, basophils, natural killer cells, mast cells, monocytes, macrophages, and dendritic cells) and soluble components (complement and coagulation factors, acute phase proteins, and cytokines), which synergize to kill and eliminate invading organisms.

Adaptive immunity (also known as acquired immunity) recognizes specific antigens through highly diverse antigen receptors on T and B lymphocytes. Upon activation, naïve T cells differentiate into various subtypes with different effector and regulatory functions. Naïve B cells differentiate into antibody-secreting plasma cells or into

memory B cells, which are rapidly reactivated when they encounter the same antigen anew.

Microbes are eliminated through co-operation between the innate and acquired immune systems. For example, antibodies (acquired immunity) activate complement (innate immunity), leading to enhanced phagocytosis and activated T cells (acquired immunity), which increase the capacity of macrophages to kill ingested bacteria.

### **The mucosa-associated lymphoid tissue**

The mucosal surfaces are protected by the mucosa-associated lymphoid system. The mouth, nose, and intestines are important sites for the induction of immune responses, both mucosal and systemic, and indeed, most foreign antigens enter the body *via* these sites.

#### ➤ *Nasopharynx-associated lymphoid tissue*

The nasopharynx-associated lymphoid tissue (NALT) comprises the adenoids, the palatine tonsils, and isolated lymphoid follicles, sometimes referred to as Waldeyer's pharyngeal ring [85]. It has also been implicated in the development of immune maturation, as intranasal exposure to a microbial extracts may induce active immune maturation [86] The tonsils are rich in FOXP3+ T cells, which are thought to be regulatory T cells, and *in vitro* activation of human tonsilla T cells induces immune suppressive activity [87].

#### ➤ *Gut-associated lymphoid tissue*

The gut-associated lymphoid tissue (GALT) comprises the mesenteric lymph nodes, the Peyer's patches, isolated lymphoid follicles in the intestinal wall, intraepithelial lymphocytes, and diffusely scattered lymphocytes in the gut lamina propria (Figure 10) [88-90]. The small intestine is a major inductive site for acquired mucosal immune responses [91]. The Peyer's patches are located in the small bowel, with approximately 200 of these structures being found in the average adult. Bacteria that reside at mucosal sites are able to translocate, i.e., remain intact and viable while crossing the mucosal barrier [89] [90]. This results in immune activation and it has been proposed that the gut

microbiota is central to immune maturation during infancy [88]. Animals that are reared germ-free have poorly developed Peyer's patches and have lower serum IgG and IgA levels, as compared with conventionally reared animals [92]. Regulatory T cells obtained from germ-free mice have poorer suppressive capacities than the regulatory T cells from mice that have an established microbiota [93].

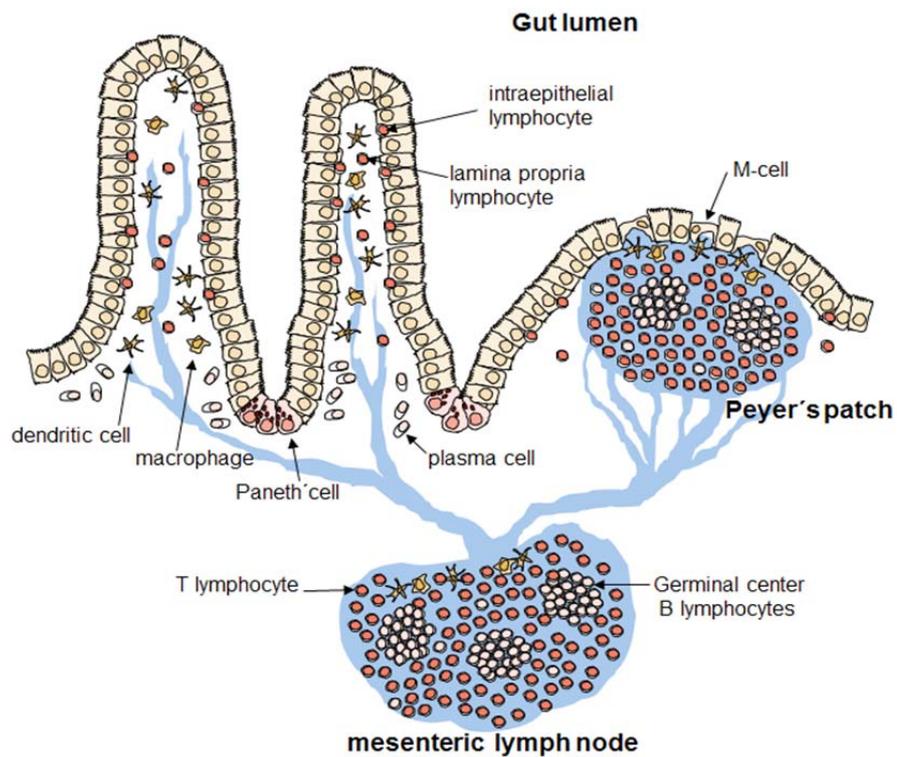


Figure 10. Gut-associated lymphoid tissue. Copy right @Ingegerd Adlerberth

### **Oral immune system**

Relatively little is known about the immune defense system of the oral cavity. Sublingual immunotherapy, which involves placing an allergen under the tongue, is a convenient way of reducing allergenic immune responses, e.g., to grass pollens.

Bacteria in the oral cavity have a pronounced capacity to translocate to the cervical lymph nodes [94] and therefore, the oral microbiota may be a major immune stimulant, albeit one that is generally overlooked.

### **Oral tolerance**

Immunologic unresponsiveness to harmless antigens, such as food proteins, is crucial for the maintenance of health. This unresponsiveness is actively induced and antigen-specific, and is termed 'oral tolerance'. Oral tolerance is defined as the ability of an orally administered antigen to suppress subsequent local and systemic immune responses [95]. Despite decades of research, the exact mechanisms underlying oral tolerance have not been defined. The small intestine is a major site of induction of oral tolerance, although the contributions of the NALT and the oral mucosa are not well understood. Oral tolerance involves both the mucosal epithelium, which produces tolerogenic packaged antigen, so-called "tolerosomes" [96], and tolerogenic antigen-presenting cells in the gut-associated lymphoid system [95]. Oral tolerance is more short-lived in animals that are reared without a microbiota than in conventionally reared animals [97], and germ-free animals have a reduced capacity to convert antigens into tolerosomes [96].

---

## **Gut microbiota and host metabolism**

### **Short-chain fatty acids**

The microbiota helps the host to metabolize otherwise indigestible dietary fibers, host mucin, and resistant starches into short-chain fatty acids (SCFAs) [98]. SCFAs are produced through bacterial fermentation and include acetate, propionate, and butyrate, isovaleric acid, and other fatty acids. Acetate is the main SCFA in the bloodstream; it is readily absorbed and transported to the liver where it is used in the production of blood lipids. Propionate is used by the liver to produce glucose, through a process termed gluconeogenesis [99]. Butyrate is produced by certain obligate anaerobic bacteria within the Firmicutes phylum, mainly those belonging to clostridial clusters IV and XIVa [100]. Butyrate serves as an important nutrient for host colonocytes [101, 102]. In fact, colonic epithelial cells derive their energy mainly from the gut lumen, rather than from the blood. Butyrate has been reported to exert anti-inflammatory [103] and anti-tumorigenic [104] effects in the colon, and Roediger proposed in the 1980s that ulcerative colitis, an inflammatory bowel disease, was due to an inability of the colonic epithelium to utilize butyrate as fuel [101].

## Microbiota and Allergy

### Allergy – “The modern plague”

The term ‘allergy’, meaning ‘changed reactivity’, was introduced in 1906 by Clemens von Pirquet [105]. Initially, allergy was equated with immunity. However, it was subsequently reserved for immune reactions that resulted in adverse symptoms, i.e., hypersensitivity. Allergy is today defined as immune-mediated hypersensitivity; in allergy, the immune system reacts to harmless substances, termed allergens, and this reaction produces symptoms. Different immune effector mechanisms are implicated in allergy, including IgE antibodies (so-called ‘atopic allergy’), T cells (contact dermatitis), and possibly also IgG antibodies.

The prevalence of IgE-mediated allergy, including asthma, atopic eczema, and hay fever, has increased dramatically in Western societies in parallel with a decrease in the prevalence of infectious diseases. The same pattern is seen for autoimmune diseases (such as multiple sclerosis) and inflammatory bowel (e.g., Crohn’s disease) (Figure 11) [106]. Today, one third of Swedish school children suffer from IgE-mediated allergy [107].

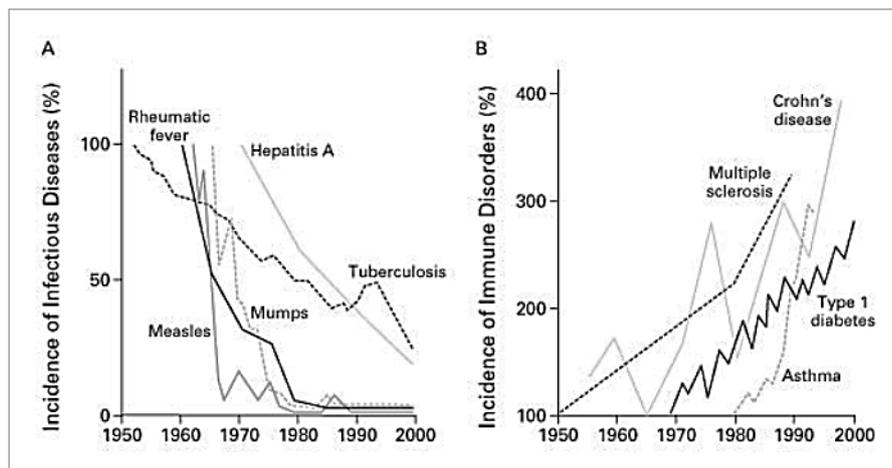


Figure 11. Inverse Relation between the Incidence of Prototypical Infectious Diseases (Panel A) and the Incidence of Immune Disorders (Panel B) from 1950 to 2000. Reproduced with permission from [106], Copyright Massachusetts Medical Society.

---

## **IgE-mediated allergy**

Allergy caused by IgE antibodies is termed atopic allergy (from the Greek *atopos*, meaning 'out of place').

In IgE-mediated allergy, the allergens are almost always proteins that are commonly present in pollens, food, and animal dander. These antibodies bind to tissue mast cells *via* Fcε receptors on the cell surface. The production of IgE antibodies directed against an allergen is termed 'sensitization'.

Allergy occurs when the sensitized individual re-encounters the same allergen at a later time-point, whereupon the allergen binds and cross-links at least two IgE antibodies and thus, the Fcε receptors on the mast cell. This induces mast cell degranulation and the release of histamine, which rapidly increases the permeability of the capillaries and venules. This early phase of allergy takes place within 10 minutes of the encounter with the allergen. Later, the activated mast cell produces and releases mast cell-specific leukotrienes (LTC<sub>4</sub>, LTD<sub>4</sub> and LTE<sub>4</sub>), which cause prolonged vascular leakage and airway smooth muscle constriction. Lastly, cytokines produced by mast cells, macrophages, and T cells at the allergenic focus synergize to attract eosinophils and more T cells to the site of allergen challenge, thereby contributing to tissue swelling and airway rigidity.

Depending on the affected organ, the involved symptoms/diseases include allergic conjunctivitis (eyes), allergic rhinitis (nose) asthma (bronchi), vomiting and diarrhea (gut), urticaria or eczema (skin). A systemic, often life-threatening allergic reaction is termed anaphylaxis [108]. There is no straightforward relationship between the route of exposure and the organ in which the symptoms appear. For example, food allergy denotes any allergic reaction to ingested food, including gastrointestinal symptoms (vomiting, diarrhea), skin manifestations (urticaria, eczema), and less commonly, bronchial constriction.

### The atopic march

Atopic allergy usually appears during childhood and early adolescence. The term 'atopic march' (also called 'allergic march') (Figure 12) refers to the sequential progression of the different manifestations of atopic allergy in susceptible children. This 'march' starts at infancy with atopic eczema and/or food allergy as the first manifestation(s). Asthma usually presents in early school-age, while allergic rhinitis is the last organ-specific manifestation of the atopic state, typically debuting in early adolescence. Asthma and allergic rhinitis usually persist throughout adult life, while food allergy and atopic eczema tend to vanish with time. However, this is not universally true and while food allergies to milk and eggs may disappear, allergies to fish and nuts may be life-long.

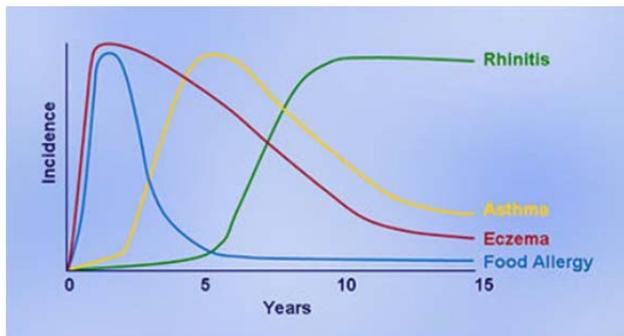


Figure 12. The atopic march. Reprinted with permission from *Childhood atopic eczema*; Copyright BMJ Publishing Group Ltd.

One-third of children with eczema develop asthma during later childhood [109] and two-thirds develop allergic rhinoconjunctivitis [110].

### Other types of allergy

Other types of allergy that do not involve IgE-mediated reactions are non-IgE-mediated food allergy and celiac disease (which appear to be T-cell-mediated, perhaps with a contribution from IgG) and contact dermatitis (T-cell-mediated). Non-IgE-mediated food allergy is most commonly to cow's milk and soya. Symptoms include diarrhea or constipation and pain. The symptoms may appear many hours and even days after consumption of the offending food, while IgE-mediated food allergy may present within minutes of food consumption.

---

## The hygiene hypothesis

Allergy is currently the most common chronic disease among children in the Western world. While the cause of allergy is unknown, epidemiologic studies point to reduced exposure to microbes of infants and young children as a risk factor. Thus, poverty, crowded housing, large families, early contact with pets or farm animals, and early exposure to foodborne microbes, are associated with a reduced risk of allergy development [111-113] (Figure 13). The "hygiene hypothesis" was proposed by Strachan in 1989 [114]. Strachan observed that small family sizes [114] and high parental social class [115] were both strongly and independently associated with high risk of hay fever at 16 years of age in two large population-based British cohorts. He proposed that some infectious agent(s) circulate(s) in poor and large families, and that exposure to this/these infectious agent(s) induced a type of immune maturation that counteracted the development of atopic disease. Twenty-five years after the formulation of the hygiene hypothesis, evidence has accumulated that an "unhygienic" lifestyle is protective against allergy development. The three major protective factors are: poverty; large family size; and animal contacts[111]. Growing up on a farm with milk production is particularly protective against allergy development [113].

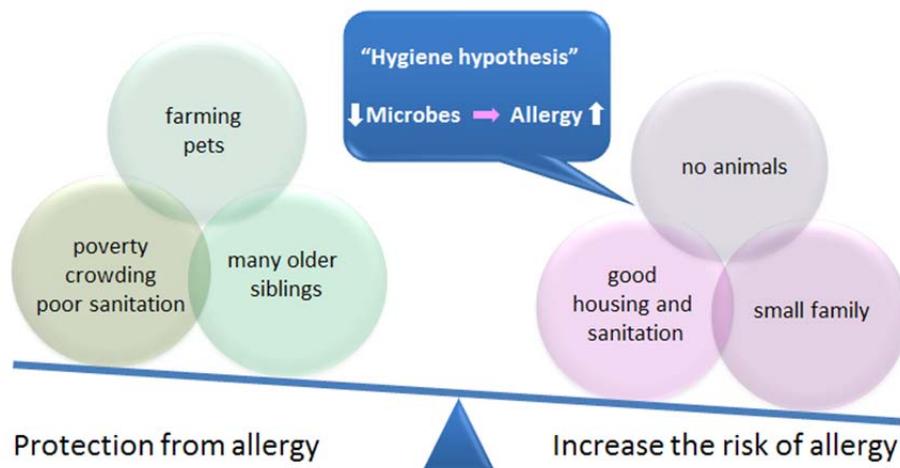


Figure 13. Lifestyle factors are associated with protection against allergy.

Birth by vaginal delivery, which exposes the neonate to the complex maternal microbiota, has been associated with a reduced risk of allergy development, as compared with Cesarean delivery, in some [116] but not all studies [117].

Factors with small or no effects on the risk of allergy development include gestational age at birth, breastfeeding or formula feeding and vaccinations.

### **Neonatal microbial exposure and the risk of allergy development**

Commensal microbial colonization is the major stimulus for the immune system of the infant and, thus, is likely to affect its development [88]. The acquisition of commensal gut bacteria is very rapid in infants born and living under poor sanitary conditions in the urban slum of Lahore, Pakistan, as compared with Swedish infants [118]. The turnover of individual strains of *E. coli* is also much higher in the former group than in the latter group. On average, an urban slum-dwelling infant in Pakistan carries 8.5 different *E. coli* strains over the first 6 months of life [119]. A study performed in the 1980's reported an average of 4.2 *E. coli* strains over a period of 11–18 months in the large bowels of Swedish infants [120], and in the Swedish FLORA study, our group has reported an average of 2.1 *E. coli* strains over 12 months in the bowels of infants who were born around year 2000 [121].

Many studies have investigated the relationship between the early gut colonization pattern and subsequent allergy development. The most consistent observation is that the presence of a gut microbiota of low complexity during the neonatal period is a risk factor for allergy development [122-125]. There is little evidence that any particular group of bacteria is particularly protective. A possible exception is *S. aureus*, a skin bacterium that has become part of the neonatal gut microbiota in recent years, probably as a consequence of limited competition from “traditional” gut bacteria that have become less prevalent in the gut microbiota in parallel with an increasingly hygienic lifestyle. *S. aureus* elaborates strongly immune-stimulating toxins, so-called ‘superantigens’,

---

which may offer broad stimulation of the immune system, equivalent to that conferred by hundreds of strains of 'normal' bacteria.

### **Oral flora and allergy development**

To the best of my knowledge, there have been no prospective studies on the relationship between early oral colonization by bacteria and allergy development. The prevalence of allergic disease in relation to antibody responses to oral pathogens was studied as part of the American National Health Program [126], antibodies were analyzed in serum samples from 9385 subjects aged  $\geq 12$  years. Individuals with high concentrations of IgG antibodies to *Porphyromonas gingivalis* had a significantly lower prevalence of physician-diagnosed asthma or hay fever, as well as self-reported wheeze (defined as a positive response to the question: "Apart from when you have a cold, does your chest ever sound wheezy or whistling"). Patient-reported wheeze was significantly negatively correlated with the concentrations of IgG antibodies against *Actinobacillus actinomycetemcomitans* [126]. In animal studies, oral administration of bacterial extracts protected mice from allergic airway inflammation *via* the recruitment of regulatory T cells to the airway [127]. Mice challenged with *P. gingivalis* showed reduced ovalbumin-induced allergic airway inflammation and airway responsiveness, concomitant with reduced levels of the Th2-type cytokines IL-4, IL-5, and IL-13 [128].

## Microbiota and inflammatory bowel diseases

Inflammatory bowel disease (IBD) comprises several diseases, of which ulcerative colitis and Crohn's disease are the most prevalent. Severe inflammation of the intestinal wall is the hallmark of these diseases. In ulcerative colitis, inflammation is confined to the colon and to the mucosal layer, while deeper parts of the gut wall are spared. In contrast, Crohn's disease can affect any part of the gastrointestinal tract and the resulting inflammation may span the entire gut wall. Furthermore, granuloma formation is a hallmark of Crohn's disease but is not seen in ulcerative colitis. Granulomas are aggregates of activated macrophages that are surrounded by scattered lymphocytes and fibrosis. Sometimes, the macrophages merge, forming multinucleated giant cells.

The etiology of IBD is unknown; interactions between the host immune system and an unbalanced intestinal microbiota are suggested to be of key importance [129]. The prevalence of IBD is highest in Northern Europe and North America at 20/100,000 persons. The incidence is increasing not only in traditionally high-incidence regions, but also in low-incidence areas, such as southern Europe and Asia.

