The development of early life gut microbiota in health and allergic disease

Harm Wopereis
Propositions

1. The rate of gut microbiome maturation in early life influences the risk of developing allergic disease.
   (this thesis)

2. Using the term ‘dysbiosis’ to describe microbial configurations in the context of complex diseases, such as allergy, is misleading.
   (this thesis)

3. If there is anything that distinguishes ‘data science’ from ‘traditional statistics’, it is the urge to learn from data.

4. Experiencing and, above all, creating music benefits your brain.

5. Colour-blindness hinders the modern microbiologist more than ever!

6. A lie may have no legs but has become faster than ever.

Propositions belonging to the thesis, entitled:

The development of early life gut microbiota in health and allergic disease

Harm Wopereis
Wageningen, 22 May 2019
The development of early life gut microbiota in health and allergic disease

Harm Wopereis
Thesis committee

Promotors
Prof. Dr J. Knol
Special Professor Intestinal Microbiology of Early Life
Wageningen University & Research

Prof. Dr W.M. de Vos
Distinguished Professor of Microbiology
Wageningen University & Research

Co-promotor
Dr C. Belzer
Assistant Professor, Laboratory of Microbiology
Wageningen University & Research

Other members
Prof. Dr H.J. Wichers, Wageningen University & Research
Prof. Dr N.G. Papadopoulos, the University of Manchester, United Kingdom
Prof. Dr S. Salminen, University of Turku, Finland
Prof. Dr C. Stanton, University College Cork, Ireland

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food technology, Agrobiotechnology, Nutrition and Health Sciences)
The development of early life gut microbiota in health and allergic disease

Harm Wopereis

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 22 May 2019
at 4 p.m. in the Aula.
Harm Wopereis
The development of early life gut microbiota in health and allergic disease, 184 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summary in English

DOI https://doi.org/10.18174/468890
# Table of Contents

Chapter 1: General introduction & thesis outline ................................................................. 7
Chapter 2: The first thousand days - intestinal microbiology of early life: establishing a symbiosis ............................................................................................................... 25
Chapter 3: Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing ............................................................................................. 51
Chapter 4: Intestinal microbiota in high-risk infants: Effects of prebiotics and role in eczema development ............................................................................................... 73
Chapter 5: A specific synbiotic-containing amino acid-based formula restores gut microbiota in non-IgE mediated cow’s milk allergic infants ....................................... 99
Chapter 6: Gut microbiota from infant with cow’s milk allergy leads to atopic orientation in a murine model .................................................................................................. 125
Chapter 7: General discussion & Summary ............................................................................ 157
Appendices.......................................................................................................................... 175
CHAPTER 1

General introduction & thesis outline
Introduction

The microscopic observations of “putrefying bodies” (fungi) by Robert Hooke and of “animalcules” (bacteria) by Antoni van Leeuwenhoek between 1665 and 1683 uncovered for the first time the microbial world we live in. The observations by these two pioneers in microbiology, both fellows of the prestigious Royal Society of London, were made possible by the first versions of the microscope, which still is a central instrument in microbiology (1). Since these first discoveries, the techniques we use to understand this “invisible” world have evolved: from isolation and cultivation techniques (end of the 19th century), to strictly anaerobic cultivation techniques (late 1960’s) and cultivation-independent molecular techniques (end of the 20th century) (2). Finally, high throughput approaches to study microbial molecules at the level of DNA and RNA (through next generation sequencing), but also at the protein (through proteomics and metaproteomics) and metabolite levels (through metabolomics and metabonomics) revolutionized the field of microbiology. They enabled researchers to describe the composition and activities of the microbial communities found everywhere around us (3). Moreover, these culture-independent methods also allow researchers to explore the microbial diversity and functionality of the large ‘not-yet culturable’ fraction (2). Since the study of microbial communities is a relatively new field in microbiology, it is important to have clear definitions of the terms used. The terms, which have been adopted in this thesis come from those proposed by Marchesi and Ravel (4):

**Microbiota:** all the microorganisms (archaea, bacteria, viruses and microscopic eukaryotes) that live in a specific environment, e.g. the gut or the root nodule of a plant. The study of it mainly relies on the phylogenetic analysis of marker genes, such as those encoding for the prokaryotic 16S ribosomal RNA (bacteria and archaea) and the eukaryotic 18S rRNA. Although viruses are an integral part of the microbiota, they are non-cellular and lack a universal marker gene that can be used for taxonomic classification. However, viruses can be assessed through metagenomics, which is a technique that enables the study of the full microbial genetic content within an environment (5). Most studies so far merely measure or consider the bacterial
members, which is also the scope for this thesis. Formerly, the microbiota was often referred to as “microflora”, but this term is misleading as it implies to describe plants.

**Microbiome**: this term refers to the ecosystem or habitat, including all the microbes and the surrounding environmental conditions (for example: pH, temperature, oxygen pressure). Often this term is limited to describing the collection of microbial genes and genomes, which however is more correctly referred to as metagenome. The prefix “meta” is used to designate the analysis of the collection of molecules from all microbes, instead of from a single organism. The profiling of the RNA-transcripts, the synthesized proteins or metabolites of the community is thus referred to as metatranscriptomics, metaproteomics, and metabolomics (or more correctly: metabolonomics), respectively. These three methods go beyond the analysis of the taxonomic composition and genetic potential of the community and provide insight into the functional activity of the genes that are expressed under certain conditions (6).

**The human microbiome**

The most intensively researched microbial community in the last two decades is probably the human microbiome (7). Microbes are found everywhere the human body is exposed to the outside world; namely the skin, the urogenital tract, the respiratory tract and the digestive tract. The composition of the microbial communities is primarily determined by body habitat but is also unique to each person and relatively stable in adults (8). The average adult has a total estimated number of microbes in the order of magnitude of 10-100 trillion cells; the majority of which is found in the lower digestive (gastro-intestinal) tract (9). Along the gastro-intestinal tract (GI tract), an increasing bacterial density is observed from the stomach ($<10^4$ cells/gram) to the small intestine ($10^4$-$10^8$ cells/gram), up to the most densely populated compartment, the large intestine ($10^{11}$-$10^{13}$ cells/gram). This progressive increase is partially explained by the differences in hydrogen potential (pH), oxygen levels and food transit times between the different digestive compartments (10).

Two large-scale research initiatives, i.e. the European Metagenomics of the Human Intestinal Tract (MetaHIT) (11) and the American Human Microbiome Project
(HMP) (12), that were joint under the International Human Microbiome Consortium (IHMC) (13) provided vast amounts of metagenomic sequence data that aimed to improve our understanding of the human microbiome, and especially that of the gut, in health and in disease. The millions of microbial genes that were identified from the faecal specimens in these initiatives can be regarded as a functional expansion of our own ~20,000 human genes (7, 11, 14). Most of the microbial genes in the gut were found to be of bacterial origin (>99%) and are estimated to represent more than 1000 bacterial species, of which an individual person harbours at least 160 species (11). Although highly subject-specific community compositions were observed, a large fraction of species and genes identified were shared between individuals, which indicated the existence of a core metagenome. The bacterial species identified mostly belong to 5 phyla, namely Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria and Verrucomicrobia (11, 14). These first descriptions of the human gut microbiome have led to increased understanding and awareness of the roles these indigenous gut microbes have in human physiology, and how that may affect human health or disease.

**The human gut microbiome in health and disease**

The digestive tract is unique among the internal organs, because it forms a large interface with our external environment. This interface is greatly enlarged by the folds, villi and microvilli of the GI tract, resulting in an estimated surface area of 32 m² (15). Naturally, it is the central organ for the digestion and absorption of the food we eat, but it is also tightly connected with the immune, neural and endocrine systems that all interact together to optimally process the food and sense the external environment (16).

The trillions of gut microbes developed several essential activities that complement our own metabolism through the digestion of complex dietary and host-derived substrates, resulting in the production of, for example, secondary bile acids, vitamins and short-chain fatty acids (SCFA), thereby contributing nutrients, energy and bioactive compounds (17, 18). Moreover, our gut inhabitants are known to react to or modify foreign chemical components (e.g. drugs, pollutants), eliminate
exogenous toxins (19), and to provide protection against the colonization of potential pathogens by nutrient competition or through direct anti-microbial mechanisms (20).

Perhaps not surprising that due to the large surface area, the gut is known to house the largest number of immune cells in the body that need to maintain immune homeostasis locally and systemically. This involves a complex interaction of the innate and adaptive immune system in response to innocuous substances (commensal microbes, food-, and self-antigens) while protecting the body from potentially harmful ones (21-23). Given the interdependency we have with our microbial inhabitants, a relationship which has co-evolved over millions of years (24), and the essential roles they have in several physiological processes, it is not surprising that they are the focus of research into a broad range of chronic diseases. Indeed, a growing body of evidence implicate an altered composition and activity of the gut microbiome that influence the onset and persistence of several diseases that have rapidly increased in prevalence over the last decades. These range from metabolic-related diseases, such as obesity (25, 26) and type 2 diabetes (27), to immune-related diseases, such as inflammatory bowel disease (28) and allergies (29), to even neurological-related diseases, such as autism spectrum disorders (30). All these diseases have a multifactorial or polygenic aetiology, which means that they are partially explained by genetic susceptibility, but that lifestyle and environmental factors, such as diet, also contribute to the onset (31). Perturbations of the gut microbiome in early life are thought to increase the risk of developing some of these multifactorial diseases, while a child’s immunity, metabolism and cognition are still developing (9).

**Early life, gut microbiota and immune development**

The host-microbe cross-talk is a crucial factor in the proper development of the immune system, a process that starts at birth, when the infant is first exposed to significant numbers of maternal and environmental microbes (32). Some studies have suggested that even within the healthy fetal milieu, exposure to microbes or microbial signals may occur (33), which is currently heavily debated in scientific literature (34, 35). Importantly, the maturation of the immune system is for a large part taking place in the first months and years of life, which is also the period when the gut microbiota gradually develops towards a mature adult-like ecosystem (36, 37). Initial microbial
inoculation strongly depends on maternal transfer of a select number of microbes, which mostly include strains from the genus *Bifidobacterium* and from the class Bacteroidia, to the vaginally born infant. These strains have been proven to originate from the maternal gut and are ecologically better adapted to persistently colonize the infant gut than strains acquired from other sources, such as the maternal skin and vagina (38). In contrast, caesarean-born infants show a striking lack of vertical transfer and delayed colonization of these pioneering microbes (32, 39, 40).

The maternally transmitted strains show persistence in infancy with an additional continuous, but rare, influx of strains from the direct family environment (32, 41). The latter type of transmission, known as horizontal transfer, seems to be facilitated by endospore formation, which is common among gut-associated Clostridia, to survive the environment outside the gut (32, 42). Human milk, as the sole source of nutrition, is of major influence on the establishment of the gut microbiota in early life. Its influence is not limited to the period of exclusively breastfeeding but is still apparent in the period of complementary feeding until full transition to family foods around the age of 3 years (43-45). In addition to the essential nutrients, vitamins and antibodies, human milk contains numerous complex glycans known as human milk oligosaccharides (HMOs). These compounds pass the infant’s digestive system, and are efficiently consumed by especially *Bifidobacterium*, which results in an infant gut microbiome that is predominated by this genus, which is in stark contrast with that of infants receiving formula based on cow’s milk (46, 47). Additionally, human milk is also found to be a source of microbes including bifidobacteria (38, 48). This specific co-development observed for vaginally-born breastfed infants leads to a relatively stable infant microbiota which probably contributes to human health throughout life. A detailed review on these ‘first 1000 days’ of gut microbiota and immune development, the pivotal role of early life nutrition herein, and the association of perturbations of gut microbiota on the development of allergic disease, is provided in Chapter 2.

**The rise of allergies**

The epidemic rise of allergies, such as food allergy, atopic eczema, allergic rhinitis and asthma in developed and developing countries during the last 60 years has led to an immense global health problem, which is expected to affect up to 4
billion individuals in the 2050’s (49). The recognition that several environmental factors and exposures, in addition to genetic risk factors, contributed to the ‘allergy epidemic’ was already postulated in 1989 by David Strachan in what became known as the *hygiene hypothesis* (50). Originally, Strachan proposed that the increased post-industrial hygiene standards and declining family sizes reduced the opportunities for cross-viral-infections, thereby increasing the risk of allergic disease, which was based on a strong inverse correlation in the incidence of hay fever in children with older siblings. More recently this hypothesis has been revisited and revived by implicating the early resident microbial communities in allergic disease (51, 52).

This renewed view of the hygiene hypothesis, which is sometimes referred to as the *biodiversity hypothesis*, proposes that significant perturbations of the resident microbial communities during early life immune development may disrupt the mechanisms of mucosal immunologic tolerance against innocuous antigens, such as food proteins (51-53). Several epidemiological observations yielded support for this hypothesis by the association of increased antibiotic-exposure (54-56), C-section delivery (57, 58), decreased breastfeeding-duration (59-61) and absence of exposure to pets (62) and farm-environment due to urbanization (63, 64) with an increased risk of developing allergic diseases, such as food allergy. Moreover, the loss of ancient members of the gut microbiota, due to compromised transmission from one generation to the next, may have compromised the context of establishing a healthy symbiosis in early life (52, 65, 66). Another important risk factor that more recently has been identified is reduced food diversity in the infants’ diets and the delayed introduction of allergenic foods (67), the latter of which has led to changing the guidelines for complementary feeding to not unnecessarily avoid the introduction of allergenic foods after 4 months of age once weaning has started, irrespective of atopic heredity (68, 69).
Dietary approaches in the primary prevention and management of cow’s milk allergy

Protein hydrolysates and amino acid-based formulas

If breastfeeding is insufficient or not possible, international guidelines recommend partially hydrolysed cow’s milk protein formulas (pHF) for infants with a family history of allergic disease (and therefore deemed ‘high-risk for allergy’) (68, 70). Clinical trials have shown the potential of reduced allergen exposure by specific pHF to reduce the risk of developing eczema in high-risk infants (71). This is, however, challenged by a 2016 meta-analysis that evaluated several different hydrolysed formulas and concluded there was no consistent evidence for prevention of allergy (72). The mixed outcomes may be explained by the combined evaluation of various types of hydrolysed formulas (71) and the differences in tolerogenic potential of the peptides present in formulas from different manufacturers (73).

For the dietary management of infants with cow’s milk allergy (CMA) and when breastfeeding is not possible, expert opinion recommends the use of hypoallergenic formula, such as extensively hydrolysed cow’s milk protein-based formula (eHF) and amino acid-based formula (AAF) to quickly resolve symptoms through allergen avoidance (74, 75). Although an eHF is suitable for most CMA infants and children, between 10–40% of patients require an AAF when allergic symptoms do not resolve with an eHF (76).

Prebiotics, probiotics and synbiotics

There is growing research into the potential role of the microbiome in patients with asthma, atopic dermatitis, and food allergy. This is complemented by an increasing number of trials that investigate prebiotics, probiotics or a combination of both (synbiotics) (77) to manipulate the microbiome to influence the development of sensitisation and allergy (51, 78), which generally reflects the move from allergen avoidance towards more active management of allergy (79). Probiotics are ‘live microorganisms that, when administered in adequate amounts, confer a health benefit on the host’, as defined by the Food and Agriculture Organization/World Health Organization in 2001 (80) and updated in 2014 by the International Scientific
Association of Probiotics and Prebiotics (ISAPP) (81). The core-group of well-studied and widely applied bacterial species are strains of *Bifidobacterium* (*adolescentis, animalis, bifidum, breve* and *longum*) and *Lactobacillus* (*acidophilus, casei, fermentum, gasseri, johnsonii, paracasei, plantarum, rhamnosus* and *salivarius*), respectively (82). The next-generation probiotics may also include more recently identified key members of the human gut microbiota, which have been associated with healthy states in metagenomic studies (81). These include *Akkermansia muciniphila* (83) and butyrate-producing bacteria, such as *Faecalibacterium prausnitzii, Roseburia* spp., *Anaerostipes* spp. and *Eubacterium hallii* (84).

In 2017, the ISAPP reviewed the definition and scope of prebiotics as originally coined by Gibson and Roberfroid in 1995 (85) and updated the definition to ‘substrates that are selectively utilized by host microorganisms conferring a health benefit’ (86). This definition broadened the initial location of action beyond the gastro-intestinal tract and added the possibility of including non-carbohydrate substances. Currently established prebiotics are mostly carbohydrate-based fructo-oligosaccharides (FOS and inulin) and galactans (galacto-oligosaccharides or GOS) that have been shown to selectively stimulate *Bifidobacterium* spp. and/or *Lactobacillus* spp. in the GI-tract (82). The mimicry of these compounds with HMOs, both in effects, building blocks and size, led at the beginning of this century to the development of the first infant formula that were enriched with specific short-chain galacto-oligosaccharides (scGOS) and long-chain fructo-oligosaccharides (lcFOS), which were added in a 9:1 ratio and at a total dosage of 0.8 g/100ml to mimic the molecular weight profile and concentrations of HMOs in breast milk (87). This specific prebiotic mixture has since its introduction been investigated in several clinical investigations, which confirmed its safety and its efficacy in shifting microbiota composition, metabolic signatures, stool frequency and stool softness closer to that of breastfed infants. Moreover, these studies provided evidence towards the prevention of allergies and infections as has been summarized recently (87).
Evidence for prebiotics, probiotics and synbiotics in the prevention and treatment of allergy

The most recent evaluation on the preventive effect of probiotics and prebiotics was performed in a joint effort of the World Allergy Organization (WAO) and the Department of Clinical Epidemiology & Biostatistics at McMaster University using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which resulted in the Guidelines for Atopic Disease Prevention (GLAD-P) (88, 89).

In relation to probiotics, the panel concluded that based on the available evidence (up to November 2014) probiotic supplementation did not reduce the risk that children will develop allergy (89). However, when considering all critical outcomes (costs, risk of adverse events), the panel determined that ‘there is a likely net benefit from using probiotics resulting primarily from prevention of eczema’ and suggested to use probiotics a) in pregnant women at high-risk of having an allergic child; b) in women who breastfeed infants at high-risk of developing allergy; and c) in infants at high-risk of developing allergy (89). Obviously, the evaluation included various types of probiotic strains (and species), plus a variety of study designs; no evaluation is currently made among the strains of the same species for example.

In relation to prebiotics, the panel suggested prebiotic supplementation in not-exclusively breastfed infants, both those at increased (based on family history of allergy) and normal risk for developing allergy. The majority of studies with infant formula (11 out of 15) included the specific scGOS:lcFOS (9:1) mixture. The effects of supplementing prebiotics during the first year of life was associated with reduced risk of developing asthma or recurrent wheezing, food allergy and probably eczema, however with low certainty due to risk of bias and imprecision of estimates (88).

Currently no recommendations are available concerning the use of prebiotics and/or probiotics in the treatment of allergy using a GRADE-like evaluation of available evidence. This probably is explained by the lack of studies that could be evaluated. However, recently a meta-analysis was performed by Chang and colleagues who evaluated the use of synbiotics for the treatment and prevention of atopic dermatitis, which offers encouraging findings based on a decrease in the
severity scoring of atopic dermatitis (SCORAD) values observed when combining 6 treatment studies (90).

Further studies are required to confirm and determine the beneficial effects of specific prebiotics, probiotics and synbiotic combinations for prevention and treatment of allergy, which may ultimately translate to specific clinical recommendations.

**Aim and outline of this thesis**

**Aim**

A suboptimal implementation of the gut microbiota in early life could contribute to the onset of allergy and/or enhancement of allergic symptoms. The main aim of this thesis was to investigate normal gut microbial trajectories in early life and to identify deviating patterns that are associated with allergic manifestations, such as atopic dermatitis and food allergy.

**Outline**

In Chapter 2, we give an overview of the development of the gut microbiota in early life, the factors impacting this development and the importance of the microbe-immune cross-talk to establish a healthy symbiosis. The review additionally includes a summary of paediatric studies on the association between gut microbiota composition and allergic disease.

Bifidobacteria are the most abundant bacteria in early life, but often under-represented in 16S rRNA-gene sequencing surveys of gut microbiota, due to poor DNA extraction techniques, poor PCR primer choice or a combination of both. When the research described in this thesis was initiated, we modified a commonly used ‘universal’ primer-set to improve the recovery of this genus and tested and validated its performance on a set of infant stools (Chapter 3).

In the research described in Chapter 4 we applied the ‘bifidobacteria-optimized’ 16S rRNA-gene sequencing method to analyse the gut microbiota of infants at high-risk of developing allergy, who participated in a clinical trial that investigated the effects of a partially hydrolysed protein formula supplemented with prebiotics on the
prevention of eczema. We investigated the effects of the nutritional intervention on the assembly of intestinal microbiome compared to infants receiving standard formula (without prebiotics) and the breastfed reference group. Additionally, in a nested case-control we investigated the association of microbiota with development of atopic dermatitis (AD).

When breastfeeding is not possible, infants suffering from CMA often require a hypoallergenic amino acid-based formula (AAF) to meet their nutritional needs and to quickly resolve allergic symptoms. In the clinical trial described in Chapter 5 we investigated the modulatory effects of an AAF supplemented with synbiotics on the gut microbiota in infants with suspected CMA with reference to healthy, breastfed infants.

In Chapter 6 we screened the intestinal microbiota of CMA-infants and healthy controls with reference to the observations done in the clinical trial described in Chapter 5. Next, we selected one representative donor per condition for faecal transfer to a murine model of CMA to investigate whether the alterations in gut microbiota composition contributed to the pathology of the disease.

Finally, the findings described in this thesis are summarized, discussed, and concluded in Chapter 7.
References

Introduction


49. Akdis CA, Agache I, Allergy EAo, Immunology C. Global Atlas of Allergy: European Academy of Allergy and Clinical Immunology; 2014.


CHAPTER 2

The first thousand days - intestinal microbiology of early life: establishing a symbiosis

Wopereis H\textsuperscript{1,2}, Oozeer R\textsuperscript{1}, Knipping K\textsuperscript{1}, Belzer C\textsuperscript{2}, Knol J\textsuperscript{1,2}.

\textsuperscript{1}Danone Nutricia Research, Utrecht, The Netherlands
\textsuperscript{2}Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

Pediatric Allergy and Immunology 2014;25(5):428-38. (minor modifications).
Abstract

The development of the intestinal microbiota in the first years of life is a dynamic process significantly influenced by early life nutrition. Pioneer bacteria colonising the infant intestinal tract and the gradual diversification to a stable climax ecosystem play a crucial role in establishing host-microbe interactions essential for optimal symbiosis. This colonisation process and establishment of symbiosis may profoundly influence health throughout life. Recent developments in microbiological cultivation-independent methods allow a detailed view of the key players and factors involved in this process and may further elucidate their roles in a healthy gut and immune maturation. Aberrant patterns may lead to identifying key microbial signatures involved in developing immunological diseases into adulthood, such as asthma and atopic diseases. The central role of early life nutrition in the developmental human microbiota, immunity and metabolism offers promising strategies for prevention and treatment of such diseases.

This review provides an overview of the development of the intestinal microbiota, its bidirectional relationship with the immune system, and its role in impacting health and disease, with emphasis on allergy, in early life.
Our microbial world

We live in a microbial world. Micro-organisms were amongst the very first life forms and still today form the greatest biomass on this planet (1). They hardly exist as single cells in nature, but rather live in complex communities coevolved and adapted to the habitats they colonise. Surveys of these complex communities are taking great advantage from the use of high-resolution cultivation-independent methods such as phylogenetic microarrays or next generation sequencing (2). Especially the study of human intestinal microbial species of which approximately 70% have not been isolated, cultivated or sequenced, due to the inability to reproduce necessary growth conditions in the lab. Sequencing of PCR amplified 16S ribosomal RNA (16S rRNA), a conserved gene routinely used for phylogenetic identification of bacteria, and whole genome surveys (metagenomic sequencing) are now starting to reveal the true microbial diversity of the human intestine and their role in health and disease. The human intestinal tract is colonised with about ten times more microbial cells than human body cells and contain about 150 times more microbial genes than the human genome (2). The intestinal microbiota is coexisting in a homeostatic relationship with the host (3). This host-microbial relationship is maintained in a bidirectional manner with the immune system. The intestinal microbiota benefits from a stable environment and nutrient supply that are provided in the intestinal tract, while the host gains products from microbial fermentation conversion of host indigestible components (dietary fibres) into short-chain fatty acids (SCFA; mainly acetate, propionate and butyrate) contributing to an estimated 10% of our energy requirement (4), vitamin K and B12 production (5, 6), and protection against potential pathogens through competitive exclusion (7, 8). The importance of the human microbiota is particularly clear as alterations of the intestinal microbiota have been associated with short and long-term health and disease issues, such as intestinal bowel disease (IBD), allergy, diabetes, obesity and autism (9).

The development of the intestinal microbiota is a dynamic process in the first years of life (10-12), a time frame that is also a critical period of gut and immune development and maturation (13). Indeed, the pioneer bacteria colonising the infant intestinal tract and the gradual diversification to a stable climax ecosystem play a
crucial role in establishing host-microbe interactions essential for optimal symbiosis; and this colonisation process may profoundly influence health throughout life (8). Future research should focus on the analyses of longitudinal data that may identify the patterns of early intestinal microbiota and functionality of not yet-cultivated species that could affect health later in life (14).

Microbial pioneers

Theodor Escherich (1857–1911) pioneered the study of intestinal microbiology in early life (15). In 1886, Escherich published his 177-page postdoctoral thesis entitled, “The Intestinal Bacteria of the Infant and Their Relation to the Physiology of Digestion” (16). Escherich demonstrated that meconium was sterile, and that bacterial intestinal colonisation is attributable to the infants’ environment and emphasized the value of breastfeeding herein.

The rise of molecular biology in the second half of the twentieth century and the more recent revolution in sequencing technologies identified the key players of the developmental intestinal microbiota in more detail. The majority is assigned to 4 phyla, namely the Actinobacteria (with genera like Bifidobacterium and Colinsella), the Bacteroidetes (with genera like Bacteroides and Prevotella), the Firmicutes (with genera like Lactobacillus, Clostridium, Eubacterium and Ruminococcus) and the Proteobacteria (e.g. Enterobacter spp.) (17). Another phylum identified throughout life is the Verrucomicrobia consisting of one major species, the mucin-degrading Akkermansia muciniphila (7).

The composition of the microbiota changes substantially at two stages in early life: from birth to weaning, and from weaning to adulthood driven by further diversification of diet (4). The pioneer species in neonates are facultative anaerobic bacteria, like Staphylococcus, Streptococcus, Enterococcus and Enterobacter spp., these bacteria create an anaerobic environment that promote the growth of obligate anaerobes, such as Bifidobacterium, Bacteroides, Clostridium and Eubacterium spp., predominating after one or two weeks. Escherich’s observation of sterile meconium and thus supposed sterile intestine at birth has only recently been opposed with molecular surveys suggesting that microbial exposure may start before birth and that
infants may already receive microorganisms from the mother during gestation (18, 19). Right after birth the early settlers are derived from the maternal microbiota (vaginal, faecal, human milk, mouth, skin) and the environment (20-22). Human milk forms an important continuous inoculum, while bacterial strains found in breast milk have also been detected in faecal samples from the corresponding infants (21, 23). These bacteria are postulated to translocate from the mothers’ intestine to the mammary gland via the mesenteric lymph nodes, suggesting a possible route of inducing immunological tolerance to these commensals (24). Another possible or contributing route may include the establishment of the mothers’ skin microbiota and infants’ oral microbiota into the mammary gland (25). Host genotype, gestational age, medical practices (i.e., antibiotic use), mode of delivery (caesarean section vs. vaginal delivery), geographical origin and linked to that, cultural traditions, especially regarding diet, are factors profoundly influencing the microbiota development (26, 27). Breastfed infants typically have a microbiota dominated by bifidobacteria, while formula-fed infants have a more diverse microbiota. Infants born preterm or by caesarean section show a reduced diversity and a delayed colonisation by bifidobacteria compared to infants born at term or vaginally (26, 28). Some studies applying PCR amplification and sequencing did not reproduce the early predominance of *Bifidobacterium* (12), however efficient DNA extraction and careful selection of PCR primers have proven to be critical to effectively detect this genus (29).

Introduction of first solid foods around 4 to 6 months of age impacts the infant microbiota considerably. Although still “infant-like”, with decreased but still dominating levels of bifidobacteria, a gradual diversification is seen towards more adult-type species, mainly *Bacteroides* spp. and *Clostridium* clusters IV and XIV, the latter two clusters known to contain numerous butyrate producers (10, 30). Interestingly the factors influencing the early colonisation process strongly influence the post-weaning colonisation pattern. Early diversification, as observed under formula-feeding not containing prebiotics, promotes earlier acquisition of an adult-type microbiota (10, 30). Further diversification of diet gradually increases diversity and abundance of Bacteroidetes and Firmicutes towards adult levels and generally low abundant levels
of *Bifidobacterium* (27). Healthy adults have a stable microbiota; unique for individuals though sharing a core microbiome with other individuals, which may change and destabilise only at older age again (4, 31). Although low abundant in adults, *Bifidobacterium* species still play important metabolic roles in adults (32). On top of this ageing is generally associated with a significant decrease of bifidobacteria, along with other rearrangements and decreasing stability, all together associated with increased susceptibility to infections in elderly (31). The exact age at which a stable adult community is established is unclear but is thought to be reached around 3 years of age (11, 27). Changes in the genetic capacity of the microbiome with human development include changes in the abundance of genes involved in access to host-derived glycans (in human milk and intestinal mucosa) and vitamin biosynthesis, i.e. infants having more genes that encode enzymes involved in folate biosynthesis and adults more encoding for vitamin B12 (27). The influence of early colonisation patterns on the composition of the adult microbiome is not yet fully understood. However these patterns have been shown to influence gut maturation, immune development and host metabolism (8), and differences in composition driven by environmental factors in infancy may affect susceptibility to metabolic (e.g. obesity), immunological (e.g. IBD and allergy) and even behavioural (e.g. autism) disorders into adulthood (Figure 1), diseases which are increasingly prevalent in developed countries (9). The central role of diet in influencing the human microbiota, immunity and metabolism offers promising strategies for prevention and treatment of such diseases.

**Establishing a healthy symbiosis**

Neonates have a limited capacity to initiate immune responses and both innate and adaptive immune responses are not yet fully functional. In the months and years after birth the immune system gradually matures (8), concurrent with the infants’ microbiota development. The largest immune component in the body, the mucosal immune system, comprised of the gut-associated lymphoid tissue (GALT); the mucosal lamina propria; and the mucosal surface, have a central role in this developmental process.
Establishing a symbiosis

Figure 1: Early life, gut microbiota and immune development – establishing a symbiosis. The establishment of the intestinal host-microbiota symbiosis is driven by both developmental and environmental signals especially in early life, profoundly influencing health throughout life. The prenatal intestine is thought to be sterile and development depends most importantly on genotype as well as on maternal factors, including nutrition and health status. The cryptopatches and lymphoid tissues (mesenteric lymph nodes and Peyer’s patches) with dendritic, T and B cells develop in preparation of the exposure to the extra-uterine world. During birth infants are inoculated with maternal and environmental microbes, and the type and patterns strongly depend on birth mode and gestational age. The gut microbial development in the neonatal period is influenced by several early life factors and especially diet (type, composition and timing) drives the further diversification towards an adult complexity, which is reached around 3 years of age. This postnatal colonization process provides several signals, known as microbe-associated molecular patterns (MAMPs), affecting the maturation of the immune system and the mucosal barrier, accompanied with increased mucus secretion. These signals also result in the proliferation of intestinal epithelial cells in crypts and the crypt-located Paneth cells, resulting in their increased depth and the production of antimicrobial peptides (defensins), respectively. Specialized epithelial cells (M cells), reside above Peyer’s patches and facilitate direct interaction of the luminal content with the underlying lymphoid cells to stimulate mucosal immunity. SIgA is the most abundant immunoglobulin on mucosal surfaces and maternal SIgA is provided by human milk during the early postnatal period along with the initiation of the infants own SIgA.
The mucosal immune system is providing protection from the external environment and directly interacts with the environmental antigens and commensal bacteria (13). The epithelial layer in neonates shows a higher permeability in both the respiratory and gastrointestinal tracts and secretion of proteases and antimicrobial peptides have not fully developed (33). The epithelial production of mucus forms an important first line of defence against microbes. The thickness and continuity of the intestinal mucus layer increases from the small intestine towards the colon correlating with increasing bacterial loads (13). Mucin glycans are nutrients for some constituents of the microbiota, such as *Bifidobacterium*, *Bacteroides* spp. and *Akkermansia muciniphila*, giving them an ecologically advantage to reside in the outer mucous layer close to the intestinal epithelial cells (IECs) (34). Niche occupation by such commensals is not only establishing a physical barrier excluding potential pathogens, also the production of acetate and lactate form an effective chemical barrier toxic for potential pathogens (35, 36). Levels of faecal SCFA of human milk fed infants are characterized by relatively higher proportions of acetate and lower proportions of propionate and almost complete absence of butyrate, when compared to adults. Also, lactate is more commonly detected in the faeces of infants, while undetected in healthy adults due to immediate onward conversion by lactate-utilizing bacteria. These elevated levels of acetate and lactate in human milk fed infants are reflecting the dominance of bifidobacteria and lactobacilli (37). Although faecal levels of butyrate are generally low in human milk fed infants, acetate and lactate may in turn be used to gradually establish butyrate producers within the Firmicutes (4, 38), which have recently been shown to be less abundant in colicky infants at 2 weeks of age at the expense of increased levels of potential pathogenic members of the Proteobacteria (39). Many of the direct effects of SCFA on epithelial cells associated with maintenance of the epithelium relate mostly to their role as an energy source and their inhibition of histone deacetylases, the latter is directly impacting human gene expression and e.g. shown to downregulate inflammation in patients with ulcerative colitis (40). The SCFA have also been shown to influence immune function beyond the gut by signalling through G-protein-coupled receptors (GPR) on IECs. Mice deficient of GPR43 have
exacerbated and poorly resolving inflammation in inflammatory models of arthritis, allergic airway inflammation and colitis (41).

An important extra layer of innate mucosal defence in neonates is derived from human milk. In addition to a unique mix of human milk oligosaccharides (HMO), and antimicrobial proteins that influence the ecology of the neonatal microbiota, human milk provides abundant secretory immunoglobulin A (sIgA), the specificities of which have been shaped by the maternal digestive system and microbiota (13). sIgA is the most abundant immunoglobulin on mucosal surfaces, where it neutralizes harmless food and microbial antigens and prevents them from penetrating the epithelium. However IgA can also function in high-affinity modes for neutralization of toxins and pathogenic microbes, and as a low-affinity system to contain the dense commensal microbiota within the intestinal lumen (42). Next to sIgA, also maternal IgG-antigen complexes play a major role in shaping the infants’ immune system. Antigen bound to IgG will be very efficiently transferred across the gut barrier using the neonatal Fc receptor (43). Therefore, both maternal sIgA and IgG may be important in the development of non-responsiveness to harmless commensals and food antigens, i.e. induction of oral tolerance (44). The developmental microbiota is essential for the initiation of an infants’ own sIgA while germ-free mice show drastically reduced mucosal IgA-secreting cells. Studies using prebiotics or synbiotics (combination of pre- and probiotics) treatment given for 6 months to infants showed increased levels of faecal sIgA (45, 46) and is linked to reduced risk of allergy before 2 years of age in one of these studies (46).