One-fifth of all cases of IBD occur in childhood, and the incidence of pediatric IBD in Sweden has increased dramatically over the last decades [130]. Family history is a well-known risk factor for IBD development. Several IBD-associated genes have been identified, the most well-known being NOD2, which is a risk-associated gene for Crohn's disease [129]. NOD2 is the intracellular receptor for bacterial cell wall peptidoglycan, a fact that links IBD with immune reactions to the microbiota. However, genetic susceptibility alone does not explain the observed increases in the prevalence of IBD over the past decades. In addition, individuals who have migrated from a country with a low prevalence of IBD to a high-prevalence country develop the disease at the same rate as that observed in the country to which they have migrated [131, 132].

---

Similar to allergy, IBD has been linked to a Western lifestyle with good sanitary conditions [133, 134]. Numerous studies have investigated the microbiota in patients with IBD. In most of those studies, the investigated patients had already received treatment for their disease and may have had the disease for many years, in which case the observed differences may be secondary to the disease process and/or the administered medication. Investigations of the microbiota in pediatric cases of new-onset IBD are summarized in Table 7. In summary, the results of these 11 studies show that children with IBD have: 1) increased levels of *Enterobacteriaceae* and other facultative anaerobic bacteria; and 2) decreased level of Firmicutes, particularly *Clostridiales* (*Roseburia*, *Ruminococcus* and *Eubacterium* spp.). The results for Bacteroidetes and *Faecalibacterium prausnitzii* are inconsistent. A common finding is that the mucosa-associated colonic microflora of patients with IBD is less diverse than that of healthy individuals. These studies have all investigated the colonic or distal ileal microbiota.

Two reasons can be put forward to explain why an altered microbiota triggers IBD: 1) that the IBD microbiota contains a higher proportion of inflammatogenic bacteria, i.e., facultative Gram-negative bacteria that live in proximity to the mucosa and have the capacity to translocate the mucosal barrier; and 2) that bacteria with protective properties have been lost.

Regarding the first possibility, colitis does not occur in a germ-free environment in animal models of IBD [135], and CD4+ T cells activated by commensal bacterial antigens can transfer the disease [136]. Monocolonization by *Bacteroides fragilis* could initiate chronic inflammatory lesions in susceptible mice, while *E. coli* induced acute-type inflammatory reactions [137]. Regarding the second possibility, i.e., the potential loss of protective bacteria, butyrate is produced by bacteria in the Firmicutes phylum, mainly those belonging to clostridial clusters IV and XIVa [100], and decreased numbers of the members of these cluster have been found in patients with IBD [138]. This observation has led to suggestions that butyric acid-producing anaerobic bacteria could be used as a novel probiotic treatment for IBD [139].

Table 7. Gut microbiota in new-onset IBD

Age(year), subject & sample	Methods	Major finding	Ref
Mean age=12, feces CD=27,UC=16, Ctrl=IBS=30	16S rRNA PCR. Cloning & sequencing	Correlation to age Young CD ↑ E. coli & ↓ Bacteroidetes Older UC ↓ E. coli & ↑ Bacteroidetes	[140]
Age=16-60, feces CD=30, UC=33 Ctrl=IBS=34	16S PCR Cloning & sequencing	↑ Escherichia/Shigella in CD ↓ Faecalibacterium in CD vs. Ctrl ↓ Faecalibacterium in CD vs. UC	[141]
Age=8-13, colonoscopy biopsy, CD=12, UC=7, Ctrl=GI symptoms=7	Culture 16S rRNA RT-PCR	↑ Aerobic & facultative bacteria in IBD vs. Ctrl ↓ Bacteroides vulgatus in IBD vs. Ctrl	[142]
Age=12-73, colonoscopy biopsy, CD=20, UC=15 Ctrl=family history of colon cancer =15	DGGE. 16S Clone libraries.	DGGE band: UC≠CD Bacteria association: UC>CD(=Ctrl) Bacteroidetes: CD>UC	[143]
Age=11-18, colonoscopy biopsy, CD=15 Ctrl=GI symptoms=26	NGS 16S, V1-V3. Fungal 18S PCR. Mycobacterium PCR.	↓ Roseburia, Ruminococcus, Eubacterium, and Subdoligranulum in CD vs. Ctrl	[144]
Age=6-16, colonoscopy biopsy, CD=13, UC=12 Ctrl= GI symptoms=6	NGS 16S, V3-V6. RT-PCR	↑ Faecalibacterium prausnitzii in CD vs. Ctrl. ↓ Diversity in CD vs. Ctrl NO disease-specific PCA clustering	[145]
Age=6-16, colonoscopy biopsy, CD=29, UC=13 Ctrl= GI symptoms =42	Culture and PCR of microaerophilic Gram-negatives	No significant difference in microaerophilic Gram-negatives bacteria	[146]
Age=3-18, Ileum biosy CD=32, Ctrl= GI symptoms =36	Pseudomonas 16S PCR	↑ Pseudomonas in CD vs. Ctrl Pseudomonas species were less diverse in CD	[147]
Age=5-18, Colonic biopsy, CD=62, UC=26 Ctrl= GI symptoms =54	Mycobacterium avium subsp. Paratuberculosis PCR	↑ Mycobacterium avium subsp. paratuberculosis in CD vs. Ctrl	[148]
Age=8-16, colonoscopy biopsy, CD=9, UC=6, Ctrl=GI symptom =8, Healthy ctrl=16 (faecal)	End-point PCR. qPCR	↓ Bdellovibrio bacteriovorus in CD vs. Ctrl	[149]
Age=3-17, colonoscopy biopsy, CD= 447 Ctrl=IBS= 221	NGS 16S, V4. NGS shotgun and metagenomics	↑ Enterobacteriaceae, Pasteurellaceae, Veillonellaceae & Fusobacteriaceae in CD vs. Ctrl. ↓ Erysipelotrichales, Bacteroidales and Clostridiales in CD vs. Ctrl. Dysbiosis reflected in biopsy but weakly in stool, Antibiotic treatment =more dysbiosis	[150]

Abbreviations: CD= Crohn's disease, UC=Ulcerative colitis, Ctrl=Control, GI=Gastrointestinal, IBS=Irritable bowel disease, ↑=Increased level of, ↓=Decreased level of,

---

## AIM

The main purpose of this thesis was to study the importance of the commensal microbiota of children in health and immunoregulatory diseases, such as allergy and inflammatory bowel disease. To this end, we have established different bacterial identification methods and compared the suitability of these methods for characterization of the infant microbiota.

*Papers I and III - Whether or not the microbiota plays a role in immune disorders*

Paper I: To investigate whether different pacifier cleaning habits affect oral microbiota composition and allergy outcomes in infants.

Paper III: To compare the duodenal microbiota at the onset of pediatric IBD in patients with that of controls using high-throughput pyrosequencing and quantitative culturing, before medication and in the absence of any antibiotic treatment.

*Papers II and IV - Infant gut flora development and comparative analysis of the methods used for studying the microbiota*

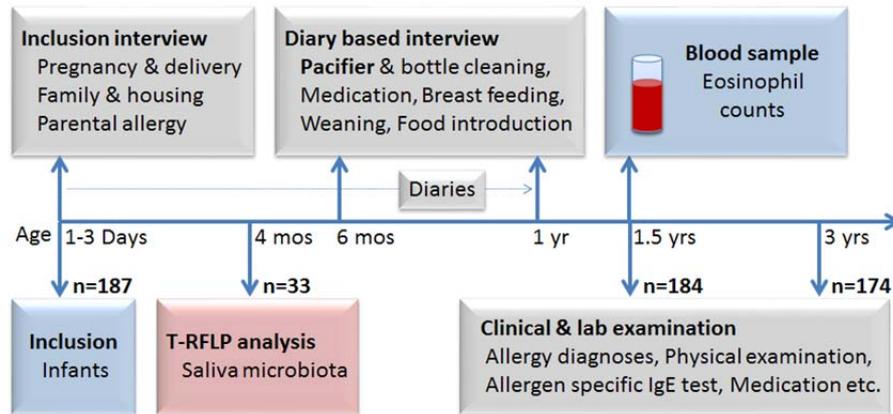
Paper II: To adapt T-RFLP to the characterization of infant fecal microbiota, including the building of a database for the identification of T-RFLP fragments. Furthermore, to compare T-RFLP with quantitative culturing for the analysis of the intestinal microbiota in infants with respect to sensitivity for detecting different groups of gut bacteria.

Paper IV: To compare pyrosequencing with T-RFLP and quantitative culturing for the detection of various groups of commensal bacteria in the intestinal microbiota of infants.

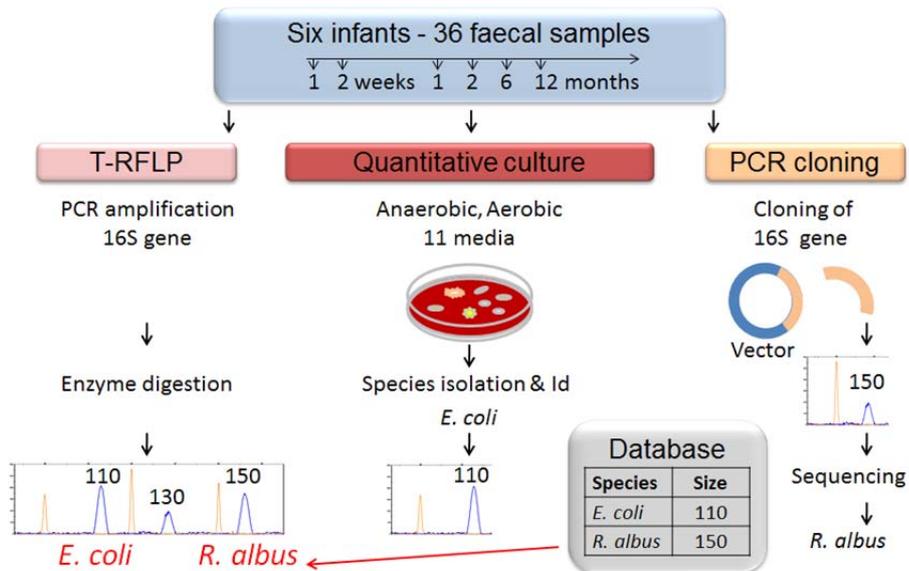
## MATERIAL AND METHODS

### General overview of the four papers

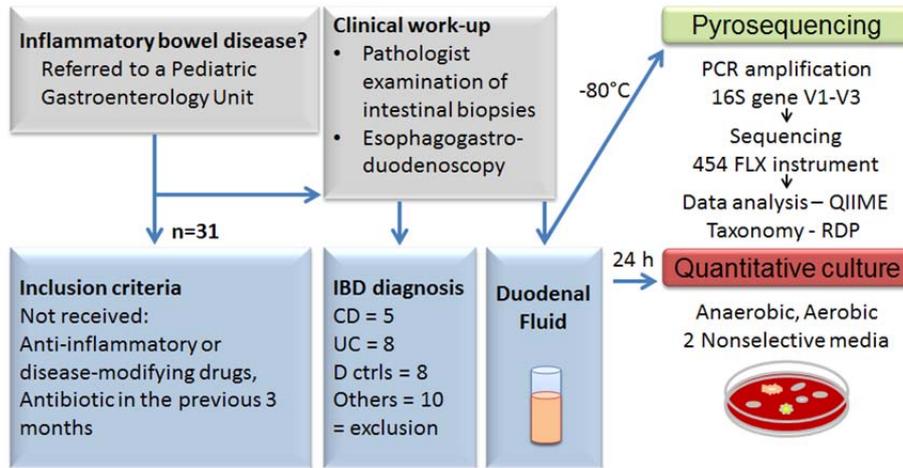
#### I. Pacifier Cleaning Practices and Risk of Allergy Development



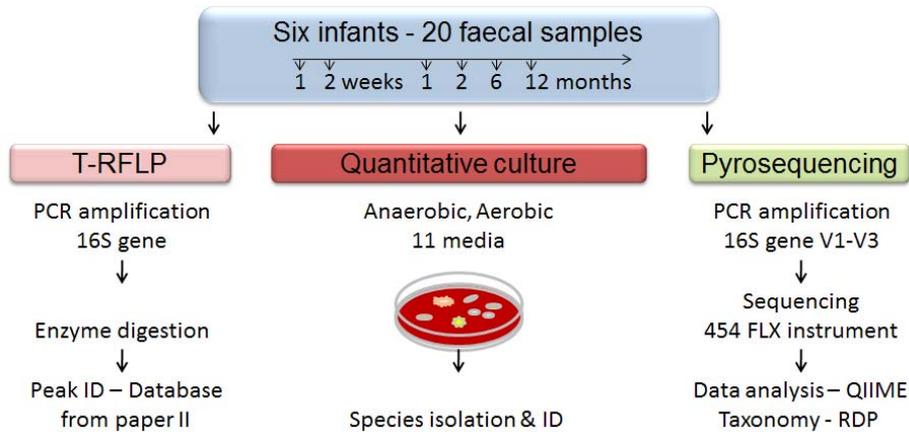
#### II. Comparison between terminal-restriction fragment length polymorphism (T-RFLP) and quantitative culture for analysis of infants' gut microbiota



### III. Altered composition of duodenal microbiota of children with newly diagnosed inflammatory bowel disease



### IV. Comparative analysis of infants' gut microbiota by next generation sequencing and quantitative culture



## Children and Samples

Detailed descriptions of the materials and methods used are given in each paper, and only those of particular relevance to the results are described below.

### The AllergyFlora birth cohort (Papers I, II and IV)

Pregnant women were recruited into the AllergyFlora study at Mölndal Hospital, in the greater Gothenburg area of South-West Sweden. A total of 187 children born in the period 1998–2002 were enrolled at 1–3 days of age to examine the relationship between the early intestinal colonization pattern and subsequent allergy development. In this cohort, families with at least one allergic parent were mainly targeted, so as to achieve a high proportion of children with allergic disease. Upon inclusion, a structured interview was conducted, which focused on the pregnancy and delivery, and on the family structure and housing conditions. Diaries covering the infant's first year of life were kept by the parents, who were asked to record food introduction, weaning, diseases and medications, and other significant events. This information was reported during structured telephone interviews with the parents when the children were 6 months old. During this interview, the parents were also asked: "Does the child use a pacifier?" and "Is it cleaned by boiling, rinsing in tap water, or by the parents sucking on it?" (more than one option could be selected). The exclusion criteria were preterm birth (<38 weeks of gestation), neonatal intensive care, and presumed inability to comply with a quite demanding study protocol, e.g., difficulties in understanding the Swedish language.

Fecal samples were collected from the 184 infants. From infant number 117 onwards, a saliva sample was obtained at 4 months of age and blood samples were obtained at 4, 18, and 36 months of age. The aim was to monitor the immune maturation of the infants in relation to the intestinal colonization pattern and allergy development.

A pediatric allergologist examined the children, reviewed their medical charts, and performed a structured interview with the

---

parents when the children were 18 and 36 months old, and whenever symptoms suggestive of the commencement of allergy were noted. Venous blood was analysed for eosinophilic granulocyte counts, which increase in patients with allergies, and for allergen-specific IgE (“sensitization”, see below).

For clinical diagnosis, the examining pediatrician used the following criteria:

*Eczema:* Diagnosed according to Williams criteria for “atopic dermatitis”. “Eczema at 18 months” denoted diagnosis at any time before or at 18 months of age; “Eczema at 36 months” required the presence of symptoms after 24 months of age.

*Asthma:* Persistent wheezing for  $\geq 4$  weeks or  $\geq 3$  episodes of wheezing combined with  $\geq 1$  minor criterion (symptoms between colds, eczema, allergic rhinoconjunctivitis or food allergy). For “asthma at 36 months”,  $\geq 1$  wheezing episode had to have occurred after 24 months of age, and responses to inhaled glucocorticoids or leukotriene antagonists were included among the minor criteria.

*Sensitization:* Presence of specific IgE against inhalant allergens (birch, timothy grass, mugwort, cat, dog, horse, *D. pteronyssinus*, *D. farinae*, and mold; Phadiatop®; Phadia AB, Uppsala, Sweden) or against food antigens (milk, egg, soy, fish, wheat, and peanut; ImmunoCAP® food-mix test; Phadia AB). A positive reaction was defined when the sum of the allergen-specific IgEs was  $\geq 0.35$  kU/L in either assay.

*Parental history of allergy:* Physician diagnosis of asthma, allergic rhinoconjunctivitis or eczema, as reported in the interviews.

In Paper I, allergic manifestation at 18 months of age was related to the parents’ habits of cleaning their baby’s pacifiers in the ALLERGYFLORA cohort. Saliva samples collected at 4 months of age were available for 33 children.

In Papers II and IV, the fecal samples from six infants were analyzed in each study. Infants were selected from the

ALLERGYFLORA cohort study. Four infants were analyzed in both Paper I and Paper II. Samples were collected by the parents on six occasions during the first year of life (at 1, 2, and 4 weeks and at 2, 6, and 12 months of age). In Paper I, all 36 samples from the six infants were analyzed. In Paper IV, a total of 20 samples were analyzed, due to the fact that for four infants, only the 1-week and 1-year samples were analyzed.

### **Inflammatory bowel disease study (Paper III)**

The study was designed to characterize the microbiota of children with IBD as early as possible, i.e., before diagnosis and initiation of treatment. Children who were referred for suspected IBD to the Gastroenterology Unit at the University Hospital in Gothenburg, Sweden, were included. The exclusion criteria were any prior treatment with anti-inflammatory, disease-modifying drugs or antibiotic treatment in the previous 3 months. The children were hospitalized for clinical work-up and their intestinal biopsies were examined by an experienced pathologist. Diagnosis of the patients was based on clinical presentation, endoscopic and histologic features, and radiologic appearance according to the Porto criteria. The following diagnostic groups were included: Ulcerative colitis (N=8); Crohn's disease (N=5); and Controls (N=8). The latter group represented those children in whom the intestinal biopsy showed no signs of inflammation; their symptoms were deemed to be caused by a functional intestinal disorder. Symptomatic children with the following diagnoses were not included in the study: allergic colitis (N=2); celiac disease (N=2); infectious colitis (N=1); unspecified colitis (N=4); and indeterminate colitis (N=1).

Duodenal contents were sampled in connection with gastroscopy. The duodenal fluid was immediately transported to the laboratory for quantitative culturing (see below) and subsequent assessment by pyrosequencing (see below).

---

## Methods

### Quantitative culture (Papers II, III, and IV)

Samples (of feces or duodenal fluids) were transported in a gas-tight sachet in which an anaerobic milieu was created using an anaerobic generator (Benex limited, County Clare, Ireland). Anaerobiosis was confirmed upon arrival to the laboratory by checking the anaerobic indicator. Culturing was performed within 24 h. The detection limit was 300 ( $10^{2.5}$ ) colony-forming units (CFU)/g for feces and 33 ( $10^{1.52}$ ) CFU/mL for duodenal contents.

In brief, a calibrated spoonful of feces was serially diluted in ten-fold steps in sterile peptone water and cultivated on two non-selective media and eight selective media (Table 1). For duodenal fluids, 100  $\mu$ l were serially diluted and cultivated on non-selective media aerobically (Columbia blood agar) and anaerobically (Brucella blood agar). Bacterial colonies with different morphologies were enumerated on appropriate media, subcultured for purity, and speciated using a combination of biochemical tests and genus- or species-specific PCR, as previously described [49]. Pure-culture isolates were stored frozen at -80°C. Anaerobic spore formers (“clostridia”) were quantified by anaerobic culture of alcohol-treated samples. Alcohol treatment kills vegetative cells, whereas bacterial spores survive. Thus, spore-forming bacteria can be detected by this method. Spore-forming bacteria include *Bacillus*, which are aerobic, and *Clostridium*, which are anaerobic. The group “Clostridia”, defined by culture does not conform to the genus Clostridia, as defined by DNA-based taxonomy. The latter comprises 19 groups (I–XIX), which include both spore-formers and non-spore-forming bacteria. A portion of a 1:10 dilution of stool sample was mixed with an equal volume of 99% ethanol on a shaker at room temperature for 30 min. The sample was diluted and plated on agar.

Table 1. Culture conditions and species identification methods

	Culture condition, agar & duration		Species identification	
Facultative bacteria	Aerobically at 37°C	Days	Morphology	Bacterial typing
<i>Enterobacteriaceae</i>	Drigalski	1	G- bacilli	API20E <sup>a</sup>
<i>Staphylococcus</i>	Staphylococcus	2	G+ cocci	Catalase+, Coagulase±
<i>Enterococcus</i>	Enterococcosel	1	G+ cocci	Esculine hydrolysis
Non-selective	Colombia blood	1		Total CFU counts
Anaerobic bacteria	Anaerobically at 37°C			
<i>Bacteroides</i>	Bacteroides bile esculin	3	G- bacilli	Rapid ID 32A <sup>a</sup>
<i>Clostridium</i> (sporeforming)	Brucella blood, alcohol treatment	3	G+ bacilli	Rapid ID 32A <sup>a</sup>
<i>Clostridium difficile</i>	CCFA <sup>b</sup>	3	G+, G± bacilli	Rapid ID 32A <sup>a</sup>
<i>Bifidobacterium</i>	Beerens	3	G+ bacilli	Species-specific PCR
<i>Lactobacillus</i>	Rogosa	3	G+ bacilli	Species-specific PCR
Non-selective	Brucella blood	3		Total CFU counts <sup>c</sup>

<sup>a</sup>Biotyping, bioMérieux.

<sup>b</sup>Cycloserine cefoxitin fructose egg yolk.

<sup>c</sup>Facultative bacteria growing on the plate were subtracted

G+=Gram-positiv, G-=Gram-negative, G±=Gram-variable

---

## **DNA-based methods (Papers I-IV)**

### **DNA extraction for T-RFLP and pyrosequencing**

Bacterial DNA was extracted from 200 µl of thawed duodenal fluid or 160 mg (wet weight) of frozen feces using the QIAamp DNA Stool Mini Kit (Qiagen, Hildens, Germany). The DNA yield was increased by adding additional steps [151], in which four glass beads (3.0 mm in diameter) and 0.5 g of zirconia beads (0.1 mm in diameter) were added to the sample suspended in ASL buffer. The suspension was homogenized at 5 m/sec twice for 30 sec using a FastPrep Cell disrupter (Savant Instruments, NY, USA), incubated at 95°C for 5 min, and shaken at 1,200 rpm (IKA Vibrax Shaker; Labortechnik, Germany) for 30 min at 5°C. DNA was quantified spectrophotometrically (NanoDrop Technologies, Wilmington, DE, USA).

### **T-RFLP (Papers I, II, and IV)**

Terminal-Restriction Fragment Length Polymorphism (T-RFLP) was performed on saliva samples and fecal samples.

### ***PCR amplification of bacterial 16S rRNA genes***

Bacterial 16S rRNA genes were amplified using the universal primers ENV1 (5'-6-FAM-AGA GTT TGA TII TGG CTC AG-3', *E. coli* nt 8-27) and ENV 2 (5'-CGG ITA CCT TGT TAC GAC TT-3', *E. coli* nt 1511-1492) [152]. The forward primer was labeled with fluorescent 6-FAM at the 5'-end. The PCR mixture contained 50 ng DNA (25 ng and 75 ng were also tested; 25 ng DNA generated smaller T-RF peaks), 25 µl of Hot Start Taq Master Mix (Qiagen), 0.3 µM of primer, 1 mM MgCl<sub>2</sub>, and water added to a final volume of 50 µl. For pure-culture strains, the final volume was 25 µl and the reagent concentrations were 50% of those used for the fecal samples. The PCR was performed in an Eppendorf Mastercycler gradient (Eppendorf, Hamburg, Germany) using the following program: an initial activation of Taq polymerase at 95°C for 15 min, 25 cycles (28 and 32 cycles were also tested) of 94°C for 1 min, 50°C for 45 s, and 72°C for 2 min, with a final extension at 72°C for 7 min. Negative controls that contained all the DNA extraction reagents but that lacked bacteria were included

throughout the whole analysis, to detect DNA contamination from the reagents or instruments. For each sample, three PCRs were run in parallel.

### ***Restriction enzyme digestion***

Purified PCR products from fecal samples (100 ng) were digested with 16 U (4, 10, and 20 U were also tested) of the *MspI* endonuclease (New England Biolabs, Ipswich, USA) at 37°C for 5 h in a final volume of 5 µl. For pure-culture bacteria, 30 ng of PCR products were digested with 16 U *MspI* in a final volume of 10 µl. The reaction was stopped by heating at 65°C for 20 min.

### ***Fragment analysis***

Fluorescently labeled T-RFs were separated and detected using the ABI PRISM 310 genetic analyzer (Applied Biosystems, California, USA) with the Pop 4 gel. The resolution (number and quality of peaks) was compared using different volumes of digested PCR amplicons (0.5, 0.7, 1, 1.5, 2.5, 3.8, and 4.8 µl), run voltages (9.9, 12.2, and 15 kV), injection times (5 s and 10 s) and size standards (ROX 500, ROX1000, combined ROX 500 and ROX1000, LIZ 600, and LIZ 1200; Applied Biosystems). Reproducibility was investigated by running samples in duplicate.

The fragment lengths were analyzed using GeneMapper ver. 4 (Applied Biosystems) and the Local Southern Method. 'True' peaks were separated from 'noise' within a fragment length range of 28–1000 bp, according to Abdo et al. [153]. Briefly, the data were standardized by dividing the area of each peak by the total peak area of the particular sample. The standard deviation (SD) of the dataset was then computed assuming that the true mean was zero. Peaks with an area three standard deviations greater than the mean were considered to be true signals. They were collected and removed; thereafter, the process was iterated until no more 'true' peaks were identified. T-RFs that differed by no more than ±1bp in different analyses were considered to be identical.

All the fecal samples were analyzed in three T-RFLP analyses. Only those fragments that were represented in two out of three runs were considered, and their peak sizes and areas were averaged.

---

### **Pyrosequencing (Papers III and IV)**

DNA extractions of the fecal samples and duodenal fluid samples were pyrosequenced on a 454 Life Sciences Genome Sequencer FLX instrument (GATC Biotech AG, Konstanz, Germany). A proofreading polymerase was used for PCR amplification of the V1-V3 region of the bacterial 16S rRNA genes using the forward primer 27F (5'-AGA GTT TGA TCC TGG CTC AG-3') and the reverse primer 534R (5'-ATT ACC GCG GCT GCT GG-3'). The PCR program used was as follows: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 56°C for 30 s, and 72°C for 10 s, and a final extension at 72°C for 1 min. The PCR products were column-purified. A second PCR was performed to add the sequencing adaptors and multiplex identifiers. Multiplex identifiers are short DNA sequences that are used to identify the different samples. The conditions for the second PCR were the same as those described above, except that steps 2–4 were only performed for five cycles. The PCR products were column-purified, quantified using the Qubit system (Invitrogen, CA, USA), and all the samples were pooled. Prior to sequencing on the GS FLX, the pool of amplicons was gel-purified and the DNA was recovered with a standard PCR and gel purification kit. After quantification with the Qubit system, the library was sequenced on the Roche GS FLX following the standard protocols from Roche.

### ***Sequence analysis***

The raw sequence reads were analyzed using the software package Quantitative Insights into Microbial Ecology (QIIME), employing default parameters for each step [154], except where specified. The raw reads were assigned to samples according to their barcodes. The singletons and possible chimeric sequences were removed before the subsequent analysis. Only high quality reads that had a minimum quality score of 25 and a read length of 200–500 bases were retained for further analysis.

Operational Taxonomic Units (OTUs) were chosen based on repetitive sequences derived by sequence clustering using the UCLUST software [155], with a minimum pairwise identity of 97%. The representative sequence (reads of the same DNA sequence) of each OTU was chosen based on abundance and length, and aligned

using PyNAST (a python-based implementation of NAST in QIIME). The taxonomy of each representative OTU was subsequently assigned using the Ribosomal Database Project (RDP) classifier nomenclature method [63]. Finally, an abundance table for each OTU was generated from the frequency read count.

### **Statistical analysis**

Orthogonal Partial Least Squares (OPLS, SIMCA P+ V13; Umetrics AB, Umeå, Sweden) was used to obtain the overall pattern of the microbiota. OPLS is a prediction and regression variety of principal component analysis which finds information in the X data that is related to Y variables. The presence or absence of each OTU was used as the X-variables, rather than the number of reads per OTU. Y represents a diagnostic group, e.g., children with ulcerative colitis or Crohn's disease, the diseased controls in Paper III, or the children whose parents "clean" their pacifier by sucking on it vis-à-vis children whose parents did not have this habit (Paper I).

The T-RFLP analysis of the fecal samples from six infants collected over the first year of life was depicted by principal component analysis.

Univariate analyses were performed to confirm the unequal distribution of the OTUs that showed the strongest contributions in the respective OPLS models used to compare the diagnostic groups. Univariate analyses were performed by using Fisher's exact tests (SPSS; IBM Corp., NY, USA) in Paper III.

One-way ANOVA was performed to compare the overall diversity at the phylum and genus levels between the diagnostic groups in Paper III (Graphpad Prism 5; San Diego, USA).

---

## RESULTS AND COMMENTS

### **Pacifier cleaning practices and allergy development in infants**

In Paper I, the effects of pacifier cleaning practices on the composition of the infant oral microbiota and allergy development were studied. In total, 184 children in the ALLERGYFLORA birth cohort were studied. Overall, 80% of the children had at least one allergic parent, and these infants were therefore considered "high-risk" with respect to allergy development. Almost all the infants were breast-fed for at least 4 months, and 74% (N=136) used a pacifier during their first 6 months of life.

All the infants were examined for sensitization (IgE antibodies to common allergens) and clinical allergy at 18 and 36 months of age. At 18 months of age, 25% had eczema and 5% had asthma. Sensitization to food antigens occurred in 15% of the infants, while sensitization to inhaled antigens was uncommon (see Table 3, Paper I).

When the child was 6 months old, the parents participated in a telephone interview that was designed to find out how they cleaned the pacifier used by the child. Several alternative responses were allowed.

As shown in Table 1, almost all the parents cleaned the pacifier by rinsing it in tap water, although approximately 50% of the parents also boiled the pacifier and almost 50% reported 'cleaning' the pacifier by sucking on it before giving it back to the infant.

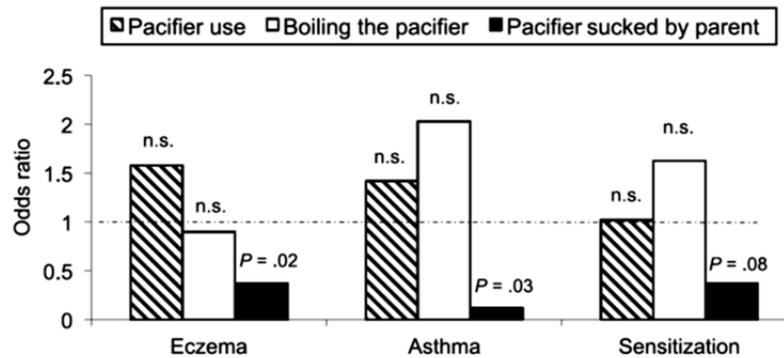
**Table 1:** Pacifier cleaning practices among the 136 infants who used a pacifier during the first 6 months of life.

Technique	No. of children	%
Rinsing in tap water	113	83
Boiling	74	54
Parental sucking of the pacifier	65	48
Not reported	13	10
Total	136	100

Techniques for cleaning the pacifier reported in a telephone interview when the child was 6 months of age.

### ***Pacifier cleaning practices and allergy at 18 months of age***

The presence of eczema, asthma or atopy at 18 months of age was related to pacifier use and pacifier cleaning practices. Figure 1 shows the risks of developing eczema, asthma, and sensitization. An Odds Ratio (OR) of 1 denotes that the risk is as high as that for infants who did not use a pacifier. Pacifier use *per se* was not significantly associated with clinical allergy or sensitization (Figure 1). Boiling the pacifier was associated with an increased prevalence of asthma, although the effect was not significant. However, parental ‘cleaning’ of the pacifier by sucking on it was strongly associated with reduced risks of eczema and asthma.



**Figure 1:** The relationship between atopy or allergy in children at 18 months of age and in relation to pacifier use and different pacifier cleaning practices. Odds Ratios are given for pacifier vs no pacifier use. For children using a pacifier the



of respiratory infections in the first 6 months of life did not differ between children whose parents sucked their pacifier (1.4) and other pacifier-using children (1.5; ANOVA,  $p=0.75$ ).

### ***Analysis of confounding factors***

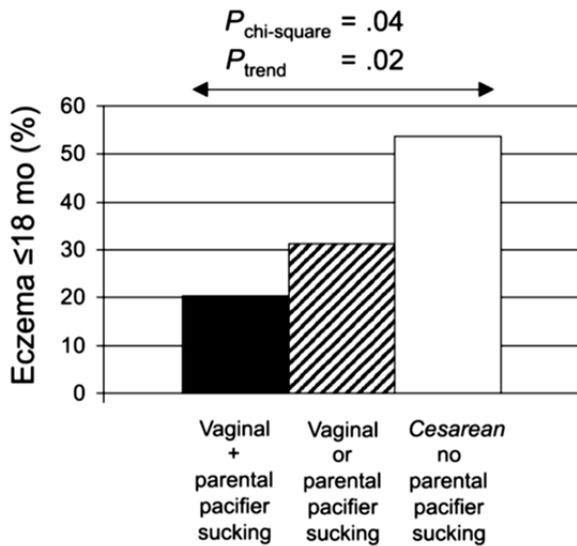
Confounders, such as parental education or delivery mode, could be responsible for the observed associations between pacifier cleaning practice and allergy development. The parents' habits of cleaning the baby's pacifier were analyzed in relation to a range of lifestyle factors. Parents of vaginally delivered infants were significantly more likely to suck the child's pacifier than parents of babies delivered by Cesarean section (Fisher's exact test,  $p=0.02$ ). Conversely, mothers with a University level degree tended to be less likely to suck the child's pacifier, although this association was not significant (see Table 4, Paper I).

Logistic regression analysis was performed to adjust for delivery mode and mother's occupation. The protective effect of parental pacifier sucking on eczema development during the child's first 18 months of life remained (OR 0.27, 95% confidence interval 0.086–0.819;  $p=0.02$ ). The protective effect on asthma could not be analysed in this way due to the low number of cases.

### ***Delivery mode and parental pacifier cleaning***

Delivery *via* the vaginal route exposes the infant to the maternal vaginal microbiota, while parent-infant pacifier-sharing exposes the child to the parental oral microbiota. We examined whether these exposures conferred additive protective effects by stratification of the cohort into three groups: (1) vaginally delivered infants with pacifier-sucking parents; (2) Cesarean-delivered infants whose parents did not suck their pacifiers; and (3) infants who were either vaginally delivered or exposed to the parents' oral microbiota by pacifier sharing. The prevalence of eczema in infants at 18 months of age was compared between the three groups (Figure 3). The group that was exposed to both the maternal vaginal and parental oral microbiota had the lowest prevalence of eczema (20%), whereas the infants who were exposed to neither the maternal vaginal nor parental oral microbiota had the highest prevalence of eczema (54%).

The group of children who were either vaginally delivered or whose parents sucked their pacifiers had an intermediate prevalence of eczema (31%) (Figure 3).



**Figure 3:** Prevalences of eczema at 18 months of age in the cohort stratified as: (1) vaginally delivered infants with pacifier-sucking parents (black bar, N=59); (2) infants who were either vaginally delivered or exposed to the parents' oral microbiota by pacifier sharing (hatched bar, N=48); and (3) Cesarean-delivered infants whose parents did not suck on their pacifier (open bar, N=13)

### ***Parental pacifier sucking and allergy during the child's first 36 months of life***

The children in the cohort were also followed up at 36 months of age. Kaplan–Meier curves were calculated for eczema, asthma, and sensitization (see Figure 4 Paper I). Development of eczema up to 36 months of age was significantly less likely in those infants whose parents sucked on their pacifiers during their first 6 months of life, as compared with other pacifier-using children ( $p=0.04$ ). The protective effect of parental pacifier sucking on asthma and sensitization was not statistically significant.

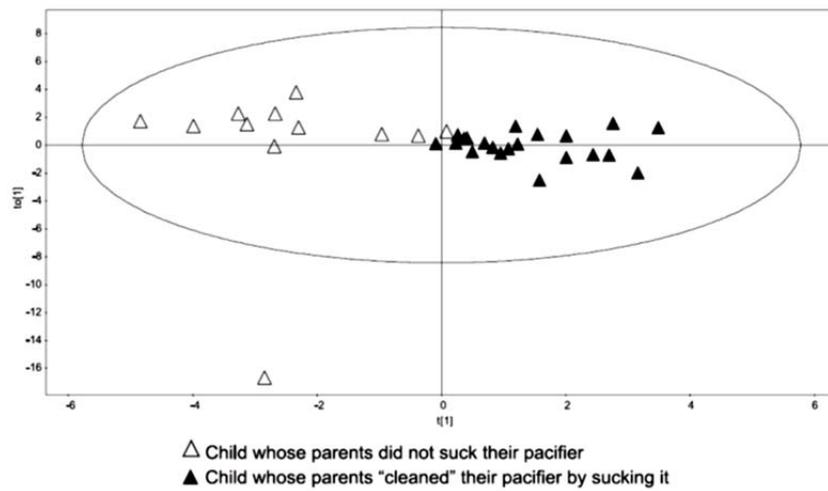
The children in the cohort also underwent clinical examinations for allergy at 8 years of age, and the results have recently been summarized. The preliminary results suggest that infants whose parents "cleaned" their pacifier by sucking on it during the child's first 6 months of life have a significantly lower prevalence of asthma and reduced bronchial hyperreactivity (measured by the metacholine test) than infants who used a pacifier that was not "cleaned" in this way (Bill Hesselmar, personal communication).

***Parental pacifier sucking and salivary microbiota of children at 4 months of age***

If the parents' salivary microbes are transferred to the infant *via* the pacifier, this may affect the development of the infant's salivary microbiota.

Salivary samples had been obtained from 64 of the infants at 4 months of age and had been stored frozen. DNA was extracted from these samples and analyzed by T-RFLP to obtain an overall picture of the microbiota. Salivary samples from Cesarean-delivered infants were excluded, since it is known that delivery mode affects oral microbiota composition [156]and, furthermore, delivery mode and pacifier cleaning mode were related in our cohort (see above). After exclusion of Cesarean-delivered infants and non-pacifier users, 33 infants remained, 21 of whom had parents who sucked on their pacifier.

Figure 4 shows an O-PLS (orthogonal projection onto latent structures) analysis of the T-RFLP results. O-PLS is a regression variety of principal component analysis that can be used to separate two groups (in this case, pacifier-using infants whose parents did, or did not, "clean" the pacifier by sucking on it) based on a very large set of variables. The presence or absence of each T-RF (usually representing a bacterial genus, see Paper II) was used as the set of X-variables. In all, 845 T-RFs were identified in the group of infants. As seen in the figure, the two groups of infants could be separated based on the salivary T-RFLP patterns. One infant (outlier, bottom left) had a microbiota that was clearly distinct from the others. As identification of T-RFLP peaks requires a substantial amount of additional work, we do not, at present, know the identity of the microbes that are more or less prevalent in the saliva samples of the two groups of infants.



**Figure 4:** The oral microbiota of children at 4 months of age in relation to parents' habit of cleaning the pacifier. Each symbol represents one infant and its position is based on the totality of bacterial taxa (genus/species) in the saliva, as generated by T-RFLP (digestion of total bacterial DNA and separation of the fragments). The data were analyzed by orthogonal projection to latent structures, which is a variation of principal component regression analysis, with parental pacifier sucking as the Y-variable, and the totality of bacterial taxa as the X-variables. The first component, representing the largest variations among the X-variables, is represented along the horizontal axis. The second component, depicted along the vertical axis, represents the second largest variations among the X-variables.

## Development of a database for the identification of T-RFs

In theory, it should be possible to deduce the identity of a bacterium that generates a fragment of a certain size, i.e., a T-RF in the T-RFLP method, since the restriction sites are known and it is possible to calculate the sizes of the fragments generated after digestion of the PCR products from amplification of the 16S rRNA gene of that bacterial taxon.

However, this has proven to be practically impossible. We noted that the relationship between fragment size and migration time was not linear (see Figure supplement S1 in Paper II), and this is also evident from Table 2, in which the theoretical "in silico" digestion results are compared with the observed apparent sizes. The larger the fragment, the greater the deviation, which makes it impossible to introduce any automatic correction factor that would make it possible to "translate" directly the migration time to T-RF size.

**Table 2:** Type-strain *in silico* digestion and T-RFLP analysis for comparison of the fragment size discrepancies between true and observed T-RFs.