More evidence is mounting in how commensal bacteria directly influence adaptive immunity and oral tolerance and is focusing on the mechanisms involved in the cross-talk between the intestinal microbiota and the host. This cross-talk is mediated through pattern-recognition receptors (PRRs), such as toll-like receptors (TLRs), specifically recognising conserved microbial molecular structures, called microbe-associated molecular patterns (MAMPs) (8). Recognition of these patterns may promote pro-inflammatory responses or repress them and seem to depend on by whom (i) and where (ii) they are triggered (13). First (i), subtle differences between commensal bacteria, probiotic and pathogenic microorganisms may mediate different
host responses (47) and second (ii), apical signalling normally promotes intestinal homeostasis, however basolateral signalling, implicating intestinal barrier disruption and infection, initiates inflammatory responses of innate and adaptive immune cells (13) and can lead to exacerbate intestinal inflammation (48). Pro-inflammatory responses are counterbalanced by specialized T cells known as regulatory T (Treg) cells and play a crucial role in maintaining immune homeostasis. Treg cells are characterized by production of IL-10, one of the main immunoregulatory cytokines required for immune tolerance of the intestinal microbiota (13). The role of TLR in sensing the microbiota in this process is evident by the absence of colonic inflammatory disease in germ-free IL-10-deficient mice and mice deficient for both IL-10 and myeloid differentiation factor 88 (MyD88). Remarkably IL-10-producing T cells can be induced to develop in response to specific commensals or their products. This was first shown for a common commensal, *Bacteroides fragilis* through its polysaccharide A (PSA) mediating through the TLR2–MyD88 pathway (49). More human symbionts are thought to exert comparable mechanisms to induce mucosal tolerance, i.e. a probiotic *Bifidobacterium breve*, but not a *Lactobacillus casei* strain, induced development of IL-10-producing Treg cells and were shown to prevent inflammation in a colitis model (50).

In these and other studies, MyD88-dependent TLR signalling has proven to be a key mediator for maintenance of intestinal homeostasis, requiring active communication among epithelial cells, immune cells, and the intestinal microbiota (48). Hill et al. (51) showed that antibiotic-mediated disruption of the microbiota is sufficient to predispose mice to allergic disease. The authors showed that the commensal microbiota modulates B-cell production of IgE antibody in a MyD88-dependent manner and that perturbation leads to high circulating levels of basophils and high serum IgE concentration. Exposure of antibiotic-treated mice with DNA motifs specific for bacteria (unmethylated cytosine-guanosine CpG oligonucleotides), a known Toll-like receptor 9 (TLR9)-dependent microbial ligand, was sufficient to reduce serum IgE as well as the frequency and total number of circulating basophils. These findings identify intriguing links between the adaptive
immune system interacting with the intestinal microbiota and will further elucidate the specific microbes involved in promoting a healthy host-microbiota symbiosis.

**Early microbiota and allergy**

The prevalence of atopic manifestations (atopic dermatitis, food allergy, allergic rhinitis and asthma) has been increasing worldwide, predominantly in the western world and particularly among children (52). Expression of an allergic phenotype is dependent on the interaction between 2 major factors: a genetic predisposition and gene-environment interactions (e.g. lifestyle, diet). Infants suffering from atopic dermatitis and or food allergy are more susceptible to develop other allergies like allergic asthma later in life, a process known as the atopic march (53). There is mounting evidence that modifications in the pattern of microbial exposure early in life represents a critical factor underlying the development of an allergic phenotype (54-57), such as the protective effects observed for exposure to siblings or a farming environment (58, 59). Accumulating preclinical studies start to reveal pathways linking aberrant microbial patterns to atopic diseases (51, 60). The classical explanation for the increasing prevalence of allergies in western countries, and a possible role of “early” microbes, was postulated in the hygiene hypothesis in 1989 (61). This hypothesis focuses merely on decreased exposure to infectious agents under improved hygiene standards to explain the hypersensitive reaction of the immune system towards normally harmless substances in the environment. The supposed mechanism proposes a lack of shifting of allergen-specific responses from the T Helper 2 to the T Helper 1 phenotype, because of reduced exposure to infectious agents in early childhood (62). More recently the possible implication of the resident human intestinal microbiota in developing allergy has been suggested to play a crucial role in the development of mucosal immunologic tolerance. The discovery of Treg cells and their role in immune suppression and self-tolerance (63), lead to an important explanatory mechanism of reduced activity of Treg to a loss of microbial symbionts, which may partly explain the increasing prevalence of other western diseases, like IBD, obesity and diabetes (9, 62). The role of the endogenous microbiota in developing allergy under this extended hygiene hypothesis is supported
by the positive correlation of environmental factors, known to impact microbial colonisation, and allergic manifestations (i.e. antibiotic use and caesarean section), correlations with an altered microbiota composition and increasing evidence of successful prevention or reduction of allergy through microbiota modulating diets (52). Altered microbial composition and activity between healthy and atopic children have been shown in several cross-sectional epidemiologic studies and have been extensively reviewed up to January 2007 by Penders et al. (64). Table 1 gives an overview of observational studies summarised in a similar approach, from 2007 onwards, considering type of atopic disease under study, the study population, design and methodology to examine the intestinal microbiota. Interestingly, differences in the intestinal microbiota composition often precede the manifestation of atopic symptoms and atopic sensitization (65-69), although two studies reported no meaningful differences between groups (70, 71), possibly explained, as the authors suggest, by the application of cultivation methods overlooking the unculturable bacteria. Reduced bacterial diversity in the early microbiome has been associated with developing atopic disease by several comparative studies (54, 65, 66, 72, 73). Abrahamsson et al., applying 16S rRNA sequencing, linked reduced bacterial diversity at 1 month to IgE-associated eczema in infants at 2 years of age, which was subsequently confined to developing asthma at 7 years of age (74), supporting the importance of pioneer microbes in early immune maturation. This early reduced diversity was mainly attributed to a decreased diversity of *Bacteroides* spp. within the Bacteroidetes phylum. At 12 months of age a decreased diversity of Proteobacteria was observed and a tendency of higher levels of the phylum Firmicutes in atopic infants, a phylum indicating development towards a more “adult-type” microbiota. Nylund et al., applying a phylogenetic microarray, reported increased diversity at 18 months of age, but not at 6 months, in eczematous versus healthy infants. This increased diversity at 18 months of age was associated with higher abundances of *Clostridium* clusters IV and XIVa, members of the Firmicutes phylum. At this age healthy infants showed increased abundance of members within Bacteroidetes (56), a group of bacteria which may have been underestimated in early life, due to molecular bias (75). Notably species within the Bacteroidetes have been shown, next to *Bifidobacterium* spp., to be
efficient fermenters of human milk oligosaccharides in contrast to species within the Firmicutes phylum (76). Interestingly a recent study showed that colonisation of germ-free mice with the faecal microbiota of a healthy infant rich in *Bifidobacterium* spp. and *Bacteroides* spp. protected against the development of cow’s milk allergy following sensitization to β-lactoglobulin (77). The genera within the phyla Bacteroidetes and Firmicutes linked to allergy in both observational studies, may thus play important roles in the gradual succession of an infant-type microbiota, dominated by bifidobacteria, towards a stable adult-type microbiota.

The observations of both decreased and increased bacterial diversity linked to allergic manifestations may seem contradictory but were made at different stages of early life and development of allergic disease. Bacterial diversity as such is difficult to interpret in early childhood, while the early microbiota is highly dynamic with high inter-individual variation. Also, bacterial diversity gradually increases towards adulthood reaching adult levels no earlier than around 3 years of age (11, 27). There is an ongoing debate whether low total diversity of the gut microbiota in early childhood is more important than the altered prevalence of particular bacterial species in allergy development (78), but more likely the combination of both may lead to identifying the key microbial signatures for developing allergy and response to nutritional strategies.

**Early microbiota and nutrition**

The initial bacterial colonisers of our gastrointestinal tract may determine the composition of our intestinal microbiota throughout life. Furthermore, this early development occurs concomitantly to the development of our metabolism, cognitive and immune systems, which have been described to be closely linked to the intestinal microbiota. Knowing that the microbiota can significantly interfere with the human metabolic, cognitive, and immune systems, the initiation of symbiosis seems a crucial step for preparing optimal health later in life. Consequently, understanding the early interaction between the intestinal microbiota and the human body opens new avenues for important nutritional innovations, particularly for infants and young children.
### Table 1: Human observational studies on the association between the gut microbiota composition and atopic diseases. Studies applying next generation sequencing or microarray technologies as microbial analysis tool are indicated in bold.

<table>
<thead>
<tr>
<th>Allergic phenotype</th>
<th>Study design (country)</th>
<th>Tool</th>
<th>Allergic vs non-allergic (ref)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcZ (Williams’ criteria) and/or sIgE+ until 18m</td>
<td>PC: 324 infants at risk for allergy (SE, UK and IT)</td>
<td>Cultivation</td>
<td>No differences observed (70)</td>
</tr>
<tr>
<td>Allergic manifestations (ISAAC questionnaire) until 2y</td>
<td>PC: 15 infants (JP)</td>
<td>qPCR</td>
<td>Increased abundance of Bacteroidaceae at the ages of 1 and 2m of age (67)</td>
</tr>
<tr>
<td>Any allergic manifestation or SPT+ until 6m</td>
<td>CC: 10 allergic and 16 non-allergic infants (JP)</td>
<td>PCR</td>
<td>Higher prevalence of <em>Bifidobacterium catenulatum</em> group at 1m and higher prevalence of <em>B. bifidum</em> at 6m of age (79)</td>
</tr>
<tr>
<td>EcZ (PD) until 6m</td>
<td>CC: 9 allergic and 12 non-allergic infants (USA)</td>
<td>DGGE</td>
<td>Lower diversity at 1 and 4m of age (73)</td>
</tr>
<tr>
<td>EcZ (PD) until 6m</td>
<td>CC: 37 allergic and 24 non-allergic infants (NZ)</td>
<td>TTGE, FISH</td>
<td>Higher prevalence of <em>Bifidobacterium pseudocatenulatum</em> (80)</td>
</tr>
<tr>
<td>API: wheezing + EcZ /wheezing + allergic heredity until 3y</td>
<td>PC: 117 infants (B)</td>
<td>Cultivation</td>
<td>Higher prevalence of <em>Bacteroides fragilis</em> at 3w of age (81)</td>
</tr>
<tr>
<td>EcZ (Williams’ criteria) and/or sIgE+ until 18m</td>
<td>CC: 15 allergic and 20 non-allergic infants (SE, UK, IT)</td>
<td>T-RFLP, TTGE, qPCR</td>
<td>Lower diversity at 1w of age (72), same cohort as (70)</td>
</tr>
<tr>
<td>Allergic manifestations and at least 1 SPT+ until 5y</td>
<td>CC: 16 allergic and 31 non-allergic infants (SE)</td>
<td>qPCR</td>
<td>Lower prevalence of Lactobacilli, <em>Bifidobacterium adolescentis</em> and <em>Clostridium difficile</em> during first 2m of life (82)</td>
</tr>
<tr>
<td>EcZ (PD) until 2y</td>
<td>CC: 3 allergic and 5 non-allergic, C-section (USA)</td>
<td>16S rRNA seq</td>
<td>Lower abundance of <em>Bifidobacterium</em>, higher abundance of Enterococcus, Klebsiella and Shigella in 1st y of life (83)</td>
</tr>
<tr>
<td>Allergic manifestations (PD) until 1y</td>
<td>CC: 24 allergic and 72 non-allergic, VLBW (NL)</td>
<td>FISH</td>
<td>Lower prevalence of <em>Bifidobacterium</em> at 1y of age (84)</td>
</tr>
<tr>
<td>SPT+ and sIgE+ and/or allergic manifestations (PD) until 6y</td>
<td>PC: 411 infants with maternal history of asthma (DK)</td>
<td>Cultivation, DGGE</td>
<td>Low diversity at 1 and 12m with SPT+/sIgE+ and allergic rhinitis, but not with asthma or AD (65)</td>
</tr>
<tr>
<td>Allergic manifestations and SPT+ and/or sIgE+ until 5y</td>
<td>CC: 16 allergic and 19 non-allergic infants (SE)</td>
<td>qPCR</td>
<td>Lower prevalence of lactobacilli (<em>L. casei, L. paracasei, L. rhamnosus</em>) in 1st 2m, lower prevalence of <em>Bifidobacterium bifidum</em> in 1st w of life (55)</td>
</tr>
<tr>
<td>Allergic manifestations (ISAAC questionnaire) until 2y</td>
<td>CC: 11 allergic and 11 non-allergic infants (JP)</td>
<td>16S rRNA seq</td>
<td>Higher abundance of <em>Bacteroides</em>, lower abundance of <em>Clostridium</em> and <em>Proteobacteria</em> (other than <em>Klebsiella</em>) at 1m and higher abundance of <em>Klebsiella</em> at 1 and 2m (85). Same cohort as (67)</td>
</tr>
<tr>
<td>Allergic phenotype</td>
<td>Study design (country)</td>
<td>Tool</td>
<td>Allergic vs non-allergic (ref)</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------</td>
<td>------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>CMPA with SPT+, slgE+ and DBPCFC+ for cow's milk, age 2 to 12m</td>
<td>CC: 46 allergic and 46 non-allergic (SP)</td>
<td>FISH</td>
<td>Increased abundance of Clostridium coccoides group and Atopobium cluster, increased concentrations of butyric acid and branched-chain SCFA (86)</td>
</tr>
<tr>
<td>API: wheezing + EcZ/wheezing and allergic heredity until 3y</td>
<td>PC: 110 infants (B)</td>
<td>DGGE</td>
<td>Association of Clostridium coccoides XIVa species and Bacteroides fragilis species at 3w (68). Same cohort as (81)</td>
</tr>
<tr>
<td>SPT+ and/or positive atopic patch test and/or radioallergosorbent test, age from 6-24m</td>
<td>CC: 10 allergic and 20 non-allergic infants (FR)</td>
<td>Cultivation TTGE BOX-PCR</td>
<td>No differences observed in bacterial groups cultivated nor in the bifidobacterial-specific fingerprinting (71)</td>
</tr>
<tr>
<td>Ecz with SPT+ and/or slgE+ until 2y</td>
<td>CC: 20 allergic and 20 non-allergic infants (SE)</td>
<td>16S rRNA seq.</td>
<td>Lower microbial diversity at 1m of age, linked to reduced Bacteroides spp. diversity (54)</td>
</tr>
<tr>
<td>Allergic manifestations (PD) and/or slgE+, age from 4-14y</td>
<td>CC: 19 allergic children and 12 non-allergic (IT)</td>
<td>16S rRNA µarray, qPCR</td>
<td>Decreased abundancies of Clostridium cluster IV with Faecalibacterium prausnitzii, Akkermansia muciniphila and increased abundance of Enterobacteriaceae (87)</td>
</tr>
<tr>
<td>Ecz (PD) with or without SPT+ until 12m</td>
<td>PC: 98 high risk infants (AUS)</td>
<td>T-RFLP</td>
<td>Low microbial diversity at 1w of age with EcZ development but not with SPT+ or parental allergic status (both/single) (66)</td>
</tr>
<tr>
<td>Allergic manifestations and SPT+ until 7y</td>
<td>CC: 47 infants (SE)</td>
<td>16S rRNA seq.</td>
<td>Low microbial diversity at 1w and 1m in infants having SPT+ associated EcZ in first 2y of life, subsequently developing asthma at 7y of age (74), same cohort as (54)</td>
</tr>
<tr>
<td>Ecz (PD) with or without SPT+ until 2y</td>
<td>CC: 15 allergic and 19 non-allergic high risk infants (FI)</td>
<td>16S rRNA µarray, qPCR</td>
<td>Higher diversity at 18m, increase of Clostridium clusters IV and XIVa and lower abundance of Bacteroidetes members (56)</td>
</tr>
<tr>
<td>Ecz (PD) and/or slgE+ until 3y</td>
<td>PC: 606 high risk infants (DE)</td>
<td>qPCR</td>
<td>Increased prevalence of Clostridium cluster I at ages 5 and 13w associated with EcZ, but not slgE+ (69)</td>
</tr>
<tr>
<td>FA and SPT+, food challenge and/or slgE+, age from 2-11m</td>
<td>CC: 34 allergic and 45 non-allergic infants (CN)</td>
<td>16S rRNA seq.</td>
<td>Lower abundance of phyla Bacteroidetes, Proteobacteria, and Actinobacteria and increase of Firmicutes (88)</td>
</tr>
</tbody>
</table>