Type strain		<i>In silico</i>	Observed	Discrepancy
		T-RF size		(nucleotide)
<i>Lactobacillus vaginalis</i>	NCTC 12197T	29	30	1
<i>Bacteroides uniformis</i>	ATCC 8492T	97	91	-6
<i>Clostridium lituseburense</i>	NCIMB 10716T	195	188	-7
<i>Clostridium ramosum</i>	ATCC 25582T	295	286	-9
<i>Clostridium difficile</i>	NCTC 12197T	457	448	-9
<i>Citrobacter braakii</i>	CDC 80-58T	496	485	-11
<i>Lactobacillus fermentum</i>	ATCC 25582T	581	570	-11

Furthermore, to judge the sizes of the T-RFs, a molecular size standard has to be run in parallel. Several molecular weight (MW) standards were tried, and Table 3 shows that, depending on the

---

MW standard, the T-RFs generated by digestion of two pure-culture bacterial strains (*Bacteroides* and *E. coli*) and a long PCR fragment had different apparent sizes.

**Table 3:** T-RFLP results for a sample with three fragments run in combination with different size standards.

Internal size standard used	Apparent T-RF size of the bacterial DNA (nucleotides)		
Rox 500	93	487	0
Rox 1000	93	486	858
Rox 500 & 1000	94	487	857
Liz 1200	92	484	878

These results clearly demonstrate that it is in practice impossible to identify the species/genus based on *in silico* digestion. Instead, the actual apparent size (in relative to the chosen MW standard) must be investigated in real life, using pure-culture isolates that are run, one at a time, and subsequent examination of the generated T-RFs.

To construct a database, we analyzed a large set of pure-culture strains by T-RFLP. A distinct advantage of this exercise was that it allowed identification of the conditions that generate a single T-RF after digestion. A problem with the method is the generation of "pseudofragments", i.e., fragments that appear if a cleavage site is not used by the restriction enzyme, generating a longer than expected fragment. If too little restriction enzyme is used, this can pose a significant problem.

Despite these precautions, we observed pseudofragments on some occasions. For samples in which *Bacteroides* were present in very high numbers, fragments were visualized that could not be attributed to any taxon, and we concluded that they were pseudofragments created by incomplete digestion (Paper II).

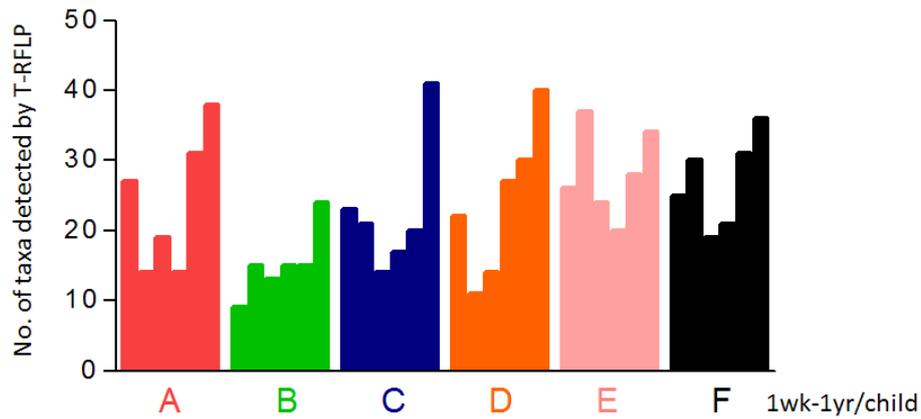
The database was developed by using pure-culture isolates from all the groups of gut bacteria that were isolated by culture the microbiota of the ALLERGYFLORA cohort were used, complemented by strains obtained from the culture collection of University of Gothenburg (CCUG) strain collection (42). Lastly, we performed cloning and sequencing on 9 fecal samples, which were found to contain a high number of T-RFs that could not be identified.

Using this method, a database that consisted of 130 species was constructed, which enabled identification of the bacteria based on the apparent sizes of their T-RFs. Through the use of this database, approximately 60% of the peaks found in infant fecal samples could be determined. T-RFLP identified bacteria to the genus level and also provided some sub-genus discrimination.

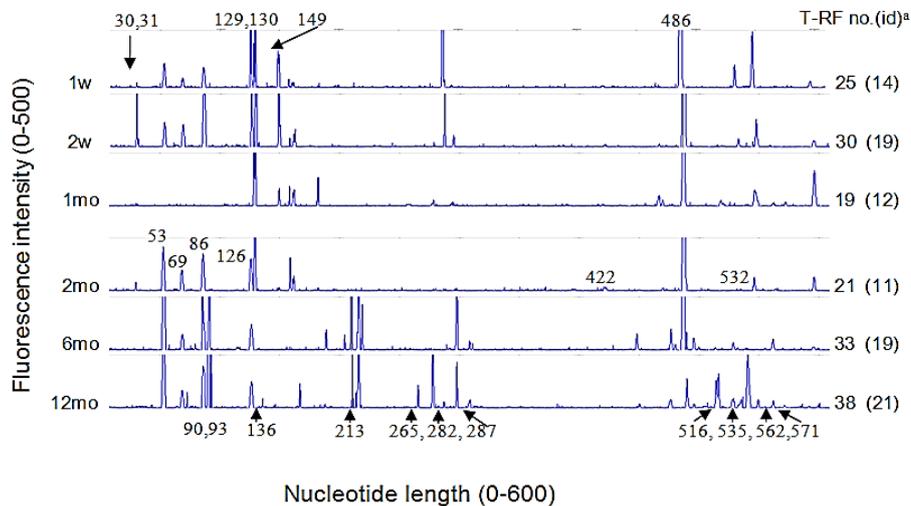
## **Infant microbiota development**

In Paper II, longitudinally sampled fecal samples from six infants were analyzed by T-RFLP. It is well known that infants acquire successively more and more bacterial species, and that the complexity of the microbiota increases with time [50]. However, when we analyzed the early infant faecal samples using T-RFLP, an interesting picture emerged. In most of the infants (Figure 5), we observed rather high bacterial diversity, as characterized in the fecal samples at 1 week and 2 weeks of age, while microbiota complexity decreased thereafter, to reach a nadir at around 1 month of age. Thereafter, complexity increased markedly once again, with the highest level of complexity detected in the last sample collected at 12 months of age. This pattern was also seen in the two infant analyzed with pyrosequencing (Paper IV).

The high microbial complexity of the samples obtained at 1 and 2 weeks of age was unexpected. Figure 6 shows an example of the T-RFLP pattern of a single infant who was followed using repeated sampling of the feces. As shown in the figure, several of the T-RFs that were found in the early fecal samples were not present in the microbiota after the first weeks of life and did not reappear subsequently.

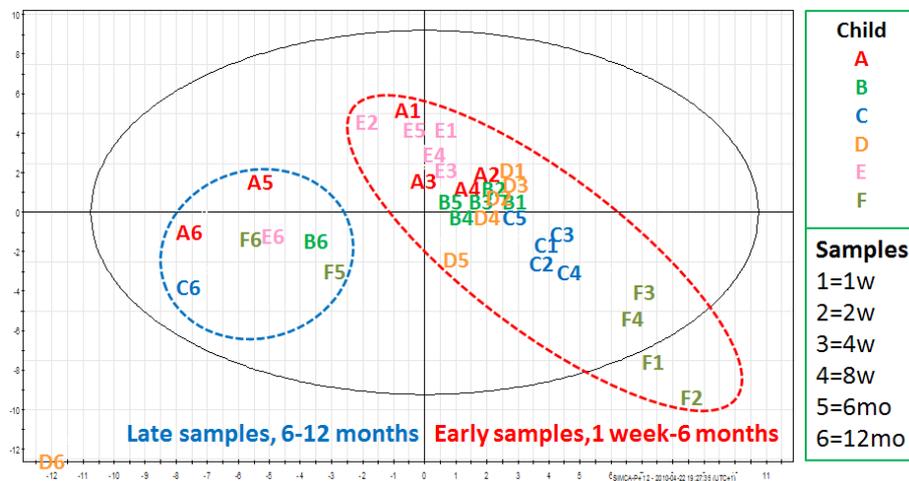


**Figure 5.** Microbiota complexity as a function of age in six children. The number of taxa detected in each sample is depicted by a bar; each infant is represented by six bars, one for each sampling occasion (at 1, 2 and 4 weeks and at 2, 6 and 12 months of age).



**Figure 6:** T-RFLP analysis of faecal microbiota composition in an infant followed during the first year of life. Bacterial DNA was extracted from faecal samples obtained at 1 and 2 weeks of age and at 1, 2, 6 and 12 months of age and analyzed by T-RFLP. Each peak represents a terminal restriction fragment of a specific length that corresponds to a bacterial phylotype, usually a genus. A database was constructed from T-RFLP analysis of known bacterial species or cloned and sequenced bacterial 16s rRNA genes, and this database was used to identify individual T-RFs in faecal samples. T-RF no. denotes the number of T-RFs identified in the sample. <sup>a</sup> the numbers of T-RF that could be identified using the database. The infant depicted corresponds to infant F in paper II.

An overall picture of the microbiotal composition, as analyzed by T-RFLP, was obtained by principal component analysis (Figure 7). Here, each sample was positioned based on the T-RFs found in that sample. In general, samples that were obtained between 1 week and 6 months of age (samples 1–5) showed clustering by infant, suggesting that each infant had a distinct microbiota. In contrast, all the 12-month samples (A6–F6) and the 6-month samples from two children (A5 and F5) appeared in vicinity of one another but far from the earlier samples. Thus, the later samples were more similar to one another than they were to the early samples from the respective infants. This convergence agrees with the results of a previous study [157].



**Figure 7:** Microbiota development over the first year of life for six infants. Fecal samples from six infants (A–F) obtained on six separate occasions over the first year of life were analyzed by T-RFLP, and the T-RF pattern was analyzed by principal component analysis. Sampling schedule: 1=1 week; 2=2 weeks; 3=1 month; 4=2 months; 5=6 months; and 6=12 months. The position of the sample is dictated by the totality of the T-RFs present in the sample.

In Paper IV, we examined by pyrosequencing fecal samples that were obtained longitudinally between 1 week and 1 year of age from six infants. Four of these infants were the same as those analyzed in Paper II, but due to insufficient remaining material, we had to select 2 additional infants (table 4). For inclusion, the selected infants had to have not been consumed of any antibiotics during the first year of life.

**Table 4:** Infants and sample in papers II and IV.

Infant	Analyzed & called		Delivery mode	Breast feeding	
	Paper II	Paper IV		Exclusive	Partial
1	<b>A</b>	<b>A</b>	Vaginal	4	8
2	<b>B<sup>a</sup></b>		Vaginal	0	0
3	<b>C</b>	<i>D</i>	Vaginal	4	12
4	<b>D</b>	<i>E</i>	Vaginal	4	0
5	<b>E</b>	<b>B</b>	Vaginal	4	4
6	<b>F</b>		Vaginal	6	7
7		<i>C</i>	Vaginal	5	10
8		<i>F</i>	Cesarean	4	9

**Bold** denotes faecal samples were analyzed at age 1 & 2 weeks, 1, 2, 6 & 12 months.

*Italic* denotes faecal samples were analyzed at age 1 week and 1 year

<sup>a</sup>Developed food allergy

The most prevalent (found in 4/6 infants) bacterial taxon in fecal samples at one week and one year by culturing, T-RFLP, and pyrosequencing are showed in table 5.

At one week of age anaerobic bacteria were less prevalent. However, at 1 year of age, many anaerobic bacteria could be found in the fecal samples from the infants such as *Roseburia*, *Eubacterium*, *Faecalibacterium*, *Veillonella* and *Ruminococcus*. This is in accordance with the results of previous studies, e.g., that of Palmer and colleagues, who reported that the infant microbiota starts to converge toward an adult profile by 1 year of age [42].

**Table 5:** Most prevalent bacterial taxon in the fecal sample of infant at one week and one year as determined by culture, T-RFLP and pyrosequencing.

	One week			One year		
	Culture	T-RFLP	Pyro	Culture	T-RFLP	Pyro
<b>Bacteroidetes</b>						
<i>Bacteroides</i>	X	X	X	X	X	X
<i>Parabacteroides</i>			X			X
<i>Bacteroidales</i>			X			
<i>Porphyromonadaceae</i>						X
<i>Alistipes</i>						X
<b>Proteobacteria</b>						
<i>Enterobacteriaceae</i>	X	X	X	X	X	X
<i>Ralstonia</i>			X			
<i>Acidovorax</i>			X			X
<i>Haemophilus</i>			X			X
<i>Pasteurellaceae</i>			X			X
<i>Sutterella</i>			X			X
<i>Acinetobacter</i>			X			
<b>Actinobacteria</b>						
<i>Bifidobacterium</i>	X	X	X	X	X	X
<i>Propionibacterium</i>			X			
<i>Eggerthella</i>						X
<b>Firmicutes</b>						
Co-neg Staph <sup>*</sup>	X	X	X	X		
<i>S. aureus</i>	X					
<i>Gemella</i>			X			
<i>Clostridium</i>	X	X		X	X	X
<i>Clostridiales</i>			X			
<i>Enterococcus</i>	X			X		
<i>Streptococcus</i>		X	X			X
<i>Lactobacillus</i>			X		X	
<i>Lactobacillales</i>			X			
<i>Coprobacillus</i>				X	X	
<i>Ruminococcus</i>					X	
<i>Veillonella</i>		X	X		X	X
<i>Veillonellaceae</i>			X			X

<i>Selenomonas</i>	X	
<i>Peptostreptococcaceae</i>		X
<i>Eubacterium</i>		X
<i>Lachnospiraceae</i>		X
<i>Roseburia</i>		X
<i>Anaerotruncus</i>		X
<i>Blautia</i>		X
<i>Butyricoccus</i>		X
<i>Dorea</i>		X
<i>Faecalibacterium</i>		X
<i>Oscillibacter</i>		X
<i>Dialister</i>		X
<i>Clostridiaceae</i>		X
<i>Verrucomicrobia</i>		
<i>Akkermansia</i>		X

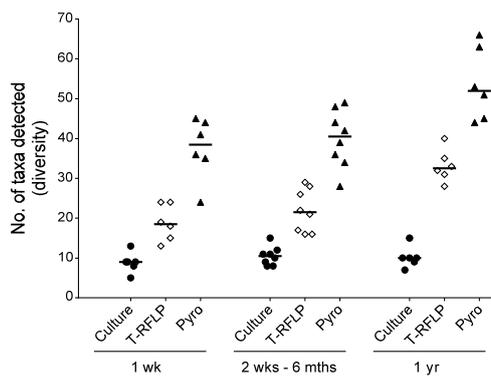
\*Coagulase-negative staphylococci, based on coagulase test

## **Methodologic considerations-Advantages and drawbacks**

In total, 40 fecal samples collected from 8 infants who were followed from 1 week to 1 year of life were analyzed. All of these samples had previously been analyzed by quantitative culturing, and the samples from six infants were analyzed using both T-RFLP and pyrosequencing. This permits us to compare the performance of the three methods for the analysis of the large bowel microbiota of infants.

### **Detection limits of the three methods**

In the 1-week samples, twice as many OTUs were identified by pyrosequencing as were detected by T-RFLP, which in turn identified twice as many OTUs as did quantitative culturing (Figure 7). With T-RFLP and pyrosequencing, an increased complexity of the microbiota over the first year of life was noted, a phenomenon that was not seen using quantitative culturing. This supports the notion that exclusively oxygen-sensitive anaerobic bacteria settle in the commensal microbiota mainly during the second half of the first year of life; the culturing conditions applied were not adapted to isolate and identify these bacteria (Table 5), as the main focus was on the early microbiota and on bacterial groups that have significant impacts on immune maturation in the infant. It is known that obligate anaerobes elicit very meager immune responses, supposedly because they do not translocate the mucosal barrier[90]. At 1 year of life, 45–70 individual genera were identified in each infant by pyrosequencing (Figure 7). It should be noted that while all six infants contributed fecal samples at 1 week and 1 year of age, we only analyzed those samples that were obtained from two infants in the intermediate period (2 and 4 weeks and 2 and 6 months of age).

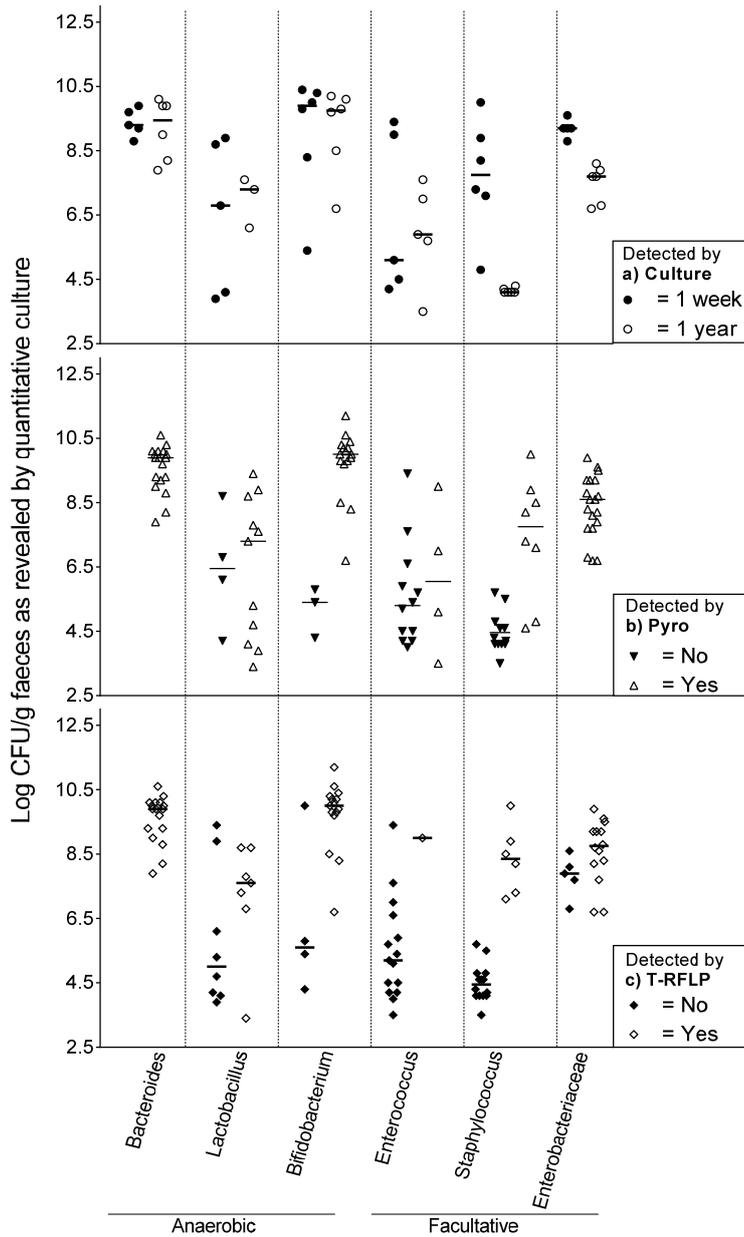


**Figure 7:** Numbers of taxa in individual fecal samples from infants, as identified by quantitative culturing, T-RFLP, and pyrosequencing, and differences in the compositions of the early (1 week) and late (1 year) microbiota, as revealed by pyrosequencing

Quantitative culture (Figure 8a) showed that *Bacteroides* were present in high numbers,  $>10^9$  CFU/g feces, both in the 1-week and 1-year samples. By culturing, it was clearly evident that *Escherichia* was the dominant genus among the Proteobacteria (*E. coli* being its single species). However, sequencing of the 16S rRNA gene cannot adequately resolve the different genera of the Enterobacteriaceae family (*Escherichia*, *Klebsiella*, *Proteus* etc.), which makes it impossible to use this type of DNA-based methodology to distinguish clinically important members of the infant microbiota. In young infants, *Klebsiella* is initially as common as *E. coli* in the commensal microbiota, although this genus and Enterobacteriaceae members other than *E. coli* become less prevalent when the microbiota "matures" toward the end of the first year of life. The inability of sequence-based methods to distinguish between different members of the Enterobacteriaceae is a distinct disadvantage given the different ecologic niches of these bacteria. The primary habitat of *E. coli* is in the gastrointestinal tracts of humans and many warm-blooded animals whereas *Klebsiella*, *Enterobacter*, and non-*E. coli* Enterobacteriaceae, are commonly found in nature. Vaginally delivered infants are, therefore, colonized by *E. coli* much earlier than Cesarean-delivered infants, while the latter more often acquire other types of enterobacteria. These differences can readily be revealed by culturing [49, 118, 119], though not by sequencing methods.