CMPA: Cow's Milk Protein Allergy; CC/CS/PC: Case-controlled study/Cross-sectional study/Prospective cohort; DBPCFC+: positive for double-blind placebo control food challenge; DGGE/TTGE: denaturing gradient gel electrophoresis/ temporal temperature gel electrophoresis, EcZ: EcZema; FA: Food Allergy; FISH: Fluorescent In Situ Hybridisation; ISAAC: International Study of Asthma and Allergies in Childhood; (q)PCR: (quantitative) Polymerase Chain Reaction; PD: Physician-diagnosed; slgE+: positive serum specific IgE; SPT+: positive skin prick test; T-RFLP: terminal-restriction fragment length polymorphism, VLBW: very low birth weight infants.
Human milk is the natural source of nutrition in early life and exclusive breastfeeding is recommended for at least 6 months by WHO. In allergy breastfeeding is thought to be protective because of both the presence of numerous allergens in human milk that are absent from artificial milks and their tolerogenic presentation due to human milk feeding related factors such as antigen handling by maternal gut, allergens found in immune complexes in milk, the presence of tolerogenic immune mediators in milk, increased gut maturation and a microbiota favouring tolerance induction in breastfed infants (44). The latter is linked to the HMO, naturally present in human milk. Consequently, a significant number of studies have been performed with different types of prebiotic oligosaccharides, defined as non-digestible carbohydrates that reach the colon intact and are known for their ability to selectively stimulate the growth and or activity of intestinal bacteria that impact health positively as postulated by Gibson & Roberfroid (89). Interestingly, intervention with infant milk formulas containing a specific mixture of short-chain galacto-oligosaccharides and long-chain fructo-oligosaccharides (scGOS/lcFOS, 9:1 ratio, 8 g/L) till 6 months of age reduced the risk of atopic dermatitis and some allergic manifestations in infants with a familiar history of atopy not only at 6 months, but also at 2 and 5 years of age (90-92) and reduced the number of infectious episodes in healthy term infants (93) or infants with a high risk of developing allergy (90), underlining the importance of early nutrition on infant health. Efforts to prevent or manage atopic dermatitis and food allergy may prevent the onset of other atopic manifestations such as allergic asthma later in life. A study applying synbiotics gives indications that this may actually be possible. The combination of scGOS/lcFOS (9:1) and Bifidobacterium breve M16-V in a 12-week intervention in infants around 5 months of age showed reduced severity of atopic dermatitis in a subgroup of infants with elevated IgE levels but not in the whole study group. However, at one year of age it was found that the synbiotic group showed attenuated use of asthma medication and lower prevalence of asthma-like symptoms in the whole study group suggesting long-term effects of the intervention early in life (94). Microbial analysis of the dominant bacterial groups affected in this 12-week intervention showed an increase of bifidobacteria at the expense of mainly adult-type clostridial clusters XIV and clostridial clusters containing potential pathogens C.
Establishing a symbiosis

dificile and C. perfringens (95). Establishing such infant-type microbiota and reducing the adult-type clusters may lead to a more gradual diversification, while e.g. clostridial cluster XIV has been associated with atopic manifestations later in life (56). The exact mechanism of this synbiotic concept remains to be elucidated. Recently induction of galectin-9 (a soluble-type lectin expressed by IEC exhibiting binding specificity for \(\beta\)-galactosides) by this synbiotic concept has been suggested to be involved in suppression of IgE-mediated allergy (96). Galectin-9 was shown to neutralize IgE and to induce Th1 and Treg type immune responses and was indeed enhanced in serum of the synbiotic treated infants. The exact mechanism underlying induction of galectin-9 expression remains to be clarified, however the synergy shown for the combination of scGOS/lcFOS and Bifidobacterium breve M-16V in enhancing serum galectin-9 levels in mice, suggest a possible interaction between microbe-induced TLR signalling and direct interaction of scGOS/lcFOS with IECs. Recently in-vitro studies confirmed that galectin-9 is secreted by IEC apically exposed to TLR9 ligand (either synthetic or DNA derived from B. breve M-16V) in the presence of scGOS/lcFOS is involved in inducing Th1 and Treg immune responses (97). These results give important mechanistic insights and may be a promising target to prevent or treat allergic disease.

Concluding remarks

Clearly the first 1000 days in life are very important, since this is the period where we encounter external stimuli for the first time and the body is trained to respond to these stimuli. Longitudinal studies of this critical period are limited and include several confounding factors that complicate the identification of specific microbes associated with e.g. atopic disease. In the light of the recent revolution of next generation sequencing technologies we can gain important new insight how early-life events like type of feeding, mode of delivery, genetic background or geographical differences, may interfere with the colonisation pattern and therefore determine a predisposition to disease later in life. The challenge will be to go from taxonomic mapping to functionality of the microbiota. Omics-technologies, like transcriptomics, proteomics or metabolomics, will certainly catalyse our further
understanding of the intestinal microbiota. Our genome is more or less fixed, but still the environment can have a major impact on the development. Processes like epigenetics are particularly interesting and we are just starting to understand how DNA methylation and histone modification mechanisms can regulate gene expression and confer phenotypical changes. And where our genome is fixed, we can still influence the epigenome and our microbiota. Knowing the importance of the intestinal microbiota for human physiology, the incredible development of infants in the first years of life, and the concurrent colonisation of the body with microbes makes it reasonable to believe that the intestinal colonisation of early life may be very important for health also in later life. Whether immunological, metabolic or neurological, all these systems are developing at this period. Therefore, it is important to understand the impact of factors like early life nutrition, but also the increase of caesarean deliveries or the increasing use of antibiotics. Disturbances in early life may lead to altered growth, immune diseases like allergy, metabolic diseases like obesity or cardiovascular diseases and maybe even brain and behavioural problems. Nutrition in early life and acquiring the essential microbes is probably a critical factor in this process.
Establishing a symbiosis

References


62. Romagnani S. Coming back to a missing immune deviation as the main explanatory mechanism for the hygiene hypothesis. Journal of Allergy and Clinical Immunology. 2007;119(6):1511-3.


Establishing a symbiosis
CHAPTER 3

Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing

Sim K¹, Cox MJ², Wopereis H³, Martin R³, Knol J³, Li M-S¹, Cookson WO², Moffatt MF², Kroll JS¹.

*Joint first

¹Imperial College London, St. Mary’s Campus, London, United Kingdom
²Imperial College London, National Heart and Lung Institute, South Kensington Campus, London, United Kingdom
³Danone Nutricia Research, Wageningen, The Netherlands

Abstract

The 16S rRNA gene is conserved across all bacteria and as such is routinely targeted in PCR surveys of bacterial diversity. PCR primer design aims to amplify as many different 16S rRNA gene sequences from as wide a range of organisms as possible, though there are no suitable 100% conserved regions of the gene, leading to bias. In the gastrointestinal tract, bifidobacteria are a key genus, but are often under-represented in 16S rRNA surveys of diversity. We have designed modified, ‘bifidobacteria-optimised’ universal primers, which we have demonstrated allow detection of bifidobacterial sequence present in DNA mixtures at 2% abundance, the lowest proportion tested. Optimisation did not compromise the detection of other organisms in infant faecal samples. Separate validation using fluorescence in situ hybridisation (FISH) shows that the proportions of bifidobacteria detected in faecal samples were in agreement with those obtained using 16S rRNA based pyrosequencing. For future studies looking at faecal microbiota, careful selection of primers will be key in order to ensure effective detection of bifidobacteria.
Introduction

With the advent of next-generation sequencing, semi quantitative, in-depth characterisation of microbial communities that has never been practically possible is now becoming increasingly accessible to researchers. In samples from the gastrointestinal (GI) tract, use of universal primers for amplification of the bacterial 16S rRNA gene followed by pyrosequencing is beginning to reveal the role of the GI microbiome in diverse diseases such as obesity (1), atopic disease (2, 3), colonic cancer (4) and necrotizing enterocolitis (5). Two of the key questions surrounding the role of the GI microbiota in health are how the microbiota is involved in immunomodulation (6, 7), and how imbalance may lead to disease states. Organisms such as the bifidobacteria, which rapidly colonise the gastrointestinal microbiota in the first year of life are thought to be central in the establishment and maintenance of a ‘healthy microbiota’.

Universal PCR primers allow amplification, and therefore detection of all the bacteria in a mixed population. A number of primer sets amplifying different regions of the 16S rRNA gene exist and are in common use (8, 9). A truly universal primer pair that binds to the 16S rRNA of all eubacteria is impossible to design since the longest number of consecutive nucleotides in the gene that are 100% conserved is 11 (Escherichia coli 16S rDNA positions 788 to 798), and in general, the number of sequential absolutely conserved nucleotides in other regions of the gene is four (10). The decreased amplification efficiency due to differential annealing of universal primers when a heterogeneous template is used leads to bias against the detection of certain taxa (11). For example, even well designed primers matching over 95% of sequences in the Ribosomal Database Project (RDP) (12) from the dominant bacterial phyla present in the gut, may miss specific taxa; primer 967F (13) will detect less than five percent of Bacteroidetes whilst primer 1492R (14) detects only 61% of Actinobacteria and 54% of Proteobacteria (15). Mismatches towards the 3’ end are likely to lead to greater amplification inefficiency than that at the 5’ end (16). Pragmatic approaches to primer use are often taken, accepting that not all bacteria will be fully represented, but that between sample comparisons making use of the
same primer pair are valid and that particular organisms of interest are successfully amplified.

In order to address this issue, different approaches may be adopted to ensure that detection of the specific taxa of interest to the study are maximised. The universal primer set used can be optimised by either introducing a degenerate base pair at the positions of mismatch. Alternatively, taxa-specific primers can be added to the primer pool. Frank et al. (16) used a primer pool consisting of seven different primer sequences (fourfold-degenerate primers and three primers specific for amplifying Bifidobacteriaceae, Borrelia and Chlamydiales) and were able to dramatically increase the detection of genera which were previously missed from clinical samples. Increasing the number of degenerate bases in the primer set may however introduce a bias in the template to product ratios when a heterogenous template is used since templates with a greater GC content at the primer site will be preferentially amplified (17). Furthermore, inclusion of a large number of degenerate bases equates to dilution of the primer pool, and the number of templates which exactly match each primer sequence is reduced, resulting in a potential decrease in the overall annealing efficiency (16). Using an inosine residue at the mismatched positions is an alternative approach (10), but as it forms a stable bond with all four nucleotides, this may lead to erroneous PCR products (16).

Bifidobacteria

Bifidobacteria are considered to be a major component of the GI microbiota in healthy breast-fed infants (18, 19). This is mainly driven by a high level of complex oligosaccharides (10-12 g/L) available as a natural prebiotic in breast-milk (20). Their use as a probiotic, or their stimulation by adding prebiotics (synbiotics) has become increasingly widespread. Specific prebiotics or synbiotics added to infant milk formula have been shown to induce a more ‘breastfed-like’ microbiota with associated physiological changes (metabolic end products and pH) compared to standard formula (21, 22). These changes are considered as an important mechanism for the inhibition of pathogens in the gut (23). Used as a prophylactic infant feed supplement bifidobacteria have been found to be effective at reducing both the severity as well as the risk of developing rotavirus diarrhoea. Their use also appears to reduce the risk of
antibiotic-associated diarrhoea (24). Moreover, bifidobacteria may be beneficial in the treatment of atopic disease (25) and a synbiotic infant formula has been found to prevent asthma-like symptoms in infants with atopic dermatitis (26).

Bifidobacteria were found to constitute only a minor component of the faecal microbiota in healthy, full term infants (27). The authors acknowledge that this was surprising and speculated that this result might arise through the 8F universal primer having a three base pair mismatch against Bifidobacterium longum, and that the genus in general does not have 100% sequence identity to the 8F primer sequence. In our study, we have therefore sought to assess the impact of using a standard ‘universal’ primer set with one exactly matched to the target region of bifidobacteria, in detecting this genus.

We designed a ‘bifidobacteria-optimised’ universal primer set by modification of a well-established primer set 357F/926R, originally designed by the Muyzer group (28, 29) for denaturing gradient gel electrophoresis. Primer set 357F/926R is one of two primer pairs recommended by the NIH Human Microbiome Project protocols (30, 31) for 16S rRNA amplicon pyrosequencing. We demonstrate that our ‘bifidobacteria-optimised’ primer set increased the bifidobacteria detection rate in both pure DNA mixtures as well as faecal samples, without compromising the detection of other genera. In addition, we have independently confirmed the relative abundance of bifidobacteria detected using fluorescence in situ hybridisation (FISH).

**Results**

**Pyrosequencing**

Pyrosequencing of the standard mixes and the faecal samples was carried out in a single multiplexed run on the GS Junior platform and resulted in 85 126 reads. After denoising and chimera-removal 60 794 high quality reads remained and these were assigned to samples using the barcode sequences, 37 977 reads for faecal samples, 22 817 for the standard DNA mixtures.
DNA mixtures

Standard universal primers detected *Streptococcus pneumoniae* and *Moraxella catarrhalis* sequences in correct relative proportions in the DNA mixtures. The primers however, consistently failed to correctly quantify the bifidobacterial sequences present. The standard universal primers failed to amplify bifidobacterial DNA to a level above 1% in four out of the five mixtures, and the maximum proportion of bifidobacteria that was detected was 1.6%, even when the bifidobacterial DNA constituted 90% of the mixture.

This was in contrast to the relative proportions of species-specific reads obtained with ‘bifidobacteria-optimised’ universal primers, which correlated far better with the original proportions of the species’ DNA in the mixture ($R^2 = 0.955$) (Table 1, Figure 1). With the ‘bifidobacteria-optimised’ primers, bifidobacterial DNA could be detected at the lowest concentration tested (2%).

![Figure 1: Proportions of 454 sequencing reads obtained using both primer sets. Increased detection rate of *Bifidobacterium dentium* demonstrated using the ‘bifidobacteria-optimised’ universal primers (b) compared to regular universal primers (u).](image)
Table 1: Proportions of DNA in each mixture.

<table>
<thead>
<tr>
<th>Sample</th>
<th><em>Bifidobacterium dentium</em></th>
<th><em>Streptococcus pneumoniae</em></th>
<th><em>Moraxella catarrhalis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2%</td>
<td>49%</td>
<td>49%</td>
</tr>
<tr>
<td>2</td>
<td>15%</td>
<td>50%</td>
<td>35%</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>4</td>
<td>75%</td>
<td>5%</td>
<td>20%</td>
</tr>
<tr>
<td>5</td>
<td>90%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Faecal samples

Operational taxonomic unit (OTU) analysis

The most abundant taxa at phylum level were the Firmicutes and Actinobacteria, followed by Proteobacteria and Bacteroidetes, irrespective of which primer set was used. The ten samples all comprised of different numbers of OTUs and OTU abundances (Figure 2), but, the most striking difference was the increased number of bifidobacterial reads present in the sample set analysed with the ‘bifidobacteria-optimised’ universal primers.

Fluorescence in situ hybridisation (FISH) analysis

Table 2 shows the proportion of faecal bifidobacteria, expressed as a percentage of the total number of bacteria in faeces as enumerated by FISH and the relative read abundances by 454-sequencing.

Comparing data obtained with the two primer sets to the FISH using Pearson correlation shows significant correlation of FISH with the pyrosequencing using the ‘bifidobacteria-optimised’ primer set (Table 3). To confirm good agreement between two methods Bland-Altman agreement tests were performed (32). The agreement between two methods is tested by comparing the differences between two methods against the average of the methods. The results from bifidobacteria-optimised pyrosequencing against the FISH method shows agreement in determining the level of bifidobacteria in the faecal samples tested (Table 4).
Figure 2: Heatmap displaying the relative abundance of OTUs per sample. Samples are grouped by hierarchical cluster analysis on the x-axis and by neighbour-joining phylogenetic tree with nearest neighbour interchange on the y-axis. Samples amplified with ‘bifidobacteria-optimised’ primers are in red and with the standard primers in blue. Bifidobacterial OTUs are highlighted in the red box.
Improved detection of bifidobacteria

Table 2: Relative proportions of faecal bifidobacteria in ten faecal samples as determined by FISH and 454-sequencing using ‘bifidobacteria-optimised’ universal primers (926Rb) or regular universal primers (926R).

<table>
<thead>
<tr>
<th>Sample</th>
<th>926Rb</th>
<th>926R</th>
<th>FISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.2%</td>
<td>0.0%</td>
<td>0.3%</td>
</tr>
<tr>
<td>P2</td>
<td>81.1%</td>
<td>1.0%</td>
<td>61.2%</td>
</tr>
<tr>
<td>P3</td>
<td>69.0%</td>
<td>1.7%</td>
<td>70.9%</td>
</tr>
<tr>
<td>P4</td>
<td>63.5%</td>
<td>0.4%</td>
<td>75.8%</td>
</tr>
<tr>
<td>P5</td>
<td>0.2%</td>
<td>0.0%</td>
<td>0.6%</td>
</tr>
<tr>
<td>P6</td>
<td>62.7%</td>
<td>4.4%</td>
<td>67.3%</td>
</tr>
<tr>
<td>P7</td>
<td>74.1%</td>
<td>10.8%</td>
<td>47.5%</td>
</tr>
<tr>
<td>P8</td>
<td>90.6%</td>
<td>5.3%</td>
<td>75.0%</td>
</tr>
<tr>
<td>P9</td>
<td>16.9%</td>
<td>0.0%</td>
<td>10.4%</td>
</tr>
<tr>
<td>P10</td>
<td>8.0%</td>
<td>0.1%</td>
<td>67.0%</td>
</tr>
</tbody>
</table>

Table 3: Correlation matrix (Pearson) shows the Pearson correlation coefficients and p-values. Values in bold are different from 0 with a significance level alpha=0.05.

<table>
<thead>
<tr>
<th>Variables</th>
<th>926Rb</th>
<th>926R</th>
<th>FISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>926Rb</td>
<td>n/a</td>
<td>0.593 (p=0.071)</td>
<td>0.761 (p=0.011)</td>
</tr>
<tr>
<td>926R</td>
<td>0.593 (p=0.071)</td>
<td>n/a</td>
<td>0.297 (p=0.404)</td>
</tr>
<tr>
<td>FISH</td>
<td><strong>0.761 (p=0.011)</strong></td>
<td>0.297 (p=0.404)</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 4: P-values resulting from Bland-Altman agreement tests. Values in bold are different from 0 with a significance level alpha=0.05.

<table>
<thead>
<tr>
<th>Variables</th>
<th>926Rb</th>
<th>926R</th>
<th>FISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>926Rb</td>
<td>n/a</td>
<td><strong>0.0026</strong></td>
<td>0.8974</td>
</tr>
<tr>
<td>926R</td>
<td><strong>0.0026</strong></td>
<td>n/a</td>
<td><strong>0.0011</strong></td>
</tr>
<tr>
<td>FISH</td>
<td>0.8974</td>
<td><strong>0.0011</strong></td>
<td>n/a</td>
</tr>
</tbody>
</table>

Principal Coordinate Analysis and statistics

In order to ensure that detection of other organisms was not compromised or that abundance levels were not altered by using ‘bifidobacteria-optimised’ primers, principal coordinate analysis (PCoA) was performed. PCoA using the weighted UniFrac metric (33) (Figure 3a) (which takes into consideration both the presence/absence as well as abundance of sequences,) demonstrates clustering of samples by primer set used except for pairs P1 and P5 (circled). On OTU analysis, (Figure 2) these are shown to have very small or only moderate numbers of bifidobacteria present. Removing bifidobacterial sequences from the principal
coordinate analysis (Figure 3b) resulted in tight clustering of all pairs of samples. This indicates that the main differences between the two principal coordinate analyses are due to the detection of bifidobacteria, and that ‘bifidobacteria-optimised’ universal primers do not compromise the quantitative detection of other organisms.