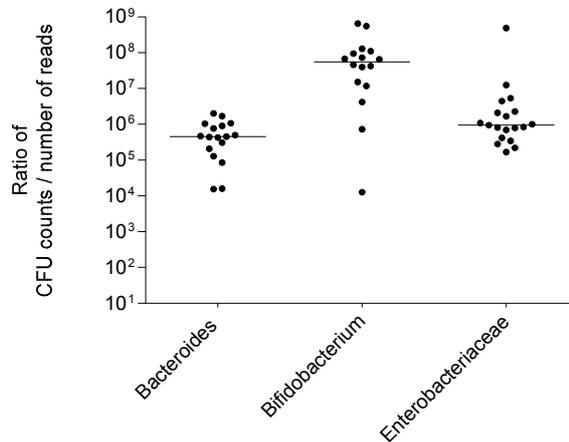
As shown in Figure 8a, quantitative culture reveals that the counts of *Enterobacteriaceae* decrease approximately ten-fold (one log unit) between 1 week and 1 year of age, which is a consequence of increased competition caused by the successively more complex microbiota and the increasingly anaerobic milieu. In contrast, obligate anaerobes do not decrease in number: bifidobacteria and *Bacteroides* remain numerous at 1 week and 1 year of age in infants who harbour the genus in question in their gut microbiota.

Interestingly, *Bacteroides* were detected equally well by all three methods. Thus, all the fecal samples that contained *Bacteroides* spp., as determined by culturing, were also found to contain the *Bacteroides* genus by pyrosequencing and T-RFLP. Notably, samples that were positive for *Bacteroides* contained high population counts of this genus, i.e.,  $>10^{7.9}$  CFU/g feces. In sharp contrast, *Bifidobacterium* spp. were detected by pyrosequencing in all the samples that contained  $>10^6$  CFU bifidobacteria/g feces but were missed in the three samples with lower bifidobacterial population counts (Figure 8b). T-RFLP also failed to detect bifidobacteria in these samples, as well as in a sample with a high count of bifidobacteria (Figure 8c).



**Figure 8:** Population levels of the major taxa in early and late fecal samples from infants, as revealed by quantitative culturing (a) and the probabilities of detecting or missing these taxa in pyrosequencing (b) and T-RFLP (c) analyses, respectively.

We determined the number of CFUs (in theory, each CFU represents one live cultivatable bacterium) that had to be present in a fecal sample for it to generate one sequence read (Figure 9). We found that for *Enterobacteriaceae* and *Bacteroides*,  $10^5$ – $10^6$  CFU/g feces were needed to generate one read. This is close to the theoretical calculation.



**Figure 9:** Numbers of CFU per read in samples that were deemed positive by both culture and pyrosequencing

However, the sensitivity for detection of bifidobacteria was 100-times lower:  $10^8$  CFU/g had to be present in order to generate one pyrosequencing read. This very low level of sensitivity was masked by the fact that the infants' feces contained very high numbers of this genus. It is plausible that bifidobacteria are often missed in the fecal samples from adults, where they are also present in high numbers ( $10^9$ /g) [50], although they are seldom reported in papers that use sequence-based methodologies.

Samples that contained *Enterococcus*, *Staphylococcus* and *Lactobacillus* spp (all Gram-positive bacteria) were often missed by pyrosequencing. This is partly due to the low populations of these bacteria, but also the fact that they were frequently missed in those samples in which they were present in relatively high

number. This was even more evident when T-RFLP was used. In contrast, all the samples that contained Gram-negative *Enterobacteriaceae*, as detected by culturing, were also found to be positive by pyrosequencing, even if they were present in numbers that were no higher than those of the Gram-positive bacteria.

Table 6 combines the data from the two methodology papers (Papers II and IV), and compares the three DNA-based methods (T-RFLP, cloning-sequencing, and pyrosequencing) with culture. The sensitivity of culture is arbitrarily set at 100%, i.e., a method that detects the genus less often than culturing is assigned a value <100%, while a method that is more sensitive than culture is assigned a value >100%. Some factors that may influence the detection limit are also included, such as the number of copies of the 16S rRNA gene carried by a single bacterial cell and the Gram-staining status.

**Table 6 Comparisons of the detection sensitivities of DNA-based methods relative to the sensitivity of culturing (set at 100%)**

Phylum	T-RFLP	Pyro	Cloning	CFU	16S copy no.	Gram-
Taxon	detectability (%) <sup>a</sup>		% <sup>b</sup>	median	median	staining
<b>Bacteroidetes</b>						
<i>Bacteroides</i>	100	105	40	9.9	6	G- bacilli
<b>Proteobacteria</b>						
<i>Enterobacteriaceae</i>	74	100	10	8.6	7	G- bacilli
<b>Actinobacteria</b>						
<i>Bifidobacterium</i>	79	84	0.2	9.9	4	G+ bacilli
<b>Firmicutes</b>						
<i>Clostridium</i>	106	38	3	6.0*	9	G+ bacilli
<i>Lactobacillus</i>	67	80	0.1	6.8	6	G+ bacilli
<i>Enterococcus</i>	6	31	0	5.3	4	G+ cocci
<i>Staphylococcus</i>	30	40	0	4.7	5	G+ cocci

<sup>a</sup> compared to the result obtained by culture, 100 %= as good as culture.

<sup>b</sup> based on the 834 clones obtained from 9 faecal samples in paper II

\* alcohol-treated faecal sample

Cloning and sequencing showed very low sensitivity for the investigated genera. This is understandable, considering the few

reads per sample that this method generates, usually in the hundreds. If the microbiota contains  $5 \times 10^{11}$  bacteria/g and 500 sequences are obtained, they will, in principle, all be derived from bacteria that are present in population levels of at least  $10^{10}$ /g, i.e., some dominant obligately anaerobic bacteria. However, it should be noted that bifidobacteria, which clearly are present at such high levels, are still not detected, once again demonstrating the inferior capacities of DNA-based methods to detect this group of bacteria. Three reasons can be put forward to explain the insensitivity to bifidobacteria: 1) they are Gram-positive (it is much easier to disrupt the Gram-negative bacterial cell wall when extracting DNA); 2) they carry relatively few copies of the 16S rRNA gene; and 3) PCR primer mismatch.

PCR primer mismatch has an important negative impact on detection sensitivity [66]. Bifidobacteria belong to the Actinobacteria phylum, previously considered to be Gram-positive bacteria with a high GC content [67]. The commonly used forward primer 8–27 (*E. coli* 16S nucleotide positions) has been shown to contain three primer mismatches [62] and is thus inadequate for amplifying the DNA of bifidobacteria, leading to low sensitivity for this quantitatively dominant bacterial genus in the gut. Degenerative primers might increase the primer affinity during PCR by incorporation of inosine (I), as in the case of the primers we used for T-RFLP: ENV-1, 5'-AGA GTT TGA TII TGG CTC AG-3'; and ENV-2, 5'-CGG ITA CCTTGT TAC GAC TT-3'. Normally, the nucleotides in DNA bind as follows: adenine (A) binds only to thymine (T); and cytosine (C) binds only to guanine (G). However, inosine can bind to any of the above-mentioned nucleoties. i.e., A, T, C or G. Analysis of the 16S rRNA genes from different bacteria has revealed the most common primer mismatch positions [67]. Thus, inosine is incorporated at these nucleotide positions in the primers. Despite this modification, our PCR primer could still not amplify the bifidobacteria DNA with satisfactory sensitivity. The fact that they were still identified in most samples can be attributed to their exceedingly high counts in infant fecal samples.

As DNA is more easily released from Gram-negative bacteria, such as *Bacteroides* and *Enterobacteriaceae*, than from Gram-positive

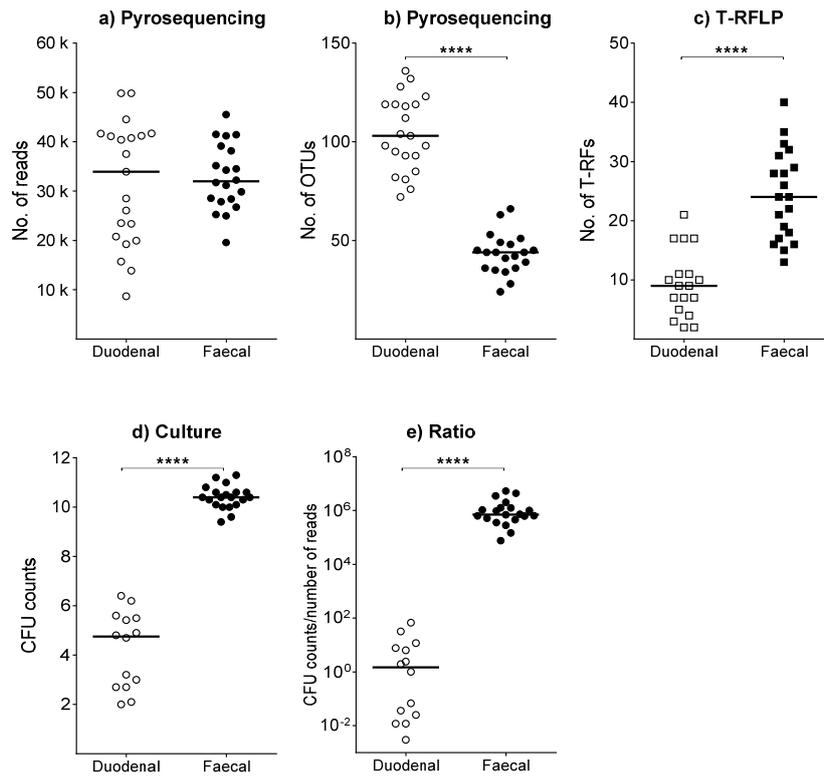
---

bacteria, such as *Enterococcus* and *Staphylococcus*, we combined enzymatic lysis with an additional mechanical bead-beating method to break the bacterial cell wall (Paper IV). However, despite this extra measure to improve the efficacy of DNA extraction, the detection sensitivity for Gram-positive bacteria did not improve. Collectively, our results suggest that Gram-negative bacteria (*Bacteroides* and *Enterobacteriaceae*) are highly overestimated when DNA-based methodologies are used to analyze complex microbiota.

### **Comparison of the performances of pyrosequencing for analyzing small and large bowel microbiota**

Pyrosequencing was used to analyze the duodenal fluid samples from children with IBD and from control children (Paper III), as well as the fecal samples from infants (Paper IV). These groups of samples were analyzed at the same time and on the same sequencing plate, which permitted direct comparisons of the performance of pyrosequencing with these different sample types. The mean numbers of sequence reads did not differ between the two cohorts: the duodenal samples gave  $32 \pm 1.2$  k reads, and the fecal samples gave  $33 \pm 7$  k reads (Figure 10a).

Culturing revealed that the duodenal samples contained maximally  $<10^6$  CFU/g feces, as compared with  $>10^{11}$  for the fecal samples. Despite this, the duodenal fluid generated an average of 104 OTUs per sample, as compared with an average of 43 OTUs for the fecal samples (Figure 10b)! It seems highly unlikely that the duodenal microbiota contains 200-times more bacterial genera than does the colonic microbiota, even if the fecal samples were from young infants and the duodenal microbiota was from older children/adolescents whose microbiota could be more mature/complex. Indeed, the colonic microbiota contained more T-RFs than the duodenal microbiota when they were both analyzed by T-RFLP (Figure 10c). Moreover, pyrosequencing did not show significantly higher numbers of OTUs at 1 week of age (average of  $38 \pm 8$  OTUs) than at 1 year of age ( $54 \pm 9$  OTUs), which represent an increase of only 16 OTUs in the average infant.



**Figure 11. Duodenal fluid samples and faecal samples analyzed using pyrosequencing, T-RFLP, and quantitative culturing.**

The major difference noted between the two sample types was the vast difference in the density of bacteria, with the faecal samples having 5 log units higher counts, as measured by culturing, than the duodenal samples (medians:  $10^{10}$  vs.  $10^{4.7}$ ) (Figure 10d). If anything, the colonic counts should be underestimated as compared to the duodenal counts, as the colonic microbiota is much more anaerobic and is thus more difficult to culture. In Figure 10e, we have calculated how many CFU counts are needed to generate one sequence read. In general, the faecal samples required  $10^6$  CFU to generate one sequence read, while the duodenal samples could generate one read already at  $10^1$  CFU (10 bacteria). The most likely explanation for the lower OTUs detected

---

in the fecal samples is that the high population levels of certain fecal bacteria hamper the detection of subdominant genera, thereby limiting the depth of pyrosequencing.

In this respect, T-RFLP has a certain advantage. Despite being less sensitive overall than pyrosequencing, it is more robust when there is an unequal distribution of population sizes, since the fragments are separated according to size before they are detected. Therefore, the detection of, for example, lactobacilli is not hampered by the presence of 10,000-times more Bacteroides.

## Duodenal microbiota and IBD

The duodenal fluids of children with newly diagnosed IBD were characterized using high-throughput sequencing and culture. We compared diversity and composition in children with IBD with symptomatic controls, i.e., children with symptoms compatible with IBD and, hence, referred to a pediatric specialist clinic for diagnosis, but whose gut mucosa displayed no signs of inflammation (thereby excluding a diagnosis IBD). To control for any effects on microbiota composition, we excluded all children who had received disease-modifying drugs or who had been taking any antibiotics in the previous 3 months. Duodenal fluids were analyzed by quantitative culturing and pyrosequencing (average of 32,000 reads/sample, average read of 500 nt). Quantitative culturing revealed bacterial levels of  $>10^5$  CFU/mL in 26% of the children (two patients with ulcerative colitis and three controls).

Pyrosequencing showed that the duodenal microbiota of children with ulcerative colitis was less diverse, i.e., contained fewer Operational Taxonomic Units (OTUs), than those of the controls. Although the microbiota of children with Crohn's disease also tended to be less complex than those of controls, there was greater variability in the Crohn's disease group and the difference did not reach statistical significance. The lower diversity of the microbiota in children with ulcerative colitis was not restricted to any particular bacterial group, but was seen across three major phyla: Actinobacteria, Bacteroidetes and Firmicutes, included bacterial genera, e.g., *Collinsella* ( $P<.001$ ), *Bacillus* ( $P=.007$ ), *Lactobacillus* ( $P=.007$ ), and unclassified *Sphingobacteriales* ( $P=.007$ ), were isolated less frequently from the children with ulcerative colitis than from the controls. *Collinsella* was also less common in patients with Crohn's disease than in control subjects ( $P=.007$ ).

Compared to control subjects, the duodenal microbiota of children with ulcerative colitis is less diverse and several bacterial groups are present at reduced frequencies. Whether these changes in the microbiota contribute to the disease process or are a consequence of the inflammatory disease remains to be determined.

---

## DISCUSSION

The commensal microbiota of the alimentary tract contains enormous amounts of antigens and inflammatogenic substances. Commensal bacteria and fungi translocate across the mucosal membranes, whereby they come in contact with the immune system, which is thereby activated. The commensal microbiota is an important stimulus for the immune system and has been suggested to play roles in several immune-mediated diseases. This has spurred interest in the composition and diversity of the microbiota during infancy and childhood, in a search for microbes that contribute either negatively or positively to the etiology and pathogenesis of allergy and inflammatory bowel disease. Traditionally, quantitative culture on selective and non-selective media has been used to map the microbiota. However, in the last couple of decades, DNA-based methodologies have become standard. These methods permit analyses of frozen samples, since they do not require that the bacteria are viable. However, there are few studies that directly compare culture-based and DNA-based methodologies in terms of their sensitivity and accuracy in characterizing complex microbiota, such as that of the alimentary tract. In this thesis, we have compared quantitative culturing of fecal samples with the DNA-based methods of T-RFLP, pyrosequencing, and cloning-sequencing. An understanding of the advantages and limitations of each method will enable more accurate analyses of the microbiota and, thus, a better understanding of its role in allergy and inflammatory bowel disease.

Terminal-Restriction Fragment Polymorphism (T-RFLP) exploits variations within the 16S rRNA genes for bacterial identification. The ubiquitous bacterial 16S rRNA gene is amplified by PCR and cleaved with a restriction enzyme (or a combination of enzymes in some applications of the method). Each bacterial taxon generates a fragment, termed the T-RF, of defined size. When applied to a complex microbiota, T-RFLP provides rapid fingerprinting of a bacterial community and the method is well suited to analyzing bacterial diversity in different environments, such as soil, water,

and the guts of humans and animals [158, 159]. A disadvantage of T-RFLP is that it is not immediately evident which bacterial taxon generates a particular T-RF. Several web tools are available to identify T-RFs [160]. These databases allow calculations of the sizes of T-RFs from 16S rRNA genes once the user enters the PCR primers and restriction enzyme (“in silico digestion”). However, in practice, the utilization of the web-based tools requires accurate fragment size determination. We show here, as previously demonstrated [161], that there are discrepancies between the true and observed T-RF sizes that are inherent to the method. We demonstrate here that one source of this discrepancy is that different fluorescein dyes are used to label the sample and the molecular weight standard, which has also been pointed out previously [70]. Moreover, the differences between the theoretically calculated and apparent fragment sizes vary across bacterial taxa, which mean that a correction algorithm cannot be applied to compensate for the discrepancy. Kaplan et al. [162] have suggested that the purine content of the 16S fragment influences the T-RFLP runs, which may be one source of the variation in T-RF migration time observed between bacterial taxa. Commonly available T-RFLP analysis programs do not take all these factors into consideration, and our results suggest that this is, indeed, impossible. Therefore, the use of automated T-RFLP programs may result in faulty identification of the bacterial taxa in a community. To identify the T-RFs generated from the analysis of the infant fecal microbiota, we developed our own database. T-RFLP analyses of pure-culture isolates of bacteria from infant fecal samples and from type strains of gut bacteria formed the basis for the analysis, which was complemented by the cloning and sequencing of a limited number of isolates from the infant fecal samples. Although we selected samples with many unidentified T-RFs for cloning and sequencing, this procedure mostly generated a large set of *Bacteroides* sequences. The limited sensitivity of the cloning and sequencing procedure means that only quantitatively dominant bacterial populations can be detected (see below).

The T-RFLP database allowed us to identify directly approximately 60% of the peaks detected in the infant fecal samples. Thus, despite extensive efforts, a substantial proportion of the infantile

---

gut microbiota still consisted of unidentified bacteria, which is a clear limitation to using the T-RFLP method for gut microbiota samples.

Using our database, we showed that T-RFLP permitted identification at the genus level in most cases. Several genera, in fact, generated more than one distinct T-RF, which is in accordance with the results of a previous study [163]. Thus, T-RFLP has almost the same taxonomic resolution potential as culturing, which by definition permits identification of the species, provided that various typing methods are applied to the pure-culture isolates.

Pyrosequencing of moderately long fragments of the 16S rRNA gene (average: 500 nt) and with an average number of reads of 30,000 per sample (today termed 'ultra-deep sequencing') was compared with culturing and T-RFLP for the analysis of infant fecal samples. As pyrosequencing does not use information on the entire 16S rRNA gene, it is impossible to determine with accuracy the species of a bacterial taxon. Even identification to the genus level may be ambiguous. Wang *et al.* calculated that when using a 400-nt primer the chances of correctly assigning a genus, family, order, class, and phylum were 88.7%, 94.6%, 97.7%, 99.2% and 99.8%, respectively [63]. This fact is seldom taken into consideration in studies that use sequence-based methods for the analysis of complex bacterial communities. As the sequence is aligned to previously published sequences, and since these sequences are attributed to species, it is tempting to believe that the sequence does contain enough information to determine accurately the species from which it is derived. However, this is not the case. We chose to use the genus level as the taxonomic "cut-off" for our pyrosequencing method. However, despite this conservative approach, there is an approximately 11.3 % risk that a sequence will be assigned to the wrong genus [63].