Using a paired T-Test to compare OTUs and read abundance of the two sample sets (‘bifidobacteria-optimised’ universal primers vs. regular universal primers) there was a highly significant difference between the read abundance of bifidobacteria using ‘bifidobacteria-optimised’ primers compared to regular primers ($P = 0.039$, $t = 0.0026$, with Bonferonni correction for multiple testing), but no significant differences between any of the other OTUs ($P > 1.4$).

**Figure 3**: Principal Coordinate Analysis using the weighted UniFrac metric. (A) Sample pairs P9, P10 and in particular P1 and P5 cluster tightly together. These samples contain small or moderate numbers of bifidobacteria reads. (B) After removing bifidobacteria sequences from the analysis, all sample pairs cluster tightly showing that the main differences between the sets are due to the bifidobacteria sequences. U = regular universal primers (926R), B – ‘bifidobacteria-optimised’ universal primers (926Rb).
**Primers**

Primer specificity of the 926Rb primer was compared *in silico* against that of 926R using the Ribosomal Database Project’s (RDP) Probe Match tool. Only sequences longer than 1200 bp, defined as good quality by the RDP were included and 92.4 % of these were hit with 0 mismatches with primer 926R compared to 94.5 % with 926Rb. Although this overall increase was modest, the difference on looking specifically at the order Bifidobacteriales was very marked and highly significant: 926R hit just 0.2 % of sequences compared to 97.1 % with the ‘bifidobacteria-optimised’ primer.

**Discussion**

Appropriate primer selection in microbiota studies using a 16S rRNA approach is essential to enable faithful representation of the organisms present in the samples. The study of Palmer *et al.* (27) revealed that the overall efficiency of amplification of DNA from bifidobacterial species was eight-fold lower than that from non-bifidobacterial species using the 8F/1391R primer pair. Our results show that even a one base pair mismatch not at the 3’ end of a primer can lead to a dramatic failure to amplify these organisms at all.

It is well known that Gram-positive organisms (such as bifidobacteria) can be underrepresented in microbial profiling studies due to the presence of their thick cell wall (34). Due to concern that poor representation of bifidobacteria from faecal samples may be due to difficulties in cell lysis during DNA extraction, we first assessed target sequence recovery from pure DNA mixtures. We were able to demonstrate with the DNA mixtures that the bias observed against the detection of bifidobacteria was due to the PCR step. This was also confirmed by using FISH analysis which does not require cell lysis. From the FISH results, the bifidobacteria proportions present in the faecal samples were in agreement with those generated from our robust DNA extraction method combined with our ‘bifidobacteria-optimised’ universal primers and pyrosequencing.
Burgeoning interest in the development of the normal GI microbiota, and its impact on child and adult health, has led to increasing numbers of studies focusing on the bacterial colonisation of the gut (7). Metchnikoff’s (35) suggestion that it is “possible to adopt measures to modify the flora in our bodies and to replace the harmful microbes by useful microbes” over a hundred years ago has led to the concept of manipulating the GI microbiota to counter disease. Furthermore, the use of probiotics as a treatment or prophylaxis strategy not only for disease, but also for modulating the immune system has now become a focus of intense attention (36). Due to the escalating use of probiotics, the World Health Organization have published specific criteria that a probiotic must fulfil (37). One important quality of a probiotic is that it must be able to survive the GI tract, even if this is transient. This means that studies assessing the effectiveness of probiotics must be able to accurately detect in at least semi-quantitative fashion these probiotics organisms in the GI microbiota.

We have demonstrated that erroneous conclusions as to the presence or absence, or relative proportions of, bifidobacteria are likely if universal primers which do not sufficiently complement the target sequence are used. The primers we have designed are able to detect bifidobacteria at low level abundance and can be used semi-quantitatively without distorting the proportions detected of other genera. This primer set can be successfully used in 16S rRNA pyrosequencing-based GI microbiota studies.

Materials and methods

PCR primer design

Primers 357F/926R (357F - CCTACGGGAGGCAGCAG, 926R - CCGTCAATTCMTTTRAGT) were assessed for specificity using the ARB software package (38) and the SILVA 108 SSU Ref 16S rRNA database release (39). Almost all bifidobacteria (as well as some closely related Actinobacteria) were found to have a one base pair mismatch (C → T) to the 926R primer (CCGTCAATTCMTTTRAGT, mismatch in bold).
A new ‘bifidobacteria-optimised’ universal primer (926Rb) was therefore synthesised in which a T/C redundancy was incorporated at the mismatch position: CCGTCAATTYMTTTTRAGT (where Y is T or C).

**Standard DNA Mixtures**

DNA was extracted from pure cultures of *Bifidobacterium dentium*, *Streptococcus pneumoniae* and *Moraxella catarrhalis* using the MP Bio Fast Soil DNA kit®. An extra bead-beating step (40 seconds, speed 6.0 m/s using the FastPrep® FP120 Instrument, MP Biomedicals) was incorporated in order to ensure efficient lysis.

Total genomic DNA concentration was measured using the Quant-iT, PicoGreen DNA assay (Invitrogen).

Pre-defined mixtures using varying proportions of *Bifidobacterium dentium*, *Streptococcus pneumoniae* and *Moraxella catarrhalis* DNAs were prepared (Table 1). All three bacterial strains have 4 copies of the 16S rRNA operon. Consequently, gene copy number is dependent only on the number of bacteria present.

**Faecal samples**

Faecal samples were collected from five healthy term infants at two time points, 4 weeks and 26 weeks of age. The samples were immediately frozen (-12°C to -20°C) prior to transfer (within one week of sampling) to -80°C prior to evaluation.

Total DNA was extracted as described by Matsuki *et al.* (40) except that DNA was re-suspended in 0.1 ml of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

**Barcoded 16S rRNA PCR and pyrosequencing**

The V3-V5 regions of the bacterial 16S rRNA gene were amplified using primers 357F with adaptor B from 454 Life Sciences for pyrosequencing: 5’ CTATCCCCTGTGTGCCTTGGCAGTCTCAGCCTACGGGAGGCAGCAG 3’, and either the standard 926R or the ‘bifidobacteria-optimised’ primer 926Rb (Y in place of C, in bold): 5’ CCATCTCATCCCTGCGTGTCTCCGACTCAG NNNNNNNNNNNNN CCGTCAATTCTMTTTTRAGT 3’. In addition the reverse primers included the 454 Life Sciences adaptor A and a unique 12 base-pair error-correcting Golay (41) barcode.
(denoted by ‘Ns’, see Table S1). This allows multiplexing of samples in a single run. Primers were obtained from Eurofins MWG Operon (Ebersberg, Germany) and HPSF purified.

PCR was carried out in quadruplicate to reduce random mispriming bias (17), and no-template PCR controls were included. Each 25µl reaction contained 1µL each of forward and reverse primers (10µM), 1µl of template DNA, 0.25µl of 5U/µl FastStart HiFi Polymerase (Roche, Mannheim, Germany), 1µl of 20g/mL BSA (Sigma, Dorset, United Kingdom), and 6.5µl of 5M Betaine (Sigma). PCR reactions were assembled within a PCR hood in under clean conditions. Thermal cycling consisted of initial denaturation at 94°C for 2 minutes followed by 30 cycles of denaturation at 94°C for 20 seconds, annealing at 50°C for 30 seconds, and extension at 72°C for 5 minutes. The replicate amplicons were pooled, PEG precipitated (42) (20%, MW 8 000 g/mol) and visualized by staining with ethidium bromide (10mg/mL) on a 1.0% agarose gel.

**Amplicon quantitation, pooling and pyrosequencing**

Amplicons were combined in a single tube in equimolar concentrations. The pooled amplicon mixture was purified twice (AMPure XP kit, Agencourt, Takeley, United Kingdom) and the cleaned pool requantified using the PicoGreen assay. This pool was then diluted in TE such that it contained 10^5 molecules/µl. 30µl of this pool was added to the emulsion PCR reaction to attain a ratio of 0.3 molecules of amplicon per bead. Pyrosequencing was carried out on a 454 Life Sciences GS Junior instrument (Roche) following the Roche Amplicon Lib-L protocol.

**Bioinformatics**

Shotgun processed data was denoised using AmpliconNoise (43) as part of the QIIME (44) (Quantitative Insights Into Microbial Ecology) package followed by chimera-removal with Perseus (43). The sequences were aligned using the Greengenes core alignment set as reference (DeSantis et al 2006) and clustered at 97 % sequence identity into OTUs. Representative sequences (most abundant) for each OTU were selected and classified using the Ribosomal Database Project Classifier. Rarefaction was performed so that the number of reads per sample would be identical. Beta diversity assessment of the reads obtained from the faecal samples
using the two primer sets was carried out using the weighted UniFrac metric to
generate principal coordinate analyses. Identification of OTUs that were significantly
different in abundance was carried out in QIIME using a paired T-test with Bonferroni
correction.

**Fluorescence in situ hybridisation**

To enumerate the *Bifidobacterium* genus by means of FISH the 16S
rRNA-targeted oligonucleotide probe: Bif164-mod 5'- CATCCGGYATTACCACCC-3'
was used (45, 46). The probe was commercially synthesized and 5'-labelled with Cy3
(Biolegio B.V., Nijmegen, the Netherlands).

The FISH analysis was performed according to the method of Thiel (47), with
some modifications. Briefly, portions of each faecal sample were fixed with 3%
paraformaldehyde at 4°C for 16 hours. Following fixation, 1 ml of the cell suspension
was centrifuged at 8 000 x g for 3 min and the cell pellet resuspended in 500 μl of PBS
buffer, mixed with 500 μl of ethanol and then stored at -20°C until use. 3 μl of the
fixed-cell suspension of the appropriate dilution (80, 160, 320 and 640-fold dilutions)
was applied to chrome gelatine coated 18-well slides (Cel-Line HTC Super cured,
Thermo Scientific Portsmouth, NH) and the cell smears were dehydrated for 3 min
each in 60%, 80% and 96% ethanol. After hybridization of the probe at 50°C for 16
hours, the slides were washed, dried, counterstained with 4',6-diamidino-2-phenylindole (DAPI) and mounted with Citifluor AF1 (Citifluor Ltd,
London, United Kingdom).

Image acquisition and image analysis was performed using the scan^R
screening station (Olympus, Hamburg, Germany). The count and percentage of
labelled bacteria per sample was determined in 25 positions divided over the well by
counting all DAPI-stained bacteria and all doubly stained bacteria (DAPI and Cy3) in
the same field of view using a quadruple band filter set (Set 84000, Chroma
Technology Corp., Brattleboro, VT, USA).
Data Availability

MIMARKS compliant (39) 16S rRNA amplicon data for the faecal samples has been deposited at MG-RAST (48) under accession numbers 4483884.3 to 4483903.3 (static link http://metagenomics.anl.gov/linkin.cgi?project=329).

Ethics Statement

The National Research Ethics Service (NRES) Committee London (Fulham) approved all protocols and procedures (ref: 10/H0711/39) and parents gave their full written consent for faecal sample collection.

Acknowledgements

Prof William Wade (King’s College, London, United Kingdom) – for providing the *Bifidobacterium dentium* stocks. Mr. Geraint Barton (Bioinformatic Support Service – Imperial College London, United Kingdom) – for producing the heatmap and phylogenetic tree figures. The Winnicott Foundation for funding this study.
**Supporting information**

Table S1: Barcoded primer sequences for 16S rRNA PCR.

<table>
<thead>
<tr>
<th>SampleID</th>
<th>BarcodeSequence</th>
<th>LinkerPrimerSequence</th>
<th>MID</th>
</tr>
</thead>
<tbody>
<tr>
<td>uP1</td>
<td>AACGCACGCTAG</td>
<td>CCGTCAATTCMTTRAG</td>
<td>MID1</td>
</tr>
<tr>
<td>uP2</td>
<td>ACTGTGTTCTAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID2</td>
</tr>
<tr>
<td>uP3</td>
<td>ACCAGACGATGC</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID3</td>
</tr>
<tr>
<td>uP4</td>
<td>ACTGCTCATGGAT</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID4</td>
</tr>
<tr>
<td>uP5</td>
<td>ACTCTGATTTAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID5</td>
</tr>
<tr>
<td>uP6</td>
<td>AGACCAGTCAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID6</td>
</tr>
<tr>
<td>uP7</td>
<td>AGCAGACGACTCA</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID7</td>
</tr>
<tr>
<td>uP8</td>
<td>ACCAGACTCTCA</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID8</td>
</tr>
<tr>
<td>uP9</td>
<td>ACCAGCGACTAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID9</td>
</tr>
<tr>
<td>uP10</td>
<td>ACGGAATCTGAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID10</td>
</tr>
<tr>
<td>uB1</td>
<td>AGCTTGTGACAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID11</td>
</tr>
<tr>
<td>uB2</td>
<td>AACTGTGATCAC</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID12</td>
</tr>
<tr>
<td>uB3</td>
<td>ACCGAGAGTCTA</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID13</td>
</tr>
<tr>
<td>uB4</td>
<td>ACGGTAGAGTCA</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID14</td>
</tr>
<tr>
<td>uB5</td>
<td>ACTCGATTTGAG</td>
<td>CCGTCAATTCTMTTRAG</td>
<td>MID15</td>
</tr>
<tr>
<td>bP1</td>
<td>AGACTGCGTACTC</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID1b</td>
</tr>
<tr>
<td>bP2</td>
<td>AGCACAGTCCAG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID2b</td>
</tr>
<tr>
<td>bP3</td>
<td>AGAGACAGACTGT</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID3b</td>
</tr>
<tr>
<td>bP4</td>
<td>AAGAGATGTCG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID4b</td>
</tr>
<tr>
<td>bP5</td>
<td>AGAGCGAGATTG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID5b</td>
</tr>
<tr>
<td>bP6</td>
<td>ACGTACTCAGTG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID6b</td>
</tr>
<tr>
<td>bP7</td>
<td>ACTCGCACAGGA</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID7b</td>
</tr>
<tr>
<td>bP8</td>
<td>AGAGAGCAAGTG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID8b</td>
</tr>
<tr>
<td>bP9</td>
<td>AGCATATGAGA</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID9b</td>
</tr>
<tr>
<td>bP10</td>
<td>AGGTACACGACG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID10b</td>
</tr>
<tr>
<td>bB1</td>
<td>AAGCTGCTGCAG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID11b</td>
</tr>
<tr>
<td>bB2</td>
<td>ACAGCTAGCTTG</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID12b</td>
</tr>
<tr>
<td>bB3</td>
<td>ACCTGTCTCTCT</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID13b</td>
</tr>
<tr>
<td>bB4</td>
<td>AGCTGCTGTAGCA</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID14b</td>
</tr>
<tr>
<td>bB5</td>
<td>AGAGCAAGAGCA</td>
<td>CCGTCAATTTYMTTRAG</td>
<td>MID15b</td>
</tr>
</tbody>
</table>
References

Improved detection of bifidobacteria
CHAPTER 4

Intestinal microbiota in high-risk infants: Effects of prebiotics and role in eczema development

Wopereis H\textsuperscript{1,2}, Sim K\textsuperscript{3}, Shaw A\textsuperscript{3}, Warner JO\textsuperscript{3}, Knol J\textsuperscript{1,2}, Kroll JS\textsuperscript{3}.

\textsuperscript{1}Danone Nutricia Research, Utrecht, the Netherlands
\textsuperscript{2}Laboratory of Microbiology, Wageningen University, Wageningen, the Netherlands
\textsuperscript{3}Department of Medicine, Section of Paediatrics, Imperial College London, London, UK

Journal of Allergy and Clinical Immunology 2018;141(4):1334-42 e5. (minor modifications).
Abstract

**Background:** The development of gut microbiota in infancy is important in the maturation of the immune system. Deviations in colonization patterns have been associated with allergic manifestations, but exact microbiome dysfunctions underlying allergies remain unclear. We studied the gut microbiota of 138 infants at increased risk of developing allergy, participating in a clinical trial investigating the effectiveness of a partially hydrolyzed protein formula supplemented with non-digestible oligosaccharides (pHF-OS) on the prevention of eczema.

**Objective:** The effects of the interventions and breastfeeding on fecal microbiota were investigated. Additionally, we aimed to identify microbial patterns associated with the onset of eczema.

**Methods:** Bacterial taxonomic compositions in the first 26 weeks of life were analyzed using 16S rRNA-gene sequencing. Additionally, fecal pH and microbial metabolites were measured.

**Results:** Fecal microbial composition, metabolites and pH of infants receiving pHF-OS was closer to breastfed infants than to infants receiving standard cow’s milk formula. Infants with eczema by 18 months showed discordant development of bacterial genera of *Enterobacteriaceae* and *Parabacteroides* spp. in the first 26 weeks, as well as decreased acquisition of lactate-utilizing bacteria producing butyrate, namely *Eubacterium* and *Anaerostipes* spp., supported by increased lactate and decreased butyrate levels at 26 weeks.

**Conclusions:** We showed that a pHF with specific prebiotics modulated the gut microbiota closer to that of breastfed infants. Additionally, we identified a potential link between the microbial activity and onset of eczema, which may reflect a suboptimal implementation of gut microbiota at specific developmental stages in infants at high-risk for allergy.
Introduction

Gut microbial development in the first years of life occurs concomitantly to the development of our cognitive, metabolic and immune systems, and form an interactive signaling network (1). The gradual diversification towards a relatively stable adult-like composition is a dynamic process influenced by several environmental factors such as birth mode, gestational age at birth and early life nutrition. Alterations in the early colonization process, e.g. through antibiotic-use or C-section delivery, have been associated with development of allergic manifestations later in life, but the exact microbiome dysfunction underlying this disease remains unclear (2).

Eczema is typically the first allergic manifestation to appear, and its incidence is still increasing in many countries around the world. Considering the concurrent maturation of the immune system, the succession of species into a stable adult-type community, a process naturally guided by human milk, may be critically important. Breastfeeding is thought to protect against the development of allergy, via the presence of allergens and immune mediators in human milk that are absent from artificial milks (3), as well as via the presence of human milk oligosaccharides stimulating a gut microbiota which may favor tolerance induction (4).

Therefore, several studies have been performed with different types of infant formulas (IF) enriched with non-digestible oligosaccharides, known as prebiotics.
Prebiotics typically reach the colon intact and selectively stimulate the growth and activity of specific beneficial members of the microbiota (5, 6). Interestingly, significant reductions in eczema risk were observed in intervention trials with IF supplemented with specific prebiotics (7-9).

The present study arises from a parent registered study (PATCH trial) investigating the effects of a partially hydrolyzed formula containing specific oligosaccharides (pHF-OS) on the prevention of eczema in infants at increased risk to develop allergy. The pHF-OS was shown to induce hypo-antigenic and immune-modulatory effects including increased regulatory T cell numbers (Treg), but did not reduce eczema incidence by 12 or 18 months, when compared with infants receiving standard cow’s milk formula (10).

In this work, 16S ribosomal RNA (rRNA)-gene sequencing was applied to obtain an in-depth characterization of the microbiota composition of feces collected at 4 and 26 weeks of age in a subset of vaginally born infants, including breastfed infants (n=30) and infants randomized to receive pHF-OS (n=51) or standard cow’s milk formula (control, n=57). In addition, fecal pH, levels of lactate and short-chain fatty acids (SCFA) were determined in the stool specimens collected at 4, 12 and 26 weeks of age.

There were two primary aims of this study: (i) to investigate whether a pHF-OS could modulate the developing gut microbiota closer to that of breastfed infants, and (ii) to identify patterns in the developing gut microbiota that may be implicated in the onset of eczema. The findings in this study confirm the major influence of early life nutrition on the assembly of the gut microbiota and provide new insights in how deviations in this assembly are associated with the development of eczema.

Materials and methods

Study design and fecal sample selection

This study arises from a double-blind, randomized, controlled parallel-group nutritional intervention trial in infants with a parental history of allergic disease, conducted in 10 specialist pediatric centers in Australia, Singapore, England and
Ireland from April 2006 to March 2011 as described elsewhere (10). In total 1047 infants were recruited and consisted of three groups of participants. The breastfed reference group consisted of exclusively breastfed infants for the first 18 weeks of life. Participants who chose to stop breastfeeding or to supplement with formula before 18 weeks of age, were randomized to receive either partially hydrolyzed whey protein-dominant infant formula containing short-chain galacto-oligosaccharides and long-chain fructo-oligosaccharides (scGOS/lcFOS) (9:1) (0.68 g/100 ml) and pectin-derived acidic oligosaccharides (pAOS) (0.12 g/100 ml), or standard cow’s milk formula (control), up to 26 weeks of age. Information of trial ethics approvals, monitoring and regulatory compliance are summarized in the Online Repository (PATCH trial, study registration ISRCTN65195597 14th February 2006).

Fecal samples for microbial analysis were selected from infants that met the following criteria: infants (I) were randomized to investigational formula before 4 weeks of age, or were part of the breastfed reference group (n=942 infants), (II) vaginally-born (normal or instrumental delivery) (n =673 infants), (III) with stool specimens available at 4 and 26 weeks of age (n=324 infants). A subset of these infants was randomly selected to obtain 60 infants for each formula group and 30 breastfed reference infants (n=150 infants).

Sample collection and preparation

The infants’ fecal samples were collected by the parents into 10 ml stool containers (Greiner Bio-One, Kremsmünster, Austria), immediately frozen (−12°C to −20°C) and transported within three months to the hospital. Upon arrival at the hospital and prior to evaluation at the laboratory, samples were kept and transported at -80°C. Frozen stool samples were defrosted on ice and stool pH was measured using a pH meter equipped with a glass-body pH electrode (Mettler-Toledo, Columbus, USA). Fecal samples were 10-fold diluted in PBS buffer (150 mM NaCl, 10 mM Na₂HPO₄, 20 mM NaH₂PO₄, pH 7.4), and 5-10 glass beads (3 mm in diameter) were added to homogenize the sample by vortexing for 3 minutes, followed by centrifugation (300×g for 1 minute). Several 1 ml portions of supernatant were stored at -80°C for downstream processing.
**16S rRNA gene sequencing and bioinformatics**

Fecal suspensions of 200 µl were mixed with 450 µl DNA extraction buffer (100 mM Tris-HCl, 40 mM EDTA, pH 9.0) and 50 µl of 10% sodium dodecyl sulfate. Phenol-chloroform extractions combined with beat-beating were performed as described by Matsuki, et al. (11) except that extracted DNA was re-suspended in 0.1 ml of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

The V3-V5 regions of the 16S rRNA gene were amplified using forward primer 357F, and a ‘bifidobacteria-optimised’ reverse primer 926Rb (12). 16S rDNA PCR was carried out in quadruplicate and pyrosequenced in three 454 GS FLX (Roche, Branford, CT, USA) runs as previously described (12). Raw pyrosequencing data for all samples has been deposited in the Sequence Read Archive under accession number PRJEB19801.

The ‘Quantitative Insights Into Microbial Ecology’ (QIIME) v1.5.0 package was used to analyse sequence data (13) as previously described (12), except that alignments were carried out using the SILVA rRNA database (SSU_REF108) (14).

**Fecal SCFA and lactic acid**

Fecal suspensions were thawed on ice and centrifuged for 10 minutes at 14,000×g. Then, 350 µl supernatant was inactivated by heating for 10 minutes at 100°C, followed by centrifugation. A portion of the supernatant was used to quantitatively determine the SCFAs: acetic, propionic, n-butyric, iso-butyric, iso-valeric and n-valeric acid by gas chromatography, as described previously (15). Another portion of the supernatant was used to enzymatically analyze levels of lactate using a D-/L-lactic acid assay kit (Megazyme, Wicklow, Ireland).

**Data handling and statistical analyses**

Multivariate statistical analyses of 16S rRNA-gene sequencing data (taxonomic dataset) and levels of SCFAs and lactate (metabolite dataset) were performed by a combination of constrained ordination methods using the Canoco 5 software (16), and differential abundance testing using the R-package MetagenomeSeq (17).

Counts of Operational Taxonomic Units (OTUs) were aggregated at genus level (resulting in 142 genera), normalized by total sum scaling and log2-transformed.
Genera present in less than 10 samples were discarded to remove sparse taxa, resulting in 58 features that were used as input for statistical analyses.

Monte Carlo Permutation tests (MCPT) were used to evaluate statistical significance (P≤0.05) of explanatory variables (constraints) in the ordination analyses performed. Benjamini–Hochberg false-discovery rate was used to account for multiple comparisons (18), with significance for adjusted P-values (P-adj) at 0.05, except for taxonomic features associated with eczema development, for which significance was considered at 0.1.

Canonical correspondence analysis (CCA) was applied on the taxonomic dataset, as the unimodal model was found to best fit the relative abundances of the bacterial genera (19). Forward selection was applied to identify sample covariates that best explained the variation in microbial taxonomic composition (20), and these were subsequently used in all multivariate models and comparisons performed, either as explanatory variables or as covariates. Temporal changes of bacterial metabolites were assessed using a linear ordination method, known as principal response curves (PRC) (21), as the linear model was found to best fit the bacterial metabolite dataset.

Univariate data analyses were performed using GraphPad Prism version 6.02 for Windows (GraphPad Software, La Jolla, California, USA), applying Mann-Whitney test for two-group comparisons, and one-way ANOVA with Bonferroni’s multiple comparisons test for three-group comparisons, with significance at 0.05.

Results

Effects of pHF-OS on microbial richness and diversity.

In total, 12 of the 150 selected infants did not have stool specimens available at 4 and 26 weeks. The characteristics of the remaining 138 participants are summarized in Table 1. Pyrosequencing was performed on stool specimens collected at 4 and 26 weeks of age, representing the specimens taken after randomization and at the end of the intervention period, respectively. Most of the randomized infants were already receiving formula before 4 weeks of age (Table 1). A total of 8 specimens collected at 4 weeks had insufficient amounts for preparation. The
remaining 268 samples were successfully sequenced with a mean sequence depth of 6211 reads and a mean read length of 521 nucleotides (Table S1).

Microbial richness and diversity of the different feeding groups was assessed at a sequence depth of 1636 reads. Richness was expressed as the number of unique OTUs (observed species) and diversity was assessed using the shannon index for diversity (22). No differences between the randomization groups were observed, but the breastfed reference group showed a significantly lower richness at 26 weeks when compared to control-fed infants (1-way ANOVA, P<0.05) (Figure 1A) and lower diversity at both 4 and 26 weeks (1-way ANOVA, P<0.01 and P<0.05, respectively) (Figure 1B).

**Figure 1:** Box-whisker plots (Tukey method) summarizing microbial intestinal richness (Fig. 1A) and shannon diversity (Fig. 1B) at 4 and 26 weeks of age in breastfed infants (B) and infants receiving pHF-OS (A) or control formula (C). The centre line shows the median, the + denotes the mean, and the boxes cover the 25th and 75th percentiles with the whiskers extending to the data points, which are no more than 1.5 times the length away from the box. Points outside the whiskers represent outlier samples. Comparisons were done at a sequencing depth of 1636 reads per sample. Statistics were performed by a one-way ANOVA with Bonferroni’s multiple comparison correction (* = P <0.05, ** = P<0.01).
Table 1: Characteristics of the study population.

<table>
<thead>
<tr>
<th></th>
<th>Control (n=57)</th>
<th>pHF-OS (n=51)</th>
<th>Breastfed (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Started formula before 4w</td>
<td>54 (95%)</td>
<td>48 (84%)</td>
<td>N/A</td>
</tr>
<tr>
<td>Male Sex</td>
<td>23 (40%)</td>
<td>30 (59%)</td>
<td>18 (60%)</td>
</tr>
<tr>
<td>Birthweight (g)</td>
<td>3317 (502)</td>
<td>3461 (442)</td>
<td>3606 (393)</td>
</tr>
<tr>
<td>Both parents allergic</td>
<td>11 (19%)</td>
<td>6 (12%)</td>
<td>11 (37%)</td>
</tr>
<tr>
<td>Only mother allergic</td>
<td>33 (58%)</td>
<td>29 (57%)</td>
<td>11 (37%)</td>
</tr>
<tr>
<td>Only father allergic</td>
<td>13 (23%)</td>
<td>16 (31%)</td>
<td>8 (27%)</td>
</tr>
<tr>
<td>Pet at home</td>
<td>11 (19%)</td>
<td>22 (43%)</td>
<td>7 (23%)</td>
</tr>
<tr>
<td>At least one sibling</td>
<td>28 (49%)</td>
<td>27 (53%)</td>
<td>22 (73%)</td>
</tr>
<tr>
<td>Vaginal delivery</td>
<td>46 (81%)</td>
<td>41 (80%)</td>
<td>25 (83%)</td>
</tr>
<tr>
<td>Instrumental delivery</td>
<td>11 (19%)</td>
<td>10 (20%)</td>
<td>5 (17%)</td>
</tr>
<tr>
<td>Caucasian ethnicity</td>
<td>16 (28%)</td>
<td>28 (55%)</td>
<td>25 (83%)</td>
</tr>
<tr>
<td>Asian ethnicity</td>
<td>35 (61%)</td>
<td>21 (41%)</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>Other ethnicity</td>
<td>6 (12%)</td>
<td>2 (4%)</td>
<td>3 (10%)</td>
</tr>
<tr>
<td>Australia</td>
<td>13 (23%)</td>
<td>14 (27%)</td>
<td>13 (43%)</td>
</tr>
<tr>
<td>Ireland</td>
<td>4 (7%)</td>
<td>10 (20%)</td>
<td>5 (17%)</td>
</tr>
<tr>
<td>Singapore</td>
<td>35 (61%)</td>
<td>21 (41%)</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>UK</td>
<td>5 (9%)</td>
<td>6 (12%)</td>
<td>10 (33%)</td>
</tr>
<tr>
<td>Eczema by 12 months</td>
<td>15 (26%)</td>
<td>17 (33%)</td>
<td>12 (40%)</td>
</tr>
<tr>
<td>Eczema by 18 months</td>
<td>18 (32%)</td>
<td>20 (39%)</td>
<td>14 (47%)</td>
</tr>
<tr>
<td>Antibiotics by 26 weeks</td>
<td>6 (11%)</td>
<td>12 (24%)</td>
<td>3 (10%)</td>
</tr>
</tbody>
</table>

Continuous data are presented as mean (standard deviation). N/A: Not Applicable.

**pHF-OS modulates the microbial composition closer to breastfed infants.**

The majority of OTUs were assigned to 4 phyla, namely Actinobacteria (65.0%), Firmicutes (24.8%), Proteobacteria (8.6%) and Bacteroidetes (1.5%) (Table S2). The weighted UniFrac distance metric, which is based on phylogenetic distances and relative abundances of bacterial taxa in a pairwise comparison of samples (23), was used to assess the overall (dis)similarity in bacterial composition comparing pHF-OS with the control group and their respective distances to the breastfed reference group. Gut microbial compositions of infants receiving pHF-OS were found to be significantly more similar to those of breast-fed infants than the compositions of control group to breast-fed infants at 26 weeks (1-way ANOVA, P<0.0001) (Figure 2A).
In order to identify sample covariates that best explained the taxonomic composition of the fecal samples, canonical correspondence analysis (CCA) was combined with forward selection of variables (20). This analysis identified time (age in weeks), ethnicity (Asian, Caucasian, other), feeding group (control, pH-OS, breastfed) and having siblings (yes/no) as the factors significantly explaining the taxonomic variation (P-adj<0.05, Table S3). MetagenomeSeq (17) was subsequently used to
assess which bacterial taxa accounted for the differences observed in weighted UniFrac at 26 weeks, while correcting for the covariates identified. Infants receiving pHF-OS were discriminated from control by increased levels of *Bifidobacterium* and decreased levels of *Clostridium* and an unassigned genus of *Lachnospiraceae* (Figure 2B-D).

**pHF-OS modulates the microbial activity closer to breastfed infants.**

Principal response curves (PRC) were used to assess temporal effects of the feeding groups on fecal compositions of SCFA and lactate in stool specimens collected at 4, 12 and 26 weeks, respectively.

---

**Figure 3:**

**A**, Principal Response Curves showing changes in bacterial metabolites (levels of SCFA and lactic acids) across time and its interaction with the different feeding groups. The horizontal axis represents time and the vertical axis the PRC score values. Fecal metabolite composition of infants receiving control formula (C) was used as reference level and has zero PRC values and so its curve lays over the horizontal axis. Changes for infants being breastfed (B) or receiving pHF-OS (A) are shown as response curves relative to this reference. Metabolite response scores are shown on the separate vertical (one-dimensional) plot. The multiple of the PRC score with the response score provides a quantitative interpretation as well as the direction of the change at the respective timepoints (4, 12 and 26 weeks). Significance of the interaction, corrected for covariates, was tested using MCPT (P=0.002, 499 permutations).

**B**, Boxplots summarizing stool pH at 4, 12 and 26 weeks for the different feeding groups. Statistical comparison was performed by a one-way ANOVA with Bonferroni’s multiple comparison correction comparing pHF-OS to control formula (*= P-adj<0.05, **= P-adj<0.01, ***= P-adj<0.001).
The interaction of feeding group with time, corrected for ethnicity and having siblings, was significant on the first constrained axis (explained variation = 80.3%, MCPT: P=0.002 with 499 permutations). Both the metabolite composition of breastfed infants and infants receiving pHF-OS were characterized by increased proportions of D- and L-lactate at 4, 12 and 26 weeks and decreased proportions of propionate, butyrate and branched-chain SCFA (iso-butyrate and iso-valerate) in contrast to the control group (Figure 3A). Moreover, infants receiving pHF-OS were found to have significantly decreased stool pH at all timepoints compared to infants receiving control formula with ranges similar as observed for breastfed infants (Figure 3B).

**Aberrant temporal dynamics in infants developing eczema in the first 18 months of life.**

In total, 52 of the 138 infants developed eczema in the first 18 months of life (Table 1). No differences in bacterial richness or diversity were observed at 4 or 26 weeks comparing eczematous to non-eczematous infants (Figure S1). PRC were used to investigate the fecal metabolite composition over time for infants developing eczema as compared to infants that did not, while correcting for the covariates identified (ethnicity, feeding group and having siblings). Significant temporal differences on the first constrained axis produced (explained variation = 84.4%, MCPT: P=0.034 with 499 permutations) were observed for the interaction (Figure 4A). Differential dynamics were most pronounced from 12 to 26 weeks of age. Infants developing eczema were characterized by decreased levels of both isomers of lactate and increased levels of propionate and butyrate at 12 weeks. This pattern was reversed at 26 weeks of age, with infants developing eczema showing increased levels of lactate and decreased levels of propionate and butyrate. No significant differences were observed in stool pH (Figure 4B).
Figure 4: A, Principal Response Curves showing changes in bacterial metabolites (levels of SCFA and lactic acids) across time and its interaction with developing eczema. The horizontal axis represents time and the vertical axis the PRC score values. The fecal metabolite composition of infants not developing eczema (NOECZ) were used as reference level and has zero PRC values. The change for infants developing eczema in the first 18 months of life (ECZ) is shown as a response curve relative to this reference. Metabolite response scores are shown on the separate vertical (one-dimensional) plot. Significance of the interaction, corrected for the effects of feeding group, ethnicity and having siblings, was tested using MCPT (P=0.034, 499 permutations). B, Boxplots summarizing stool pH at 4, 12 and 26 weeks in infants developing eczema vs. healthy infants. Statistical comparison was performed by a one-way ANOVA with Bonferroni’s multiple comparison correction (not significantly different).