Sequencing the entire 16S rRNA gene, as in the cloning and sequencing method, in most cases generates sufficient information to assign a species to the sequence in question. However, this method is of limited use in analyses of very dense bacterial communities, such as the fecal microbiota. As it is immensely

labour-intensive (each DNA fragment must be cloned into a bacterium, which must subsequently be propagated and lysed and its DNA subjected to sequencing), the number of DNA fragments that can be analyzed per sample is fairly limited (usually <1000 per samples). It is clear that analysis of a random collection of 1000 bacteria among a population of  $10^{11}$  gives little information about the overall community, as only the most quantitatively dominant bacteria are likely to be analysed, for purely statistical reasons. This was evident in our attempts to identify non-cultivable bacteria in a few infant fecal samples by cloning and sequencing; few bacterial taxa were identified and most of the sequences were derived from *Bacteroides*.

In some instances, not even sequencing of the entire 16S rRNA gene yields the genus, the reason being the close genetic relatedness of these bacterial genera. Examples include the streptococci (which include pneumococci, which are highly virulent bacteria, as well as numerous species that are collectively referred to as the alpha-streptococci), and the Enterobacteriaceae which contains many of the most clinically relevant gut bacteria, such as *E. coli* and *Klebsiella*, which are also highly prevalent in the infant microbiota. DNA-based differentiation of the *Enterobacteriaceae* family usually requires sequencing of both the 16S gene and additional housekeeping genes, e.g., *gyrB* [164] and *dnaJ* [165]. In contrast, *Escherichia*, *Klebsiella*, *Enterobacter*, *Citrobacter*, and *Proteus* are easily distinguished using culturing and simple biotyping tests. These species are of clinical importance in that they are major causes of urinary tract infections and also quite common causes of sepsis and local infections that emanate from the gastrointestinal tract.

For some applications, identification to the genus level is inadequate. For example, the clinical and immunologic effects of being colonized by *S. aureus* are profoundly different from the effects of being colonized by any of the species collectively referred to as coagulase-negative staphylococci. While the former is one of the most common invasive pathogens, the latter cause infections only in immunocompromised hosts, including premature infants. Furthermore, *S. aureus* produces a range of

---

extremely potent T-cell-stimulating toxins, the superantigenic toxins, as well as the B-cell-superantigenic protein A. We have shown that early gut colonization by *S. aureus* is associated with decreased risk of developing food allergy within the first 18 months of life, and that gut colonization with superantigen-producing strains is associated with higher serum levels of IgA [166] and soluble CD14 [167] in the colonized infants.

In conclusion, the taxonomic accuracies of T-RFLP and pyrosequencing are similar, i.e., they permit, in most cases, taxonomic identification to the genus level. When a more detailed taxonomic distinction is required, these methods are inadequate.

Apart from the identification of specific bacterial taxa in a complex microbial community, it is often of interest to determine the general complexity of the microbial community, i.e., how many bacterial taxa it contains. Thus, the feature that is most often associated with protection against allergy development is to have many different bacterial groups making up the early microbial microbiota [122, 124, 125, 156, 168]. The determination of microbiota complexity is by no means straightforward. It is often reported that thousands of individual taxa are found in fecal samples analyzed by DNA-based methods, such as pyrosequencing. However, in our hands, pyrosequencing of infant fecal samples detected on average only 43 OTUs per sample and a mere 126 unique OTUs in all 20 infant fecal samples combined, despite our "ultra-deep" sequencing rate of 30,000 reads per sample and average length of 500 nucleotides.

Although the microbiota of an infant is less complex than that of an adult, it is difficult to reconcile our results with those previously obtained, which suggests that thousands of OTUs are generated by sequencing one microbiota from one individual. A distinct possibility is that previous DNA-based studies overestimated the numbers of different taxa [30] and falsely identified novel taxa in the gut flora due to methodologic errors [54, 66]. A number of pitfalls are linked to pyrosequencing, and faults may be introduced at several steps, i.e., the PCR step, sequencing step, and alignment/annotation step. PCR has an inherent imperfection, in

that roughly  $1.2 \times 10^{-6}$ /PCR cycle of the bases is incorrectly incorporated [79]. Another problem with PCR is the formation of chimeras, which occurs when two DNA sequences from two different bacteria become amplified to generate a single PCR product, which is a chimera of the two bacterial sequences but is interpreted as being from “a new species”, the chimera rate was more than 70% of the less-abundant species were chimera [31]. Thus, extreme care must be taken to avoid the appearance of false sequences that may create a false impression of the wide diversity of a bacterial community. Classical culture-based studies employing the best available strict anaerobic culture methodology in the 1970s revealed that about 200 distinct species could be identified in one individual [169]. As microscopic evaluation of the fecal sample revealed a number of bacterial cell types that had not appeared in the culture-based analysis, it was assumed that the meticulous culturing approach had succeeded in culturing half of the bacterial species present, and the estimate of 400 species in the intestinal microbiota of an adult individual was proposed [169]. It should be noted that this culture-based approach identified facultative bacteria that are normally missed by DNA-based methodologies, such as enterococci. The number of taxa that may be identified by sequencing alone is likely to be lower than 400.

A very interesting observation was made by comparing the number of OTUs detected by pyrosequencing of the duodenal microbiota (from children with IBD and age-matched controls), as compared with the microbiota of the fecal samples (from infants). Quite remarkably, an average of 100 OTUs was found per duodenal sample, which is more than twice as many as were identified in the colonic microbiota. It is very unlikely that the duodenum contains a greater variety of bacterial taxa than the colon, the latter being regarded as one of the most complex microbial ecosystems known. Indeed, T-RFLP identified 2.5-times more T-RFs in the average fecal sample than in the average duodenal fluid sample. The most likely explanation for this is that the low numbers of bacteria in the duodenal fluid meant that the DNA from almost every single bacterial cell could be analyzed (30,000 reads would cover  $3 \times 10^4$  bacteria, which is a typical population level in the duodenum).

---

These findings provide a clear indication that pyrosequencing, despite being labeled as “ultra-deep”, only skims the surface of the complex colonic microbiota, identifying quantitatively the dominant groups, while missing quite a high number of sub-dominant taxa.

The modest number of different taxa detected in our study may reflect the fact that the gut microbiota of Swedish infants is, indeed, quite meagre, due to prolonged exclusive breast-feeding, small family size, and high standards of hygiene. We have previously reported delayed colonization of Swedish infants, as compared with Pakistani infants, by typical commensal gut bacteria [118], and sluggish turnover of individual *E. coli* strains in the gut microbiota [119]. These findings suggest that infants in affluent Western countries are exposed to low microbial loads, and that this contributes to low microbiota complexity. To compare adequately the performance of pyrosequencing in elucidating the complexity of the duodenal and colonic microbiota, samples from the same individuals must be analyzed. Such studies are underway, as we have corresponding fecal samples for all the duodenal samples from children with IBD and their controls.

We and other groups have shown that culture is substantially more sensitive than DNA-based methods for the detection of both obligate and facultative anaerobic bacteria, provided that selective media and appropriate culturing conditions are available [170] [69]. Culture-based analysis can detect cultivatable bacteria down to population levels of  $10^{2.5}$  bacteria/g feces, while our study suggests that the current sequencing techniques detect fecal bacteria when they are present at  $\geq 10^6$  CFU/g sample, as approximately 1 in  $10^5$ – $10^6$  fecal bacteria yielded one pyrosequencing read. This is in accordance with a previous report [171] and the theoretical calculations showing that at least 100,000 sequences must be analyzed to obtain one read of a bacterium present at a level of  $10^6$  CFU/g within a total population of  $10^{11}$  CFU/g. Thus, the high population levels of certain bacteria, which are readily detected by PCR and sequencing, e.g., *Bacteroides*, hamper the detection of subdominant less-dominant genera. Thus, much of the depth of pyrosequencing is “wasted” on

the masses of sequences obtained from quantitatively dominant genera, such as *Bacteroides*.

The fact that bacterial genera that are present in relatively low numbers are missed by pyrosequencing was readily demonstrated by comparing the results from the analyses of fecal samples taken from 1-week-old and 1-year-old infants. OPLS analysis based solely on the pyrosequencing data showed that *Staphylococcus*, *Enterococcus*, and *Lactobacillus* were detected more frequently in the 1-week samples than in the 1-year samples. In fact, staphylococci and enterococci were present at 1 year of age in practically all the infants, and lactobacilli were more prevalent at 1 year than at 1 week of age. However, as their population counts decreased, concomitant with increased competition from the microbiota over the first year of life, the numbers of these bacteria fell below the detection limit of pyrosequencing and “disappeared” from the microbial analysis.

There is a general belief that DNA-based methods, especially next-generation sequencing, are able to reveal “all” the bacteria in the gut microbiota. As shown in our studies and as discussed above, this is a misconception; all methods have certain limitations. Culture is, by definition, substantially more sensitive than any other method, provided that selective media and appropriate culture conditions are available. This includes all facultatives, which may easily be isolated from among billions of obligate anaerobes using the simple technique of culturing the sample under aerobic conditions, which few of the gut anaerobes can tolerate. As many facultative bacteria have important functions as immune modulators or pathogens, we believe that it is important to point out that such bacteria often escape detection when DNA-based methods are employed. Therefore, in future studies, combining both culture-based and DNA-based methods with greater sequencing depth may pave the way for a better understanding of both the bacterial composition and physiology of the microbiota living within us.

Our interest in microbiota analysis derives from the hypothesis that the infantile microbiota is a major stimulus for the developing

---

immune system and that microbial complexity and composition may affect the risk of allergy development. Indeed, a low complexity of the early microbiota is a risk factor for subsequent allergy development. In Paper II, analysis of the T-RFLP patterns of infant fecal samples yielded an interesting picture, which, to the best of our knowledge, has not been reported before. In most of the infants, a quite high bacterial diversity characterized the 1-week and 2-week fecal samples, although microbial complexity decreased thereafter to reach a nadir at around 1 month of age. Thereafter, the complexity of the microbiota increased markedly again, being highest in the last sample collected at 12 months of age. It is well known that microbial complexity increases over the first year of life as more and more anaerobic bacterial species become established in the gut [50]. However, the high bacterial complexity of the samples obtained at 1 and 2 weeks of age was an unexpected finding. Several bacterial species that were found in the early fecal samples disappeared from the microbiota after the first weeks and did not reappear. Some of these may have come from the birth canal, as all the infants examined here were vaginally delivered, or from the environment. Such non-professional gut colonizers may have a chance to expand during the first weeks but they are not able to withstand the competition as more “professional” gut bacteria are acquired. Furthermore, they may be more sensitive to the antibacterial peptides produced in the mucosa in response to colonization. Previous reports have shown that bacteria that originate from the maternal vaginal flora can be cultured from the baby's gastric contents immediately after birth [47], although many of these bacteria cannot colonize the infant's gut.

In Paper I, we demonstrate that a common parental practice - oral cleaning of their infant's pacifier before it is given back to the infant, is associated with protection against early eczema development and asthma symptoms. Early acquisition of a complex intestinal microbiota has been identified as providing protection against allergy development, which suggests that commensal bacteria provide the developing immune system with crucial signals for its proper maturation [122, 123, 168] [125]. We propose that the underlying mechanism involves immune

stimulation by microbes, which are transferred to the baby *via* the pacifier. Undoubtedly, this habit allows for close oral contact between the parents and child, facilitating bacterial transfer at a very young age, before the child starts to use feeding utensils. Parents who do not suck their infant's pacifier may have a more cautious style of living with respect to bacterial transfer to their children. Using T-RFLP to characterize the infant's salivary bacteria, we have gathered evidence that this practice influences the composition of the infant's oral microbiota. At present, we do not know the identities of the microbes that are more or less prevalent in the saliva of the two groups of infants, as development of an oral bacteria database for the identification of these microbes requires substantial additional work (cloning and sequencing). The T-RFLP peak identification database developed in Paper II was designed primarily to identify bacteria in fecal samples and is not suitable for oral microbiota identification. We have attempted to construct an oral bacteria database, but this database at present contains only 50 species and this is clearly not sufficient to identify more than a small fraction of the vast number of bacterial genera present in the oral cavity.

Oral tolerance is actively induced immunologic unresponsiveness to harmless proteins, and it has been known for decades that the presence of commensal microbiota improves oral tolerance [97]. Hitherto, attention has focused almost exclusively on the small intestine as the site of induction of oral tolerance and the influence of the gut microbiota on the handling of dietary antigens. However, the oral cavity is actually the first site of encounters between foreign antigens and the lymphoid system, and it allows the lymphoid system to sample antigens that have not been denatured by acid or degraded by digestive enzymes. The oropharynx is surrounded by dense lymphoid tissues, i.e., the adenoids and the palatine and pharyngeal tonsils. Similar to the gut-associated lymphoid tissues, these are covered with M cells, which specialize in antigen uptake and delivery to antigen-presenting cells and lymphocytes beneath the epithelium [172]. Furthermore, dendritic cells capture antigens in the oral mucosa and migrate to the cervico-mandibular lymph nodes [173], similar to the way in which dietary antigens are carried to the mesenteric lymph nodes.

---

The tonsils are rich in T cells that have a regulatory phenotype [87], and applications of contact allergens to the oral mucosa lead to active tolerance induction [174]. Thus, there is no reason to believe that active oral tolerance is not induced already in the oral cavity. Exposure of the infant to parental saliva might accelerate the development of a complex oral/pharyngeal microbiota, which, similar to a complex gut microbiota [122, 123, 168] [125], might beneficially affect tolerogenic handling of antigens by the oral/pharyngeal lymphoid tissues. Moreover, oral bacteria are swallowed, thereby also affecting the composition of the microbiota in the small intestine, which may in turn regulate tolerance development in the gut. Further studies are now required to establish if parental pacifier sucking represents a simple and safe method to reduce allergy development in infants and young children, as our study suggests.

In Paper III, the duodenal fluids of children with new-onset IBD were characterized using pyrosequencing and culturing. These children had not received either antibiotics or disease-modulating drugs. The duodenal microbiota of children with IBD was distinctly different from that of diseased control children. A potential weakness of our study is the limited number of collected samples. However, this permitted in-depth sequence analysis, with a mean number of 32,000 sequence reads per sample. Thus, the present study represents the most comprehensive examination of IBD-associated microbiota DNA sequences performed to date, including previous studies of the colonic/fecal flora in patients with IBD [150]

In particular, a lower level of diversity was found to characterize the duodenal microbiota of children with IBD. As there have been no previous studies of the duodenal microbiota of patients with IBD, comparisons are difficult. In dogs with spontaneous IBD, duodenal biopsies contained fewer species than those collected from healthy dogs, and the percentages of *Fusobacteria*, *Bacteroidaceae*, *Prevotellaceae*, and *Clostridiales* were lower. Instead, their duodenal microbiota contained more members of the Proteobacteria phylum, e.g., *Enterobacteriaceae*, *Diaphorobacter*, and *Acinetobacter* [175] [176].

The lower complexity of the microbiota in children with IBD also meant that a number of bacterial genera (e.g., *Collinsella*, *Sphingobacteriales*, *Enterococcus*, and *Lactobacillus*) were present at lower frequencies in their duodenal fluids. *Collinsella*, which is a Gram-positive anaerobe that belongs to the Actinobacterium phylum, was present in reduced numbers in both the children with ulcerative colitis and the children with Crohn's disease. This genus consists of four species, with *C. aerofaciens* being frequently found in the human large bowel [177] [178], albeit at lower levels in patients with severe irritable bowel syndrome [179]. The numbers of *Collinsella* have previously been reported to be lower in the bowel microbiota of persons in the Western hemisphere than in people from Asia and Africa [180] [181], and *Collinsella* is also found at lower frequencies in populations at high risk for colon cancer than in low-risk populations [182]. It remains to be elucidated whether *Collinsella* has a direct protective effect on gut health and whether this genus is particularly sensitive to the altered conditions prevailing in patients with IBD.

An OTU that belongs to the *Sphingobacteriales* order of the phylum Bacteroidetes was also detected less frequently in the IBD patients, particularly in those with ulcerative colitis. This OTU could not be assigned to any genus, family or class. The name *Sphingobacteriales* alludes to the production of sphingolipids by species within the phylum Bacteroidetes [183]. Sphingolipids are recognized by invariant natural killer T (iNKT) cells and may exert immunomodulatory effects [184]. Invariant NKT cells are important regulators of immune system. Animal studies have shown a protective effect of inhibitory sphingolipids against experimental iNKT cell-mediated, oxazolone-induced colitis [185]. *Lactobacillus* was also detected more frequently in the diseased controls. Lactobacilli are widely present in food products, such as yoghurts, and various probiotic lactobacilli have been shown to have immune-stimulating properties [186]. These bacteria are also commonly found in the oral [15], small intestinal, and colonic microbiota.

A drawback of our study is that we lacked duodenal fluid samples from healthy children. Esophagogastro-duodenoscopy is a

---

challenging intervention that cannot be performed on healthy children for purely research purposes. In addition, in a study of IBD conducted in dogs, ethical concerns precluded the inclusion of healthy control animals [176]. The few studies of the human duodenal microbiota that have been carried out have focused primarily on specific bacterial groups in a limited number of diseases, i.e., small intestinal bacterial overgrowth, irritable bowel syndrome, and celiac disease. These studies have been reviewed by Wang [187], who concludes that little is known regarding the duodenal microbiota, either in healthy individuals or in patients. Our diseased controls showed the anticipated typical healthy small intestinal flora, which is dominated by the genera of *Lactobacillus*, *Enterococcus*, and *Streptococcaceae* and comprises members of all four dominant intestinal phyla. With the sequence-based analysis, we demonstrate that the microbiota of control children show greater diversity across all phyla and contain more OTUs that cannot be assigned to any known bacterial phylum or that can even be classified as bacterial in origin, than the microbiota of children with ulcerative colitis.

Hopefully, our studies can be used to improve methodologies for the analysis of the intestinal microbiota of infants and young children, as these advances are needed to highlight the importance of microbial complexity and composition for the development of immunoregulatory diseases. such as allergy and inflammatory bowel disease. Although the hygiene hypothesis fits with all available epidemiologic data, we still do not know exactly which types of microbial stimuli are required to provide the appropriate maturation signals to the developing immune system so that it develops functional tolerance to harmless antigens.

## ACKNOWLEDGEMENTS

*"Nog finns det mål och mening i vår färd - men det är vägen, som är mödan värd"* Denna dikt av Karin Boyes stämmer så väl med den lyxiga tiden som doktorand. Jag har vuxit enormt som person och haft roligt under hela min doktorandperiod, aldrig haft en enda måndagsångest. Många har bidragit till denna arbetsglädje och hjälpt mig genomföra detta arbete! TUSEN TACK till er alla!

Särskilt tack till alla familjer, barn, läkare och sjuksköterskor som deltog i forskningstudieerna på Mölndals Sjukhus resp. Drottning Silvias Barnsjukhus.

**Ia** och **Agnes** för att ni skyddar mig under era vingar och guidat mig genom hela arbetet! **Ia** för din visdom, snällhet och rättvisa, din dörr är alltid öppen för mig. **Agnes** det finns ingen handledare som du, du har alltid uppmuntrat mig och får mig att se saker från andra perspektiv. Jag har fått fria händer att jobba med det som intresserat mig.

**Forough** som fick mig att doktorera och lärde mig allt du kunde. **Ed** for the thousand times I barged into your office with hundreds of questions, every time you taught me something new. **Vincent**, I really appreciate your professional help in editing my thesis.

Damerna **Eva**, **Ingela**, **Jolanta** - tack för att ni finns utan er hade det inte gått! **Eva** för din klokhet, **Ingela** för din positiva livssyn, **Jolanta** (min psykolog) för alla komplimanger och personliga terapier.

**Sofia** det finns så många ord men det räcker nog med: jag tycker så mycket om dig. **Anna-Carin** för alla fina komplimanger, men egentligen så jag tycker du är en "Natural Beauty". **Hardis** för trevlig sällskap i Taos. **Erika L** som alltid är glad och positiv. **Nahid Karami** för alla diskussioner och plasmidsamarbete. Du lärde mig Nanodropp även under ditt avhandlingstid, det säger så mycket om dig som person. **Eva J**, **Merna** och **Nahid Kondori** för alla trevliga luncher och sällskap under åren. Tacka alla kloka Post-doc. **Anna** och **Jenny**. **Madeleine** för din prisbelönta avhandling och

andra förstklassiga föreläsningar som du delat med dig. **Åsa** för dina gedigna kunskaper i Prism. **Ignacio** chefen! För alla trevliga samarbeten, alla mina dumma frågor och dina nyanserade svar. Tidigare doktorander & kollegor **Anna Ö, Anna S, Cillan, Erica B.** Ni är underbara som förgyllde vardagen, för all hjälp och alla diskussioner under åren.