Decreased microbial conversion of lactic acid into butyrate in infants developing eczema.

MetagenomeSeq was used to assess differential abundances of bacterial taxa over time and which of those were discordant in infants with eczema compared to infants without eczema. The comparisons were corrected for the identified covariates (Figure 5A). Decreases of *Staphylococcus* and *Streptococcus* over time were observed in both healthy infants and infants developing eczema. Two genera of *Lachnospiraceae* (*Blautia* and an unassigned genus designated as “Other”) and *Erysipelotrichaceae* were found to increase over time in both groups. The genus *Bifidobacterium* was found to increase over time in healthy infants and to decrease in infants developing eczema, however this pattern was not significantly different comparing the two groups (P-adj=0.115). Discordant patterns over time were
observed for two genera of *Enterobacteriaceae*, namely *Enterobacter* (P-adj<0.001) and an unassigned genus of *Enterobacteriaceae* (P-adj<0.1).

**Figure 5:** A. Plot with bacterial taxa that change over time (from 4 to 26 weeks of age) in healthy infants (ECZ) and in infants developing eczema (NOECZ), as well as taxa that are differential over time in infants developing eczema compared to healthy infants. Data are shown as fold changes with standard deviations of log2-transformed data. Taxa are detailed at phylum (B=Bacteroidetes, F=Firmicutes, P=Proteobacteria), family and genus level, respectively. The contrasts were computed by MetagenomeSeq and corrected for the effects of feeding group, ethnicity and having siblings. Adjustment of significance values for multiple comparisons was done using Benjamini–Hochberg false discovery rate. Differential taxa with an adjusted P-value (P-adj) <0.05 for all contrasts, if present in >15% of the samples and a mean abundance >0.1%, are shown. Taxa plotted in dark grey background showed no significant (ns) temporal differences when comparing healthy infants with eczematous infants. Taxa plotted in light grey and white background were different with P-adj<0.1 and P-adj<0.05, respectively. 

B. Fecal concentration of D-lactic acid (D-LA) at 26 weeks in healthy infants compared to infants developing eczema.

C. Fecal concentration of L-lactic acid (L-LA) at 26 weeks in healthy versus eczema.

D. Fecal concentration of butyrate (BA) at 26 weeks in healthy versus eczema. Significance was tested with a Mann-Whitney two-group comparison (* = P-value<0.05, ** = P-value<0.001).
These discordances were mostly driven by increased levels of both genera in healthy infants at 4 weeks of age, and a subsequent decrease over time (Figure S2A and S2B). This pattern was less pronounced in infants developing eczema. Discordant development over time was also observed for *Parabacteroides* (P-adj<0.0001), with decreasing relative abundances in healthy infants, and generally increasing relative abundances in infants developing eczema (Figure S2C). Furthermore, the increased relative abundances over time observed for *Eubacterium* (P-adj<0.05) and *Anaerostipes* spp. (P-adj<0.1), were more pronounced in healthy infants as compared to infants developing eczema (Figure S2D and S2E). Both genera are associated with a specialist group of microbes known to convert lactate together with acetate into mainly butyrate, hence referred to as lactate-utilizing and butyrate producing bacteria (LUB) (24, 25). Interestingly, decreased levels of both D- and L-lactate (Figure 5B and 5C) and increased levels of butyrate (Figure 5D) were observed at 26 weeks of age in healthy infants compared to infants developing eczema.

**Discussion**

In this study, we found that a partially hydrolyzed protein formula, supplemented with a specific oligosaccharide mixture (pHF-OS), modulates the developing gut microbiota of infants towards a pattern closer to that of breastfed infants, both in bacterial taxonomic composition as in metabolite composition. Infants receiving pHF-OS from 4 to 26 weeks of age, showed increased relative abundances of the genus *Bifidobacterium*, which was contrasted by decreases of *Clostridium* and an unassigned genus of *Lachnospiraceae*, when compared to infants receiving standard cow’s milk formula. These modulations were reflected in marked differences in gut physiology, characterized by lower stool pH, increased proportions of lactate and decreased proportions of propionate, butyrate, iso-butyrate and iso-valerate.

Furthermore, we found that infants developing eczema in the first 18 months of life showed aberrant gut microbiota development in the first 26 weeks of life with significant temporal differences of the genus *Parabacteroides* and two genera of *Enterobacteriaceae*. These were found to decrease over time in healthy infants, a
pattern which was reversed or less evident in infants with eczema. Additionally, infants with eczema showed a lower establishment over time of lactate-utilizing bacteria known to produce butyrate (LUBs), namely *Eubacterium* and *Anaerostipes* spp., supported by significantly increased fecal concentrations of lactate, and decreased concentrations of butyrate at 26 weeks of age, a pattern which was independent of feeding group, ethnicity or having siblings.

To assess modulatory effects of nutrition and the role of specific microbes in developing eczema, it is of critical importance to consider the dynamic nature of the infant gut microbiota, and the environmental factors that influence its assembly (26). Hence, we excluded infants born by caesarean section to eliminate its confounding effects (27). Moreover, multivariate comparisons were controlled for sample covariates that best explained the fecal compositional variation, namely age, feeding group, ethnicity and having siblings.

Other studies investigating the gut microbiota in relation to developing allergy, typically reported differences at specific points in time, especially in the first weeks of life, with most commonly reduced diversity and richness associated, but often no specific microbes being identified (2). We did not observe differences in diversity or richness, but rather identified differences in the colonization patterns of infants with eczema compared with those without eczema by 18 months of age.

Some of these patterns were driven by early differences at 4 weeks of age, like the increased levels of two genera within the family of *Enterobacteriaceae* in subsequent healthy infants, which was followed by sharp decreases over time. Facultative anaerobic bacteria, like the *Enterobacteriaceae*, are typically high in the early colonization process before replacement with anaerobic bacteria. This pattern being less pronounced in infants developing eczema may indicate a reduced immune-stimulus from species within this family. These include several potential pathogens, which may be needed for adequate development of the adaptive immune system. A recent study by West, et al. (28) indeed associated low abundance of *Enterobacteriaceae* at 1 month, with an exaggerated immune response at 6 months. This deviating pattern was confirmed in infants subsequently developing allergy (28-30).
It is known, and confirmed in this study, that breastfed infants compared to formula fed infants have a lower stool pH, increased amounts of colonic acetate and lactate, and decreased amounts of propionate and butyrate. This reflects the dominance of *Bifidobacterium* species that produce acetate and L-lactate (31). This specific colonic environment, also observed for infants receiving pHF-OS, is known to form an effective chemical barrier against potential pathogens (32, 33), and is thought to contribute to the protective effects of breastfeeding against infections (34, 35).

We hypothesize that this typical infant-type colonic environment may also be crucial for the establishment of LUB, like *Eubacterium* and *Anaerostipes* spp, which were found to be poorly established at 26 weeks of age in infants developing eczema. Indeed, these infants showed first decreased levels of lactate at 12 weeks, followed by increased levels of lactate and decreased levels of butyrate at 26 weeks of age. This may indicate the importance of LUB in guiding the transition around the weaning-period (between 4-6 months of age) from a lactate and acetate-rich environment, towards a more adult-like butyrogenic milieu (2, 36). Moreover, the consumption of lactate may contribute to the development of a stable and healthy microbial ecosystem (37). Indeed lactate, although commonly detected in infant' feces, is undetectable in healthy adults (38), while accumulation of colonic lactate in adults has been associated with ulcerative colitis (39, 40).

The weaning-period also exposes the infant to an increasing level of antigens from the diet, which requires the immune system to adequately respond to these harmless substances. A study investigating the correlation of specific microbial signatures with the severity of eczema in 6-month old infants found an inverse correlation with the levels of butyrate-producing bacteria and suggested a role in the observed alleviation of symptoms (41). The role of SCFA on the host, especially butyrate, has received increased interest due to accumulating evidence from murine studies showing their regulatory effects on host immunity, including anti-inflammatory and anti-allergic effects (42-45). The establishment of butyrate-producing bacteria around 26 weeks of age, possibly specifically those cross-feeding on lactate and acetate, may therefore prove important for establishing and maintaining homeostasis with our immune system during this critical stage of development.
In conclusion, this study confirms the impact of early life nutrition on the establishment of the infant gut microbiota. Moreover, it indicates a potential link between the activity of the microbiota and the expression of eczema in early life. It emphasizes the importance of the microbial succession of species and metabolite cross-feeding, to develop a gut physiology that supports gut development, and supports development of normal immune responses towards environmental triggers. These observations could aid the development of optimal nutritional strategies to support the timely gut colonization of keystone species in the gradually diversifying infant gut.

Acknowledgement

We want to thank all infants and their families who took part in the study and all the PATCH investigators for their collaboration in this study.
Supporting information

Table S1: Sample size and sequence details for sequences passing quality filters summarized per feeding group.

<table>
<thead>
<tr>
<th>Feeding Group</th>
<th>Age</th>
<th>Count of subjects</th>
<th>Sum of reads</th>
<th>Average seq. depth</th>
<th>SD of seq. depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTROL</td>
<td>4w</td>
<td>55</td>
<td>326 169</td>
<td>5930</td>
<td>3300</td>
</tr>
<tr>
<td></td>
<td>26w</td>
<td>57</td>
<td>345 935</td>
<td>6069</td>
<td>2959</td>
</tr>
<tr>
<td>pHF-OS</td>
<td>4w</td>
<td>49</td>
<td>305 530</td>
<td>6235</td>
<td>4499</td>
</tr>
<tr>
<td></td>
<td>26w</td>
<td>51</td>
<td>299 308</td>
<td>5869</td>
<td>4113</td>
</tr>
<tr>
<td>BREASTFED</td>
<td>4w</td>
<td>26</td>
<td>166 838</td>
<td>6417</td>
<td>3950</td>
</tr>
<tr>
<td></td>
<td>26w</td>
<td>30</td>
<td>220 892</td>
<td>7363</td>
<td>4014</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>268</td>
<td>1 664 672</td>
<td>6211</td>
<td>3767</td>
</tr>
</tbody>
</table>

Table S2: Summary of gut microbial compositions of breastfed infants (B), pHF-OS (A) and control (C) at 4 and 26 weeks of age.

<table>
<thead>
<tr>
<th>Taxa</th>
<th>4 weeks</th>
<th>26 weeks</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td><strong>Phylum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Actinobacteria</strong></td>
<td>55.28</td>
<td>61.83</td>
<td>69.66</td>
</tr>
<tr>
<td>A</td>
<td>Bifidobacteriaceae</td>
<td>51.39</td>
<td>60.18</td>
</tr>
<tr>
<td>(32.66)%</td>
<td>(31.18)%</td>
<td>(35.13)%</td>
<td>(27.59)%</td>
</tr>
<tr>
<td>A</td>
<td>Coriobacteriaceae</td>
<td>3.56</td>
<td>1.19</td>
</tr>
<tr>
<td>(9.53)%</td>
<td>(3.14)%</td>
<td>(1.01)%</td>
<td>(6.92)%</td>
</tr>
<tr>
<td>A</td>
<td>Propionibacteriaceae</td>
<td>0.28</td>
<td>0.4</td>
</tr>
<tr>
<td>(0.86)%</td>
<td>(1.24)%</td>
<td>(1.01)%</td>
<td>(0.04)%</td>
</tr>
<tr>
<td>A</td>
<td>Other</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>(0.09)%</td>
<td>(0.27)%</td>
<td>(0.23)%</td>
<td>(0.04)%</td>
</tr>
<tr>
<td><strong>Total Bacteroidetes</strong></td>
<td>1.88</td>
<td>1.15</td>
<td>3.06</td>
</tr>
<tr>
<td>F</td>
<td>Bacteroidaceae</td>
<td>1.77</td>
<td>1.0</td>
</tr>
<tr>
<td>(7.9)%</td>
<td>(3.17)%</td>
<td>(5.21)%</td>
<td>(1.39)%</td>
</tr>
<tr>
<td>B</td>
<td>Porphyromonadaceae</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>(0.63)%</td>
<td>(0.83)%</td>
<td>(0.73)%</td>
<td>(0.04)%</td>
</tr>
<tr>
<td>B</td>
<td>Other</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>(0.02)%</td>
<td>(0.03)%</td>
<td>(1.9)%</td>
<td>(0.04)%</td>
</tr>
<tr>
<td><strong>Total Firmicutes</strong></td>
<td>31.09</td>
<td>23.34</td>
<td>8.63</td>
</tr>
<tr>
<td>F</td>
<td>Staphylococcaceae</td>
<td>0.52</td>
<td>0.44</td>
</tr>
<tr>
<td>(1.44)%</td>
<td>(1.17)%</td>
<td>(1.57)%</td>
<td>(0.06)%</td>
</tr>
<tr>
<td>F</td>
<td>Enterococcaceae</td>
<td>8.69</td>
<td>1.19</td>
</tr>
<tr>
<td>(13.3)%</td>
<td>(2.53)%</td>
<td>(2.02)%</td>
<td>(8.28)%</td>
</tr>
<tr>
<td>F</td>
<td>Streptococcaceae</td>
<td>12.68</td>
<td>12.42</td>
</tr>
<tr>
<td>(16.91)%</td>
<td>(16.31)%</td>
<td>(4.92)%</td>
<td>(16.43)%</td>
</tr>
<tr>
<td>F</td>
<td>Eubacteriaceae</td>
<td>0.06</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.25)%</td>
<td>(0.01)%</td>
<td>(0.01)%</td>
<td>(1.38)%</td>
</tr>
<tr>
<td>F</td>
<td>Lachnospiraceae</td>
<td>2.74</td>
<td>5.7</td>
</tr>
<tr>
<td>(8.59)%</td>
<td>(15.79)%</td>
<td>(5.52)%</td>
<td>(22.48)%</td>
</tr>
<tr>
<td>F</td>
<td>Clostridiaceae</td>
<td>0.29</td>
<td>0.05</td>
</tr>
<tr>
<td>(1.21)%</td>
<td>(0.15)%</td>
<td>(0.51)%</td>
<td>(0.46)%</td>
</tr>
<tr>
<td>F</td>
<td>Peptostreptococcaceae</td>
<td>0.08</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.29)%</td>
<td>(0.01)%</td>
<td>(0.00)%</td>
<td>(0.35)%</td>
</tr>
</tbody>
</table>
Table S2 (continued)

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Family</th>
<th>4 weeks</th>
<th>26 weeks</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>F</td>
<td>Ruminococcaceae</td>
<td>0.01</td>
<td>(0.05)%</td>
<td>0.00</td>
</tr>
<tr>
<td>F</td>
<td>Veillonellaceae</td>
<td>0.87</td>
<td>(2.23)%</td>
<td>0.05</td>
</tr>
<tr>
<td>F</td>
<td>Erysipelotrichaceae</td>
<td>2.51</td>
<td>(9.88)%</td>
<td>0.63</td>
</tr>
<tr>
<td>F</td>
<td>Other</td>
<td>0.07</td>
<td>(0.18)%</td>
<td>0.03</td>
</tr>
<tr>
<td>Total Proteobacteria</td>
<td>11.72</td>
<td>(19.76)%</td>
<td>13.67</td>
<td>(22)%</td>
</tr>
<tr>
<td>P</td>
<td>Enterobacteriaceae</td>
<td>11.54</td>
<td>(19.24)%</td>
<td>13.61</td>
</tr>
<tr>
<td>P</td>
<td>Other</td>
<td>0.18</td>
<td>(0.83)%</td>
<td>0.06</td>
</tr>
<tr>
<td>V</td>
<td>Verrucomicrobiacea</td>
<td>0.01</td>
<td>(0.09)%</td>
<td>0.01</td>
</tr>
<tr>
<td>Other</td>
<td>0.02</td>
<td>(0.06)%</td>
<td>0.01</td>
<td>(0.02)%</td>
</tr>
</tbody>
</table>

Relative abundances with SDs in percentages for taxa present in more than 15% of the samples and average relative abundances of greater than 0.1% at the bacterial family-level. Taxa of lower abundance are grouped as “other.” Phylum-levels are indicated as follows: A, Actinobacteria; B, Bacteroidetes; F, Firmicutes; P, Proteobacteria; V, Verrucomicrobia. Boldface text represents subtotals.

Table S3: Results of forward selection based on canonical correspondence analysis (CCA) of fecal taxonomic compositions.