**Liselott** för all hjälp under åren, allt från lån av kemikalier till löneförhandling. Personalen på CCUG, **Kent, Kristel, Maria, Susann** och **Elisabeth** för alla stammar som jag fick och det känns alltid välkomnade att komma ned till er. **Sofia** som lärde mig allt om sekvensringskonst. **Gunnar Dahlen** och **Lisbeth** tack för allt samarbete, ni lärde mig så mycket om munflora. **Robert Saalman** för alla intressanta diskussioner om IBD.

Alla trevliga och snälla läkare på våning 6, **Annika, Susann** och **Brynja. Lisa** för kurs sällskapet och din omtänksamhet, **Bodil** för alla trevliga svamplockningsturer, fina blombilder och alla botaniska diskussioner. **Inger Mattsby** för alla diskussioner mellan himlen och jord samt ditt intresse för kinesisk kultur och mat. **Christine Wennerås** för att du alltid har varit så trevlig samt dina fina färger som tröstat mina ögon.

Alla doktorand kollegor, **Johanna** som har fraktat proviant till mig ända från Kina. Jag kan skriva flera rader om hur snäll, trevlig, hjälpsam och klok Johanna är, men framförallt är hon en ängel mot sina patienter. **Kicki** för du är så rolig och klok, **Johan** för att du står ut med mina "lille" trakasserier, egentligen vill jag bara få dig att le. **Maria Bankvall** för allt samarbete med RAS, för din tålmodighet, noggrannhet och din starka vilja att lära dig T-RFLP. **Huamei** och **Li** för allt kinesiskt snack och trevligt lunchsällskap.

谢谢父母的养育之恩！妈谢谢您让我上最好的英文学校。您一直都是我的好榜样！

Mina fina barn **Bella** och **Felicia** som är tröst och glädje i mitt liv. Min trygga man **Mikael**, som villkorslöst ställt upp och stöttat mig genom hela resan, utan dig hade jag aldrig kommit så här långt!

★ ★ ★ ★ ★ ★ ★

## REFERENCES

1. Parte AC: **LPSN--list of prokaryotic names with standing in nomenclature**. *Nucleic Acids Res* 2014, **42**(Database issue):D613-616.
2. **Book**. In: *International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision*. Edited by Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA. Washington (DC); 1992.
3. Pace NR: **A molecular view of microbial diversity and the biosphere**. *Science* 1997, **276**(5313):734-740.
4. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G *et al*: **ARB: a software environment for sequence data**. *Nucleic Acids Res* 2004, **32**(4):1363-1371.
5. Hugenholtz P: **Exploring prokaryotic diversity in the genomic era**. *Genome biology* 2002, **3**(2):REVIEWS0003.
6. Oren A, Garrity GM: **Then and now: a systematic review of the systematics of prokaryotes in the last 80 years**. *Antonie Van Leeuwenhoek* 2014, **106**(1):43-56.
7. Rossello-Mora R, Amann R: **The species concept for prokaryotes**. *FEMS microbiology reviews* 2001, **25**(1):39-67.
8. Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP: **The bacterial species challenge: making sense of genetic and ecological diversity**. *Science* 2009, **323**(5915):741-746.
9. Johnson JR: **Shigella and Escherichia coli at the crossroads: machiavellian masqueraders or taxonomic treachery?** *J Med Microbiol* 2000, **49**(7):583-585.
10. Murray RG, Stackebrandt E: **Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes**. *International journal of systematic bacteriology* 1995, **45**(1):186-187.
11. Stackebrandt E EJ: **Taxonomic parameters revisited: tarnished gold standards**. *Microbiol Today* 2006, **33**:152-155.

12. Kim M, Oh HS, Park SC, Chun J: **Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes.** *International journal of systematic and evolutionary microbiology* 2014, **64**(Pt 2):346-351.
13. **Bergey's Manual of Systematic Bacteriology**, vol. 2: Williams & Wilkins; 1986.
14. Lemon KP, Earl AM, Vlamakis HC, Aguilar C, Kolter R: **Biofilm development with an emphasis on *Bacillus subtilis*.** *Current topics in microbiology and immunology* 2008, **322**:1-16.
15. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG: **The human oral microbiome.** *J Bacteriol* 2010, **192**(19):5002-5017.
16. Jenkinson HF, Lamont RJ: **Oral microbial communities in sickness and in health.** *Trends in microbiology* 2005, **13**(12):589-595.
17. Human Microbiome Project C: **Structure, function and diversity of the healthy human microbiome.** *Nature* 2012, **486**(7402):207-214.
18. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE: **Defining the normal bacterial flora of the oral cavity.** *Journal of clinical microbiology* 2005, **43**(11):5721-5732.
19. Savage DC: **Microbial ecology of the gastrointestinal tract.** *Annu Rev Microbiol* 1977, **31**:107-133.
20. Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, Perez-Perez G, Blaser MJ, Relman DA: **Molecular analysis of the bacterial microbiota in the human stomach.** *Proc Natl Acad Sci U S A* 2006, **103**(3):732-737.
21. Ou G, Hedberg M, Horstedt P, Baranov V, Forsberg G, Drobni M, Sandstrom O, Wai SN, Johansson I, Hammarstrom ML *et al*: **Proximal small intestinal microbiota and identification of rod-shaped bacteria associated with childhood celiac disease.** *The American journal of gastroenterology* 2009, **104**(12):3058-3067.

22. Cheng J, Kalliomaki M, Heilig HG, Palva A, Lahteenoja H, de Vos WM, Salojarvi J, Satokari R: **Duodenal microbiota composition and mucosal homeostasis in pediatric celiac disease.** *BMC gastroenterology* 2013, **13**:113.
23. Finegold SM: **Normal human intestinal flora.** *Annali dell'Istituto superiore di sanita* 1986, **22**(3):731-737.
24. O'Hara AM, Shanahan F: **The gut flora as a forgotten organ.** *EMBO reports* 2006, **7**(7):688-693.
25. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA: **Diversity of the human intestinal microbial flora.** *Science* 2005, **308**(5728):1635-1638.
26. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R: **Bacterial community variation in human body habitats across space and time.** *Science* 2009, **326**(5960):1694-1697.
27. Yatsunencko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP *et al*: **Human gut microbiome viewed across age and geography.** *Nature* 2012, **486**(7402):222-227.
28. Finegold SM, Attebery HR, Sutter VL: **Effect of diet on human fecal flora: comparison of Japanese and American diets.** *The American journal of clinical nutrition* 1974, **27**(12):1456-1469.
29. Finegold SM SV, Mathisen GE.: **Human intestinal microflora in health and disease-normal indigenous intestinal flora**, vol. 3-31. London: Academic Press; 1983.
30. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T *et al*: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**(7285):59-65.
31. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E *et al*: **Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons.** *Genome Res* 2011, **21**(3):494-504.
32. Hayashi H, Sakamoto M, Benno Y: **Phylogenetic analysis of the human gut microbiota using 16S rDNA clone**

- libraries and strictly anaerobic culture-based methods.** *Microbiol Immunol* 2002, **46**(8):535-548.
33. Duncan SH, Louis P, Flint HJ: **Cultivable bacterial diversity from the human colon.** *Lett Appl Microbiol* 2007, **44**(4):343-350.
  34. Zoetendal EG, Rajilic-Stojanovic M, de Vos WM: **High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota.** *Gut* 2008, **57**(11):1605-1615.
  35. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI: **Viruses in the faecal microbiota of monozygotic twins and their mothers.** *Nature* 2010, **466**(7304):334-338.
  36. Kononen E: **Development of oral bacterial flora in young children.** *Ann Med* 2000, **32**(2):107-112.
  37. Rotimi VO, Duerden BI: **The development of the bacterial flora in normal neonates.** *J Med Microbiol* 1981, **14**(1):51-62.
  38. McCarthy C, Snyder ML, Parker RB: **The Indigenous Oral Flora of Man. I. The Newborn to the 1-Year-Old Infant.** *Archives of oral biology* 1965, **10**:61-70.
  39. Kononen E, Asikainen S, Jousimies-Somer H: **The early colonization of gram-negative anaerobic bacteria in edentulous infants.** *Oral microbiology and immunology* 1992, **7**(1):28-31.
  40. Socransky SS, Manganiello SD: **The oral microbiota of man from birth to senility.** *Journal of periodontology* 1971, **42**(8):485-496.
  41. Bradshaw DJ, Marsh PD, Watson GK, Allison C: **Role of *Fusobacterium nucleatum* and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration.** *Infect Immun* 1998, **66**(10):4729-4732.
  42. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO: **Development of the human infant intestinal microbiota.** *PLoS Biol* 2007, **5**(7):e177.

43. Tapiainen T, Ylitalo S, Eerola E, Uhari M: **Dynamics of gut colonization and source of intestinal flora in healthy newborn infants.** *Apmis* 2006, **114**(11):812-817.
44. Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS, Sears MR, Becker AB, Scott JA, Kozyrskyj AL *et al*: **Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months.** *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2013, **185**(5):385-394.
45. Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C: **Mode of delivery affects the bacterial community in the newborn gut.** *Early human development* 2010, **86 Suppl 1**:13-15.
46. Brook I, Barrett CT, Brinkman CR, 3rd, Martin WJ, Finegold SM: **Aerobic and anaerobic bacterial flora of the maternal cervix and newborn gastric fluid and conjunctiva: a prospective study.** *Pediatrics* 1979, **63**(3):451-455.
47. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R: **Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.** *Proc Natl Acad Sci U S A* 2010, **107**(26):11971-11975.
48. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, van den Brandt PA, Stobberingh EE: **Factors influencing the composition of the intestinal microbiota in early infancy.** *Pediatrics* 2006, **118**(2):511-521.
49. Adlerberth I, Strachan DP, Matricardi PM, Ahrne S, Orfei L, Aberg N, Perkin MR, Tripodi S, Hesselmar B, Saalman R *et al*: **Gut microbiota and development of atopic eczema in 3 European birth cohorts.** *J Allergy Clin Immunol* 2007, **120**(2):343-350.
50. Adlerberth I, Wold AE: **Establishment of the gut microbiota in Western infants.** *Acta Paediatr* 2009, **98**(2):229-238.
51. El Aidy S, Van den Abbeele P, Van de Wiele T, Louis P, Kleerebezem M: **Intestinal colonization: how key microbial players become established in this dynamic**

- process: microbial metabolic activities and the interplay between the host and microbes.** *BioEssays : news and reviews in molecular, cellular and developmental biology* 2013, **35**(10):913-923.
52. Ellis-Pegler RB, Crabtree C, Lambert HP: **The faecal flora of children in the United Kingdom.** *J Hyg (Lond)* 1975, **75**(1):135-142.
  53. Adlerberth I. HLÅ, Wold AE., : **Development of the gastrointestinal tract chapter 17 Ontogeny of the intestinal flora.** Ontario, Canada: B. C. decker inc. Hamilton, Ontario; 1999.
  54. Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vazquez-Baeza Y, Jansson JK, Gordon JI, Knight R: **Meta-analyses of studies of the human microbiota.** *Genome Res* 2013, **23**(10):1704-1714.
  55. Elberse KE, Nunes S, Sa-Leao R, van der Heide HG, Schouls LM: **Multiple-locus variable number tandem repeat analysis for Streptococcus pneumoniae: comparison with PFGE and MLST.** *PLoS One* 2011, **6**(5):e19668.
  56. Flint HJ, Duncan SH, Scott KP, Louis P: **Interactions and competition within the microbial community of the human colon: links between diet and health.** *Environ Microbiol* 2007, **9**(5):1101-1111.
  57. Adams E: **Studies in gram staining.** *Stain technology* 1975, **50**(4):227-231.
  58. Morrison WR, Smith LM: **Preparation of Fatty Acid Methyl Esters and Dimethylacetals from Lipids with Boron Fluoride--Methanol.** *Journal of lipid research* 1964, **5**:600-608.
  59. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D: **Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.** *Clin Infect Dis* 2009, **49**(4):543-551.
  60. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H: **Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.** *Cold Spring Harbor symposia on quantitative biology* 1986, **51 Pt 1**:263-273.

61. Woese CR: **Bacterial evolution.** *Microbiological reviews* 1987, **51**(2):221-271.
62. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ: **Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes.** *Appl Environ Microbiol* 2008, **74**(8):2461-2470.
63. Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.** *Appl Environ Microbiol* 2007, **73**(16):5261-5267.
64. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF: **Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons.** *J Bacteriol* 2004, **186**(9):2629-2635.
65. Ueda K, Seki T, Kudo T, Yoshida T, Kataoka M: **Two distinct mechanisms cause heterogeneity of 16S rRNA.** *J Bacteriol* 1999, **181**(1):78-82.
66. Jumpstart Consortium Human Microbiome Project Data Generation Working G: **Evaluation of 16S rDNA-based community profiling for human microbiome research.** *PLoS One* 2012, **7**(6):e39315.
67. Farris MH, Olson JB: **Detection of Actinobacteria cultivated from environmental samples reveals bias in universal primers.** *Lett Appl Microbiol* 2007, **45**(4):376-381.
68. Liu WT, Marsh TL, Cheng H, Forney LJ: **Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA.** *Appl Environ Microbiol* 1997, **63**(11):4516-4522.
69. Sjoberg F, Nowrouzian F, Rangel I, Hannoun C, Moore E, Adlerberth I, Wold AE: **Comparison between terminal-restriction fragment length polymorphism (T-RFLP) and quantitative culture for analysis of infants' gut microbiota.** *J Microbiol Methods* 2013, **94**(1):37-46.
70. Hahn M, Wilhelm J, Pingoud A: **Influence of fluorophore dye labels on the migration behavior of polymerase chain reaction--amplified short tandem repeats during**

- denaturing capillary electrophoresis. *Electrophoresis* 2001, **22**(13):2691-2700.
71. Rajilic-Stojanovic M, Heilig HG, Molenaar D, Kajander K, Surakka A, Smidt H, de Vos WM: **Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults.** *Environ Microbiol* 2009, **11**(7):1736-1751.
  72. Wilson KH, Wilson WJ, Radosevich JL, DeSantis TZ, Viswanathan VS, Kuczmariski TA, Andersen GL: **High-density microarray of small-subunit ribosomal DNA probes.** *Appl Environ Microbiol* 2002, **68**(5):2535-2541.
  73. Group NHW, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA *et al*: **The NIH Human Microbiome Project.** *Genome Res* 2009, **19**(12):2317-2323.
  74. Walker AW, Duncan SH, Louis P, Flint HJ: **Phylogeny, culturing, and metagenomics of the human gut microbiota.** *Trends in microbiology* 2014, **22**(5):267-274.
  75. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K: **Optimization and evaluation of single-cell whole-genome multiple displacement amplification.** *Human mutation* 2006, **27**(5):496-503.
  76. McLean JS, Lombardo MJ, Ziegler MG, Novotny M, Yee-Greenbaum J, Badger JH, Tesler G, Nurk S, Lesin V, Bami D *et al*: **Genome of the pathogen *Porphyromonas gingivalis* recovered from a biofilm in a hospital sink using a high-throughput single-cell genomics platform.** *Genome Res* 2013, **23**(5):867-877.
  77. Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, Daroub SH, Camargo FA, Farmerie WG, Triplett EW: **Pyrosequencing enumerates and contrasts soil microbial diversity.** *ISME J* 2007, **1**(4):283-290.
  78. Schloss PD: **The effects of alignment quality, distance calculation method, sequence filtering, and region on**

- the analysis of 16S rRNA gene-based studies.** *PLoS Comput Biol* 2010, **6**(7):e1000844.
79. Patin NV, Kunin V, Lidstrom U, Ashby MN: **Effects of OTU clustering and PCR artifacts on microbial diversity estimates.** *Microbial ecology* 2013, **65**(3):709-719.
80. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT: **Accurate determination of microbial diversity from 454 pyrosequencing data.** *Nat Methods* 2009, **6**(9):639-641.
81. Bomar L, Maltz M, Colston S, Graf J: **Directed culturing of microorganisms using metatranscriptomics.** *mBio* 2011, **2**(2):e00012-00011.
82. Renesto P, Crapoulet N, Ogata H, La Scola B, Vestris G, Claverie JM, Raoult D: **Genome-based design of a cell-free culture medium for *Tropheryma whipplei*.** *Lancet* 2003, **362**(9382):447-449.
83. Karami N, Martner A, Enne VI, Swerkersson S, Adlerberth I, Wold AE: **Transfer of an ampicillin resistance gene between two *Escherichia coli* strains in the bowel microbiota of an infant treated with antibiotics.** *The Journal of antimicrobial chemotherapy* 2007, **60**(5):1142-1145.
84. Adlerberth I, Huang H, Lindberg E, Aberg N, Hesselmar B, Saalman R, Nord CE, Wold AE, Weintraub A: **Toxin-producing *Clostridium difficile* strains as long-term gut colonizers in healthy infants.** *J Clin Microbiol* 2014, **52**(1):173-179.
85. Brandtzaeg P: **Function of mucosa-associated lymphoid tissue in antibody formation.** *Immunological investigations* 2010, **39**(4-5):303-355.
86. Kasahara K, Matsumura Y, Ui K, Kasahara K, Komatsu Y, Mikasa K, Kita E: **Intranasal priming of newborn mice with microbial extracts increases opsonic factors and mature CD11c+ cells in the airway.** *American journal of physiology Lung cellular and molecular physiology* 2012, **303**(9):L834-843.
87. Siegmund K, Ruckert B, Ouaked N, Burgler S, Speiser A, Akdis CA, Schmidt-Weber CB: **Unique phenotype of**

- human tonsillar and in vitro-induced FOXP3+CD8+ T cells. *Journal of immunology (Baltimore, Md : 1950)* 2009, **182**(4):2124-2130.**
88. Crabbe PA, Bazin H, Eyssen H, Heremans JF: **The normal microbial flora as a major stimulus for proliferation of plasma cells synthesizing IgA in the gut. The germ-free intestinal tract.** *Int Arch Allergy Appl Immunol* 1968, **34**(4):362-375.
89. Steffen EK, Berg RD, Deitch EA: **Comparison of translocation rates of various indigenous bacteria from the gastrointestinal tract to the mesenteric lymph node.** *J Infect Dis* 1988, **157**(5):1032-1038.
90. Berg RD: **Bacterial translocation from the gastrointestinal tract.** *Adv Exp Med Biol* 1999, **473**:11-30.
91. Macdonald TT, Monteleone G: **Immunity, inflammation, and allergy in the gut.** *Science* 2005, **307**(5717):1920-1925.
92. Carter PB, Pollard M: **Host responses to "normal" microbial flora in germ-free mice.** *Journal of the Reticuloendothelial Society* 1971, **9**(6):580-587.
93. Ostman S, Rask C, Wold AE, Hultkrantz S, Telemo E: **Impaired regulatory T cell function in germ-free mice.** *European journal of immunology* 2006, **36**(9):2336-2346.
94. Von Bultzingslowen I, Adlerberth I, Wold AE, Dahlen G, Jontell M: **Oral and intestinal microflora in 5-fluorouracil treated rats, translocation to cervical and mesenteric lymph nodes and effects of probiotic bacteria.** *Oral microbiology and immunology* 2003, **18**(5):278-284.
95. Pabst O, Mowat AM: **Oral tolerance to food protein.** *Mucosal immunology* 2012, **5**(3):232-239.
96. Karlsson M, Lundin S, Dahlgren U, Kahu H, Pettersson I, Telemo E: **"Tolerosomes" are produced by intestinal epithelial cells.** *Eur J Immunol* 2001, **31**(10):2892-2900.
97. Moreau MC, Corthier G: **Effect of the gastrointestinal microflora on induction and maintenance of oral tolerance to ovalbumin in C3H/HeJ mice.** *Infect Immun* 1988, **56**(10):2766-2768.

98. Koropatkin NM, Cameron EA, Martens EC: **How glycan metabolism shapes the human gut microbiota.** *Nature reviews Microbiology* 2012, **10**(5):323-335.
99. Wong JM, de Souza R, Kendall CW, Emam A, Jenkins DJ: **Colonic health: fermentation and short chain fatty acids.** *Journal of clinical gastroenterology* 2006, **40**(3):235-243.
100. Louis P, Flint HJ: **Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine.** *FEMS microbiology letters* 2009, **294**(1):1-8.
101. Roediger WE: **Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man.** *Gut* 1980, **21**(9):793-798.
102. Velazquez OC, Lederer HM, Rombeau JL: **Butyrate and the colonocyte. Production, absorption, metabolism, and therapeutic implications.** *Adv Exp Med Biol* 1997, **427**:123-134.
103. Segain JP, Raingeard de la Bletiere D, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottiere HM, Galmiche JP: **Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease.** *Gut* 2000, **47**(3):397-403.
104. Avivi-Green C, Polak-Charcon S, Madar Z, Schwartz B: **Apoptosis cascade proteins are regulated in vivo by high intracolonic butyrate concentration: correlation with colon cancer inhibition.** *Oncology research* 2000, **12**(2):83-95.
105. Kay AB: **100 years of 'Allergy': can von Pirquet's word be rescued?** *Clin Exp Allergy* 2006, **36**(5):555-559.
106. Bach JF: **The effect of infections on susceptibility to autoimmune and allergic diseases.** *The New England journal of medicine* 2002, **347**(12):911-920.
107. Aberg N, Hesselmar B, Aberg B, Eriksson B: **Increase of asthma, allergic rhinitis and eczema in Swedish schoolchildren between 1979 and 1991.** *Clin Exp Allergy* 1995, **25**(9):815-819.

108. Johansson SG, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF, Motala C, Ortega Martell JA, Platts-Mills TA, Ring J *et al*: **Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003.** *J Allergy Clin Immunol* 2004, **113**(5):832-836.
109. van der Hulst AE, Klip H, Brand PL: **Risk of developing asthma in young children with atopic eczema: a systematic review.** *J Allergy Clin Immunol* 2007, **120**(3):565-569.
110. Spergel JM: **Epidemiology of atopic dermatitis and atopic march in children.** *Immunology and allergy clinics of North America* 2010, **30**(3):269-280.
111. Hesselmar B, Aberg N, Aberg B, Eriksson B, Bjorksten B: **Does early exposure to cat or dog protect against later allergy development?** *Clin Exp Allergy* 1999, **29**(5):611-617.
112. Matricardi PM, Rosmini F, Riondino S, Fortini M, Ferrigno L, Rapicetta M, Bonini S: **Exposure to foodborne and orofecal microbes versus airborne viruses in relation to atopy and allergic asthma: epidemiological study.** *BMJ* 2000, **320**(7232):412-417.
113. Riedler J, Braun-Fahrlander C, Eder W, Schreuer M, Waser M, Maisch S, Carr D, Schierl R, Nowak D, von Mutius E: **Exposure to farming in early life and development of asthma and allergy: a cross-sectional survey.** *Lancet* 2001, **358**(9288):1129-1133.
114. Strachan DP: **Hay fever, hygiene, and household size.** *Bmj* 1989, **299**(6710):1259-1260.
115. Strachan DP: **Family size, infection and atopy: the first decade of the "hygiene hypothesis".** *Thorax* 2000, **55 Suppl 1**:S2-10.
116. Bager P, Melbye M, Rostgaard K, Benn CS, Westergaard T: **Mode of delivery and risk of allergic rhinitis and asthma.** *J Allergy Clin Immunol* 2003, **111**(1):51-56.
117. McKeever TM, Lewis SA, Smith C, Hubbard R: **Mode of delivery and risk of developing allergic disease.** *J Allergy Clin Immunol* 2002, **109**(5):800-802.

118. Adlerberth I, Carlsson B, de Man P, Jalil F, Khan SR, Larsson P, Mellander L, Svanborg C, Wold AE, Hanson LA: **Intestinal colonization with Enterobacteriaceae in Pakistani and Swedish hospital-delivered infants.** *Acta paediatrica Scandinavica* 1991, **80**(6-7):602-610.
119. Adlerberth I, Jalil F, Carlsson B, Mellander L, Hanson LA, Larsson P, Khalil K, Wold AE: **High turnover rate of Escherichia coli strains in the intestinal flora of infants in Pakistan.** *Epidemiol Infect* 1998, **121**(3):587-598.
120. Kuhn I, Tullus K, Mollby R: **Colonization and persistence of Escherichia coli phenotypes in the intestines of children aged 0 to 18 months.** *Infection* 1986, **14**(1):7-12.
121. Nowrouzian F, Hesselmar B, Saalman R, Strannegard IL, Aberg N, Wold AE, Adlerberth I: **Escherichia coli in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage.** *Pediatr Res* 2003, **54**(1):8-14.
122. Wang M, Karlsson C, Olsson C, Adlerberth I, Wold AE, Strachan DP, Martricardi PM, Aberg N, Perkin MR, Tripodi S *et al*: **Reduced diversity in the early fecal microbiota of infants with atopic eczema.** *J Allergy Clin Immunol* 2008, **121**(1):129-134.
123. Bisgaard H, Li N, Bonnelykke K, Chawes BL, Skov T, Paludan-Muller G, Stokholm J, Smith B, Krogfelt KA: **Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age.** *J Allergy Clin Immunol* 2011, **128**(3):646-652 e641-645.
124. Ismail IH, Oppedisano F, Joseph SJ, Boyle RJ, Licciardi PV, Robins-Browne RM, Tang ML: **Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants.** *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 2012, **23**(7):674-681.
125. Abrahamsson TR, Jakobsson HE, Andersson AF, Bjorksten B, Engstrand L, Jenmalm MC: **Low gut microbiota**

- diversity in early infancy precedes asthma at school age.** *Clin Exp Allergy* 2014, **44**(6):842-850.
126. Arbes SJ, Jr., Sever ML, Vaughn B, Cohen EA, Zeldin DC: **Oral pathogens and allergic disease: results from the Third National Health and Nutrition Examination Survey.** *J Allergy Clin Immunol* 2006, **118**(5):1169-1175.
  127. Navarro S, Cossalter G, Chiavaroli C, Kanda A, Fleury S, Lazzari A, Cazareth J, Sparwasser T, Dombrowicz D, Glaichenhaus N *et al*: **The oral administration of bacterial extracts prevents asthma via the recruitment of regulatory T cells to the airways.** *Mucosal immunology* 2011, **4**(1):53-65.
  128. Card JW, Carey MA, Voltz JW, Bradbury JA, Ferguson CD, Cohen EA, Schwartz S, Flake GP, Morgan DL, Arbes SJ, Jr. *et al*: **Modulation of allergic airway inflammation by the oral pathogen *Porphyromonas gingivalis*.** *Infect Immun* 2010, **78**(6):2488-2496.
  129. Khor B, Gardet A, Xavier RJ: **Genetics and pathogenesis of inflammatory bowel disease.** *Nature* 2011, **474**(7351):307-317.
  130. Hildebrand H, Finkel Y, Grahnquist L, Lindholm J, Ekblom A, Askling J: **Changing pattern of paediatric inflammatory bowel disease in northern Stockholm 1990-2001.** *Gut* 2003, **52**(10):1432-1434.
  131. Tsironi E, Feakins RM, Probert CS, Rampton DS, Phil D: **Incidence of inflammatory bowel disease is rising and abdominal tuberculosis is falling in Bangladeshis in East London, United Kingdom.** *The American journal of gastroenterology* 2004, **99**(9):1749-1755.
  132. Yang SK, Loftus EV, Jr., Sandborn WJ: **Epidemiology of inflammatory bowel disease in Asia.** *Inflammatory bowel diseases* 2001, **7**(3):260-270.
  133. Feillet H, Bach JF: **Increased incidence of inflammatory bowel disease: the price of the decline of infectious burden?** *Current opinion in gastroenterology* 2004, **20**(6):560-564.
  134. Radon K, Windstetter D, Poluda AL, Mueller B, von Mutius E, Koletzko S: **Contact with farm animals in early life and**

- juvenile inflammatory bowel disease: a case-control study.** *Pediatrics* 2007, **120**(2):354-361.
135. Taurog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, Fernandez-Sueiro JL, Balish E, Hammer RE: **The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats.** *The Journal of experimental medicine* 1994, **180**(6):2359-2364.
136. Aranda R, Sydora BC, McAllister PL, Binder SW, Yang HY, Targan SR, Kronenberg M: **Analysis of intestinal lymphocytes in mouse colitis mediated by transfer of CD4+, CD45RBhigh T cells to SCID recipients.** *Journal of immunology (Baltimore, Md : 1950)* 1997, **158**(7):3464-3473.
137. Medina C, Santana A, Llopis M, Paz-Cabrera MC, Antolin M, Mourelle M, Guarner F, Vilaseca J, Gonzalez C, Salas A *et al*: **Induction of colonic transmural inflammation by Bacteroides fragilis: implication of matrix metalloproteinases.** *Inflammatory bowel diseases* 2005, **11**(2):99-105.
138. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, Blugeon S, Bridonneau C, Furet JP, Corthier G *et al*: **Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients.** *Proc Natl Acad Sci U S A* 2008, **105**(43):16731-16736.
139. Van Immerseel F, Ducatelle R, De Vos M, Boon N, Van De Wiele T, Verbeke K, Rutgeerts P, Sas B, Louis P, Flint HJ: **Butyric acid-producing anaerobic bacteria as a novel probiotic treatment approach for inflammatory bowel disease.** *J Med Microbiol* 2010, **59**(Pt 2):141-143.
140. Nwosu FC, Thorkildsen LT, Avershina E, Ricanek P, Perminow G, Brackmann S, Vatn MH, Rudi K: **Age-dependent fecal bacterial correlation to inflammatory bowel disease for newly diagnosed untreated children.** *Gastroenterology research and practice* 2013, **2013**:302398.
141. Thorkildsen LT, Nwosu FC, Avershina E, Ricanek P, Perminow G, Brackmann S, Vatn MH, Rudi K: **Dominant fecal microbiota in newly diagnosed untreated**

- inflammatory bowel disease patients. *Gastroenterology research and practice* 2013, **2013**:636785.**
142. Conte MP, Schippa S, Zamboni I, Penta M, Chiarini F, Seganti L, Osborn J, Falconieri P, Borrelli O, Cucchiara S: **Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease.** *Gut* 2006, **55**(12):1760-1767.
  143. Bibiloni R, Mangold M, Madsen KL, Fedorak RN, Tannock GW: **The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients.** *J Med Microbiol* 2006, **55**(Pt 8):1141-1149.
  144. Kellermayer R, Mir SA, Nagy-Szakal D, Cox SB, Dowd SE, Kaplan JL, Sun Y, Reddy S, Bronsky J, Winter HS: **Microbiota separation and C-reactive protein elevation in treatment-naive pediatric granulomatous Crohn disease.** *Journal of pediatric gastroenterology and nutrition* 2012, **55**(3):243-250.
  145. Hansen R, Russell RK, Reiff C, Louis P, McIntosh F, Berry SH, Mukhopadhyaya I, Bisset WM, Barclay AR, Bishop J *et al*: **Microbiota of de-novo pediatric IBD: increased Faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis.** *The American journal of gastroenterology* 2012, **107**(12):1913-1922.
  146. Hansen R, Berry SH, Mukhopadhyaya I, Thomson JM, Saunders KA, Nicholl CE, Bisset WM, Loganathan S, Mahdi G, Kastner-Cole D *et al*: **The microaerophilic microbiota of de-novo paediatric inflammatory bowel disease: the BISCUIT study.** *PLoS One* 2013, **8**(3):e58825.
  147. Wagner J, Short K, Catto-Smith AG, Cameron DJ, Bishop RF, Kirkwood CD: **Identification and characterisation of Pseudomonas 16S ribosomal DNA from ileal biopsies of children with Crohn's disease.** *PLoS One* 2008, **3**(10):e3578.
  148. Kirkwood CD, Wagner J, Boniface K, Vaughan J, Michalski WP, Catto-Smith AG, Cameron DJ, Bishop RF: **Mycobacterium avium subspecies paratuberculosis in**

- children with early-onset Crohn's disease. *Inflammatory bowel diseases* 2009, 15(11):1643-1655.**
149. Iebba V, Santangelo F, Totino V, Nicoletti M, Gagliardi A, De Biase RV, Cucchiara S, Nencioni L, Conte MP, Schippa S: **Higher prevalence and abundance of *Bdellovibrio bacteriovorus* in the human gut of healthy subjects.** *PLoS One* 2013, **8(4):e61608.**
  150. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M *et al*: **The treatment-naive microbiome in new-onset Crohn's disease.** *Cell host & microbe* 2014, **15(3):382-392.**
  151. Salonen A, Nikkila J, Jalanka-Tuovinen J, Immonen O, Rajilic-Stojanovic M, Kekkonen RA, Palva A, de Vos WM: **Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis.** *J Microbiol Methods* 2010, **81(2):127-134.**
  152. Gutell RR: **Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994.** *Nucleic Acids Res* 1994, **22(17):3502-3507.**
  153. Abdo Z, Schuette UM, Bent SJ, Williams CJ, Forney LJ, Joyce P: **Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes.** *Environ Microbiol* 2006, **8(5):929-938.**
  154. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI *et al*: **QIIME allows analysis of high-throughput community sequencing data.** *Nat Methods* 2010, **7(5):335-336.**
  155. Edgar RC: **Search and clustering orders of magnitude faster than BLAST.** *Bioinformatics* 2010, **26(19):2460-2461.**
  156. Nelun Barfod M, Magnusson K, Lexner MO, Blomqvist S, Dahlen G, Twetman S: **Oral microflora in infants delivered vaginally and by caesarean section.** *Int J Paediatr Dent* 2011, **21(6):401-406.**

157. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP *et al*: **Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes.** *DNA Res* 2007, **14**(4):169-181.
158. Kibe R, Sakamoto M, Hayashi H, Yokota H, Benno Y: **Maturation of the murine cecal microbiota as revealed by terminal restriction fragment length polymorphism and 16S rRNA gene clone libraries.** *FEMS microbiology letters* 2004, **235**(1):139-146.
159. Ozutsumi Y, Hayashi H, Sakamoto M, Itabashi H, Benno Y: **Culture-independent analysis of fecal microbiota in cattle.** *Biosci Biotechnol Biochem* 2005, **69**(9):1793-1797.
160. Kent AD, Smith DJ, Benson BJ, Triplett EW: **Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities.** *Appl Environ Microbiol* 2003, **69**(11):6768-6776.
161. Pandey J, Ganesan K, Jain RK: **Variations in T-RFLP profiles with differing chemistries of fluorescent dyes used for labeling the PCR primers.** *J Microbiol Methods* 2007, **68**(3):633-638.
162. Kaplan CW, Kitts CL: **Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content.** *J Microbiol Methods* 2003, **54**(1):121-125.
163. Nieminen TT, Vihavainen E, Paloranta A, Lehto J, Paulin L, Auvinen P, Solismaa M, Bjorkroth KJ: **Characterization of psychrotrophic bacterial communities in modified atmosphere-packed meat with terminal restriction fragment length polymorphism.** *Int J Food Microbiol* 2011, **144**(3):360-366.
164. Fukushima M, Kakinuma K, Kawaguchi R: **Phylogenetic analysis of Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene sequence.** *Journal of clinical microbiology* 2002, **40**(8):2779-2785.
165. Pham HN, Ohkusu K, Mishima N, Noda M, Monir Shah M, Sun X, Hayashi M, Ezaki T: **Phylogeny and species**

- identification of the family Enterobacteriaceae based on dnaJ sequences.** *Diagnostic microbiology and infectious disease* 2007, **58**(2):153-161.
166. Lundell AC, Hesselmar B, Nordstrom I, Saalman R, Karlsson H, Lindberg E, Aberg N, Adlerberth I, Wold AE, Rudin A: **High circulating immunoglobulin A levels in infants are associated with intestinal toxigenic Staphylococcus aureus and a lower frequency of eczema.** *Clin Exp Allergy* 2009, **39**(5):662-670.
167. Lundell AC, Adlerberth I, Lindberg E, Karlsson H, Ekberg S, Aberg N, Saalman R, Hock B, Steinkasserer A, Hesselmar B *et al*: **Increased levels of circulating soluble CD14 but not CD83 in infants are associated with early intestinal colonization with Staphylococcus aureus.** *Clin Exp Allergy* 2007, **37**(1):62-71.
168. Abrahamsson TR, Jakobsson HE, Andersson AF, Bjorksten B, Engstrand L, Jenmalm MC: **Low diversity of the gut microbiota in infants with atopic eczema.** *J Allergy Clin Immunol* 2012, **129**(2):434-440 e432.
169. Moore WE, Holdeman LV: **Human fecal flora: the normal flora of 20 Japanese-Hawaiians.** *Appl Microbiol* 1974, **27**(5):961-979.
170. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M *et al*: **Microbial culturomics: paradigm shift in the human gut microbiome study.** *Clin Microbiol Infect* 2012, **18**(12):1185-1193.
171. Lagier JC, Million M, Hugon P, Armougom F, Raoult D: **Human gut microbiota: repertoire and variations.** *Frontiers in cellular and infection microbiology* 2012, **2**:136.
172. Cutler CW, Jotwani R: **Dendritic cells at the oral mucosal interface.** *J Dent Res* 2006, **85**(8):678-689.
173. Eriksson K, Ahlfors E, George-Chandy A, Kaiserlian D, Czerkinsky C: **Antigen presentation in the murine oral epithelium.** *Immunology* 1996, **88**(1):147-152.
174. Van Hoogstraten IM, Boos C, Boden D, Von Blomberg ME, Scheper RJ, Kraal G: **Oral induction of tolerance to nickel**

- sensitization in mice.** *The Journal of investigative dermatology* 1993, **101**(1):26-31.
175. Xenoulis PG, Palculict B, Allenspach K, Steiner JM, Van House AM, Suchodolski JS: **Molecular-phylogenetic characterization of microbial communities imbalances in the small intestine of dogs with inflammatory bowel disease.** *FEMS Microbiol Ecol* 2008, **66**(3):579-589.
  176. Suchodolski JS, Dowd SE, Wilke V, Steiner JM, Jergens AE: **16S rRNA gene pyrosequencing reveals bacterial dysbiosis in the duodenum of dogs with idiopathic inflammatory bowel disease.** *PLoS One* 2012, **7**(6):e39333.
  177. Gossling J, Slack JM: **Predominant gram-positive bacteria in human feces: numbers, variety, and persistence.** *Infect Immun* 1974, **9**(4):719-729.
  178. Rey FE, Gonzalez MD, Cheng J, Wu M, Ahern PP, Gordon JI: **Metabolic niche of a prominent sulfate-reducing human gut bacterium.** *Proc Natl Acad Sci U S A* 2013, **110**(33):13582-13587.
  179. Malinen E, Krogius-Kurikka L, Lyra A, Nikkila J, Jaaskelainen A, Rinttila T, Vilpponen-Salmela T, von Wright AJ, Palva A: **Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome.** *World journal of gastroenterology : WJG* 2010, **16**(36):4532-4540.
  180. Peach S, Fernandez F, Johnson K, Drasar BS: **The non-sporing anaerobic bacteria in human faeces.** *J Med Microbiol* 1974, **7**(2):213-221.
  181. Benno Y, Suzuki K, Suzuki K, Narisawa K, Bruce WR, Mitsuoka T: **Comparison of the fecal microflora in rural Japanese and urban Canadians.** *Microbiol Immunol* 1986, **30**(6):521-532.
  182. Moore WE, Cato EP, Holdeman LV: **Some current concepts in intestinal bacteriology.** *The American journal of clinical nutrition* 1978, **31**(10 Suppl):S33-42.
  183. Kato M, Muto Y, Tanaka-Bandoh K, Watanabe K, Ueno K: **Sphingolipid composition in Bacteroides species.** *Anaerobe* 1995, **1**(2):135-139.

184. Wu L, Gabriel CL, Parekh VV, Van Kaer L: **Invariant natural killer T cells: innate-like T cells with potent immunomodulatory activities.** *Tissue antigens* 2009, **73**(6):535-545.
185. An D, Oh SF, Olszak T, Neves JF, Avci FY, Erturk-Hasdemir D, Lu X, Zeissig S, Blumberg RS, Kasper DL: **Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells.** *Cell* 2014, **156**(1-2):123-133.
186. van Baarlen P, Wells JM, Kleerebezem M: **Regulation of intestinal homeostasis and immunity with probiotic lactobacilli.** *Trends in immunology* 2013, **34**(5):208-215.
187. Wang ZK, Yang YS: **Upper gastrointestinal microbiota and digestive diseases.** *World journal of gastroenterology : WJG* 2013, **19**(10):1541-1550.