<table>
<thead>
<tr>
<th>Sample covariates (variables)</th>
<th>Explains %</th>
<th>P</th>
<th>P-adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (4 weeks, 26 weeks)</td>
<td>5.2</td>
<td>0.002</td>
<td>0.01733</td>
</tr>
<tr>
<td>Ethnicity (asian, caucasian, other)</td>
<td>3.7</td>
<td>0.002</td>
<td>0.0104</td>
</tr>
<tr>
<td>Feedinggroup (control, pHF-OS, breastfed)</td>
<td>2.6</td>
<td>0.002</td>
<td>0.00867</td>
</tr>
<tr>
<td>Siblings (yes, no)</td>
<td>1.6</td>
<td>0.002</td>
<td>0.00867</td>
</tr>
<tr>
<td>Antibiotics before 26 weeks (yes, no)</td>
<td>1.2</td>
<td>0.022</td>
<td>0.08089</td>
</tr>
<tr>
<td>Birth weight (continuous data)</td>
<td>0.6</td>
<td>0.028</td>
<td>0.104</td>
</tr>
<tr>
<td>Gender (male, female)</td>
<td>0.8</td>
<td>0.206</td>
<td>0.44</td>
</tr>
<tr>
<td>Weaning before 18 weeks (yes, no)</td>
<td>0.8</td>
<td>0.226</td>
<td>0.41947</td>
</tr>
<tr>
<td>Parental allergic history (mother, father, both)</td>
<td>1</td>
<td>0.266</td>
<td>0.43225</td>
</tr>
</tbody>
</table>

The sample covariates shown were included based on the percentage explained of the total taxonomic variation (Explains %). Inclusion of sample covariates was stopped to prevent overfitting of the ordination model leading to the identification of the 9 factors (and variables) listed in the table. False discovery rate (FDR) method for P-value correction was used to identify the major covariates with an adjusted P-value (P-adj) < 0.05 (in bold) and shown alongside with the uncorrected P-value (P).
Figure S1: Box-whisker plots (Tukey method) summarizing the microbial intestinal richness (Fig. S1A) and shannon diversity (Fig. S1B) at 4 and 26 weeks of age in infants developing eczema (ECZ) compared to infants not developing eczema (NOECZ). The centre line shows the median, the + denotes the mean, and the boxes cover the 25th and 75th percentiles with the whiskers extending to the data points, which are no more than 1.5 times the length away from the box. Points outside the whiskers represent outlier samples. Comparisons were done at a sequencing depth of 1636 reads per sample. Statistics were performed by a one-way ANOVA with Bonferroni’s multiple comparison correction (not significantly different).
Figure S2: Taxa identified with MetagenomeSeq that show differential relative abundances over time (P-adj<0.1) comparing infants not developing eczema (NO ECZ, n=82 sample-pairs) with infants developing eczema (ECZ, n=48 sample-pairs). Before-after plots of relative abundances (%) are shown from 4 to 26 weeks for Enterobacter spp. (Fig. S2A), an unknown genus of Enterobacteriaceae (Fig. S2B), Parabacteroides spp. (Fig. S2C), Eubacterium spp. (Fig. S2D), and Anaerostipes spp. (Fig. S2E), respectively.
References

CHAPTER 5

A specific synbiotic-containing amino acid-based formula restores gut microbiota in non-IgE mediated cow’s milk allergic infants

Wopereis H\(^1,2\), Van Ampting MTJ\(^1\), Yavuz AC\(^1\), Slump R\(^1\), Candy DCA\(^3\), Butt AM\(^3\), Peroni DG\(^4\), Vandenplas Y\(^5\), Fox AT\(^6\), Shah N\(^7\), Roeselers G\(^1\), Harthoorn LF\(^1\), Michaelis LJ\(^8\), Knol J\(^1,2\), West CE\(^9\), on behalf of the ASSIGN study group\(^10\)

\(^1\)Danone Nutricia Research, Utrecht, the Netherlands
\(^2\)Laboratory of Microbiology, Wageningen University, Wageningen, the Netherlands
\(^3\)Royal Alexandra Children’s Hospital, Brighton, UK
\(^4\)University Hospital Verona, Verona, Italy (current affiliation: Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy)
\(^5\)KidZ Health Castle, UZ Brussel, Vrije Universiteit Brussel, Brussels, Belgium
\(^6\)Guy’s & St Thomas’ Hospitals NHS Foundation Trust, London, UK
\(^7\)Great Ormond Street Hospital, London, UK
\(^8\)Great North Children’s Hospital, Royal Victoria Infirmary, Newcastle upon Tyne, UK
\(^9\)Umeå University, Umeå, Sweden
\(^10\)Additional investigators to the ASSIGN study group: Francoise Smets, U.C.L. Saint-Luc, Brussels, Belgium; Sandra Mullier, HUDERF, Brussels, Belgium; Lee Noimark, Barts/Royal Hospital, London, United Kingdom; Antonella Muraro, University Hospital of Padova, Padova, Italy.

Submitted
Abstract

Background: Altered gut microbiota is implicated in cow’s milk allergy (CMA) and differs markedly from healthy, breastfed infants. Infants who suffer from severe CMA often rely on cow’s milk protein avoidance and, when breastfeeding is not possible, on specialised infant formulas such as amino-acid based formulas (AAF). Herein, we report the effects of an AAF including specific synbiotics on oral and gastrointestinal microbiota of infants with non-IgE mediated CMA with reference to healthy, breastfed infants.

Methods: In this prospective, randomized, double-blind controlled study, infants with suspected non-IgE mediated CMA received test or control formula. Test formula was AAF with synbiotics (prebiotic fructo-oligosaccharides and probiotic *Bifidobacterium breve* M-16V). Control formula was AAF without synbiotics. Healthy, breastfed infants were used as a separate reference group (HBR). Bacterial compositions of faecal and salivary samples were analysed by 16S rRNA-gene sequencing. Faecal analysis was complemented with the analysis of pH, short-chain fatty acids (SCFAs) and lactic acids.

Results: The trial included 35 test subjects, 36 controls, and 51 HBR. The 16S rRNA-gene sequencing revealed moderate effects of test formula on oral microbiota. In contrast, the gut microbiota was substantially affected across time comparing test with control. In both groups bacterial diversity increased over time but was characterised by a more gradual increment in test compared to control. Compositionally this reflected an enhancement of *Bifidobacterium* spp. and *Veillonella* sp. in the test group. In contrast, the control-fed infants showed increased abundance of adult-like species, mainly within the *Lachnospiraceae* family, as well as within the *Ruminococcus* and *Alistipes* genus. The effects on *Bifidobacterium* spp. and *Lachnospiraceae* spp. were previously confirmed through enumeration by fluorescent *in situ* hybridization and were shown for test to approximate the proportions observed in the HBR. Additionally, microbial activity was affected as evidenced by an increase of L-lactate, a decrease of valerate, and reduced concentrations of branched-chain SCFAs in test vs control.
Conclusions: The AAF including specific synbiotics effectively modulates the gut microbiota and its metabolic activity in non-IgE mediated CMA infants bringing it close to a healthy breastfed profile.

Introduction

The prevalence of food allergy in infancy and childhood is increasing in many countries worldwide. Cow's milk allergy (CMA) is among the most common food allergies in early life and is associated with growth retardation throughout childhood, particularly in children suffering from persistent milk allergy (1). Comorbidity is common, and many children develop other allergic conditions over time, also referred to as the allergic march (2). The microbes that colonize the mucosal tissues after birth have a pivotal role in both innate and adaptive immune development (3) and may have long-term effects both on the susceptibility and the persistence of allergic disease (4, 5).

Breastfeeding provides the infant gastrointestinal (GI) tract with a plethora of bioactive factors and has profound effects on gut microbiota composition and functions (6-8) and, as more recently reported, on oral microbiota development (9, 10). Infants who suffer from severe CMA rely on cow's milk protein avoidance and, when breastfeeding is not possible, require specialised infant formulas such as extensively hydrolysed formula (eHF) or amino acid-based formula (AAF)(11). Incorporation of prebiotics, probiotics, or their combination (synbiotics) in these formulas offers a safe, suitable and effective strategy for both the dietary management and for potentially optimizing microbiota development in both IgE- and non-IgE-mediated CMA infants (12, 13).

In a randomized controlled trial with non-IgE-mediated CMA infants (ASSIGN study), an improvement of gut microbiota was observed in infants receiving an AAF with specific synbiotics (test) compared to infants receiving the same AAF without synbiotics (control). This improvement was based on an enhancement of bifidobacteria and a decrease of the Eubacterium rectale/Clostridium coccoides (ER/CC) group; in both test levels were close to the levels observed for a separate
healthy, breastfed reference (HBR) group (14). The fluorescence in situ hybridization (FISH) method used is an effective approach to quantify specific bacterial groups, but it does not provide information on the full bacterial composition and diversity of the community. For this, application of next-generation sequencing technologies is typically needed (15, 16). We hypothesized that synbiotics would have a more comprehensive effect on the microbiota composition and activity. Therefore, we performed an in-depth characterisation of the microbial compositions of both faecal and saliva specimens collected in the ASSIGN study through 16S ribosomal RNA-gene sequencing, and in addition investigated the effects on gut physiology and bacterial metabolic activity by analysis of faecal pH, short-chain fatty acids (SCFA) and lactate.

Materials and methods

Study design

ASSIGN was a prospective, randomized, double-blind controlled study (Netherlands Trial Register NTR3979) including infants with suspected non-IgE mediated CMA and a separate non-randomized healthy, breastfed reference group (HBR). Detailed methods on how the trial was conducted, and the primary and secondary outcome measures, have been published previously (14).

In brief, subjects <13 months of age with non-IgE mediated CMA were stratified based on predominant, investigator-assessed symptoms (skin or gastrointestinal) and randomly allocated to receive test (n=35) or control formula (n=36). Study duration was 26 weeks with allocation to study product for at least 8 weeks. After 8 weeks, randomized subjects continued to use the allocated study product or switched to commercially available eHF, or other milk substitute as per clinical practice guidelines of each medical centre. Subjects in the HBR group were age-matched to week 8 of the randomized groups (n=51). Infants in the test group received an AAF (Neocate LCP; Nutricia Advanced Medical Nutrition, Liverpool, UK) including a prebiotic blend of chicory-derived neutral oligofructose and long-chain inulin (BENEO-Orafti SA, Oreye, Belgium) (9:1 ratio at a total concentration of 0.63g/100ml) and a probiotic
strain *Bifidobacterium breve* M-16V (Morinaga Milk, Tokyo, Japan) at a concentration of $1.47 \times 10^9$ colony-forming units (CFU)/100 mL formula. The control formula was a commercially available AAF (Neocate LCP; Nutricia Advanced Medical Nutrition, Liverpool, UK).

**Collection of saliva and stool samples**

Saliva samples were collected from randomized infants at baseline and 8 weeks by a healthcare professional using the SalivaBio Children’s Swab method (Salimetrics, Carlsbad, USA) at least 1 hour after feeding. Stool samples from randomized infants were collected, as reported previously (14), by parents/guardians at baseline, 8, 12 and 26 weeks. Parents/guardians of the age-matched non-randomized infants in the HBR group were asked to collect stool samples only. Saliva and stool specimens collected in the clinic were immediately frozen at $-80^\circ$C. Stool specimens collected at home were immediately frozen in home-freezers and transported with ice-packs to the clinic by parents/guardians or by courier for storage at $-80^\circ$C. Thereafter, both saliva and stools were transported on dry-ice (solid CO2) to Nutricia Research and stored at $-80^\circ$C until analysis.

**DNA extraction**

DNA extraction from saliva samples was performed with DNeasy Blood & Tissue Kits (Qiagen, Venlo, the Netherlands) according to the manufacturer’s protocol, except for an adaptation in the enzymatic lysis step and the addition of a mechanical lysis step as pre-treatment before the DNA isolation procedure. In brief, 150 µl of saliva sample was diluted up to 350 µl in PBS buffer (150 mM NaCl, 10 mM Na2HPO4, 20 mM NaH2PO4, pH 7.4) to which 50 µl of lytic enzymatic cocktail was added (50 mg/mL lysozyme, Sigma-Aldrich, St. Louis, Missouri, United States, USA and 20 µl proteinase K from Qiagen kit) and 300 mg of 0.1 mm glass beads (Biospec, Bartlesville, Oklahoma, USA). This suspension was incubated at 37°C for 30 minutes, followed by one round of bead-beating for 10 minutes at 25 Hz (Tissuelyser I, Qiagen, Venlo, the Netherlands) and followed by the QIAcube isolation procedure (Qiagen).

DNA extraction from stools samples was performed with QIAmp DNA Stool Mini Kit (Qiagen, Venlo, the Netherlands) according to the manufacturer’s protocol except
for the addition of two bead-beating steps as described before (17). Extracted DNA from stools were purified from extraction impurities using spin columns (DCC™, Zymo research, Irvine, California, USA).

**Microbiota profiling**

Faecal and salivary microbiota compositions were profiled by sequencing the hypervariable V3-V4 regions of the 16S rRNA gene. Sequencing was performed by LifeSequencing S.L. (Valencia, Spain) on an Illumina MiSeq instrument (San Diego, California, USA). The V3-V4 region was PCR-amplified with universal primers S-D-Bact-0341-b-S-17 primer (forward 5’-CCTACGGGNGGCWGCAG-3’) and S-D-Bact-0785-a-A-21 primer (reverse 5’-GACTACHVGGGTATCTAATCC-3’) (18) designed for dual indexing following the Illumina 16S Metagenomic Sequencing Library Preparation protocol (Part # 15044223 Rev. B). In brief, PCR amplification was performed in two steps: i) In a first step, the V3-V4 region was amplified with the addition of universal adaptors to the amplification products. All amplicons were purified (AMPure XP, Beckman, Danvers, MA) to remove short amplification products and quantified using the Quant-iT PicoGreen dsDNA kit (Invitrogen, Carlsbad, California, USA). ii) In the second PCR step, the amplicons from the first step were amplified by targeting the universal adapters and with the addition of sample specific indexes and sequencing adaptors. The final amplicons were purified (AMPure XP) and quantified using the Quant-iT PicoGreen ds DNA kit (Invitrogen). All samples were pooled in equal amounts and sequenced in a 300bp paired-end mode.

**Bioinformatic analysis of sequence data**

Illumina reads were trimmed (removal of primers) and quality filtered by removing all reads with a mean q-score lower than 20 with ‘cutadapt v1.4.1’ (19). Paired-end reads were merged using the program ‘PEAR v0.9.6’ (20). Merged reads with q>15 over a window of 5 bases, no ambiguous bases and a minimal length of 300 were retained and analysed with the ‘Quantitative Insights Into Microbial Ecology’ (QIIME) v1.9.0 package (21). Sequences were clustered into Operational Taxonomic Units (OTUs) based on 97% sequence identity as proxy for bacterial species using VSEARCH v2.03 with exclusion of chimeric sequences identified against the RDP
gold database (22, 23). Taxonomic assignment was performed using the RDP classifier (24) against the SILVA119 database (25). Singleton OTUs, and OTUs with eukaryotic assignments, as well as OTUs with a low relative abundance (Counts of an OTU as proportion of total reads of a sample) up to 0.005% were excluded from further downstream analysis. Representative sequences of OTUs were aligned using PyNAST (26) and used to build a phylogenetic tree with FastTree (27). Rarefaction of the OTU tables was applied to account for the differences in sequencing depths (number of reads per sample) between the samples with default settings (10 equal depths from 10 sequences/sample up to the median number of sequences/sample with 10 iterations at each sequencing depth). The tree and rarefied OTU tables were used to calculate the species diversity (α-diversity) of the samples using Faith’s phylogenetic diversity (PD) (28) and the Shannon index for diversity (29).

The sequences within an OTU of interest (i.e. identified as differentially abundant from the statistical comparisons performed) were further partitioned into homogenous nodes with high sequence identity using the MED v2.1 algorithm (30). Taxonomic assignment of the MED nodes were performed using the RDP classifier (24) against the SILVA119 database (25). The assignment of the node with the largest number of reads and highest sequence identity was subsequently used as a more accurate proxy to the taxonomy of that OTU.

**Additional faecal sample parameters**

To assess overall bacterial metabolic activity, the following faecal sample parameters were measured as described previously (31): pH, concentrations of short-chain fatty acids (SCFA) (i.e., acetate, propionate, butyrate, iso-butyrate, valerate, and iso-valerate), and D- and L-lactate.

FISH was applied to quantify the *Bifidobacterium* genus and *Eubacterium rectale/Clostridium coccoides* group (ER/CC) as described previously (16) using the 16S rRNA-targeted oligonucleotide probes S-G-Bif-0164m-a-A-18 (5′-CATCCGGYATTACCACCC-3′) (32, 33) and S-*=Erec-0482-a-A-19 (5′-GCTTCTTAGTCARGTACCG-3′) (15), respectively.
Data handling and statistical analyses

All analyses were performed on intention-to-treat data set (ITT), defined as all randomized subjects. Overall, the statistical analyses were performed comparing test with the control group per specimen analysed (saliva or faecal). The HBR group data was used as reference only and not as an intervention group. Statistical analyses were performed by using SAS® (SAS Enterprise Guide version 4.3 or higher) for Windows (SAS Institute Inc., Cary, NC) unless indicated otherwise. Results are expressed and presented as mean values and standard deviations unless stated otherwise.

16S rRNA-gene sequencing data:

The species diversity (α-diversity) indexes calculated in QIIME from the 16S rRNA-gene sequencing data were analysed at one single rarefied sequencing depth. The sequence depth which was selected for comparison was based on the maximum rarefaction depth where all or most of the samples were still included. Differences between treatment groups across time were tested using a random intercept mixed model including baseline in the outcome vector, adjustment for stratification factor (skin or gastrointestinal symptoms), treatment, time and treatment by time interaction as fixed factors. For assessing the treatment effect over time, significance of the treatment by time interaction was used.

The non-rarefied OTU tables obtained from QIIME were trimmed, removing sparse OTUs with at most 10 non-zero observations. Statistical analysis of the bacterial compositions was performed by applying a combination of multivariate analysis with Canoco 5 software (34), followed by differential abundance testing using the two-part statistics method (35). Firstly, the constrained ordination method, known as redundancy analysis (RDA), was used to test time-dependent treatment effects with adjustment for stratification factor. The Monte Carlo Permutation test (MCPT) with 1000 permutations was used to evaluate statistically significant differences (P≤0.05) of the resulting model. If found significant, the top-15 responding OTUs identified from the RDA model were subsequently tested for differential abundance at the different timepoints using the two-part statistics method (35). If the two-part statistics method
could not be applied due to a small number of non-zero observations, then only presence-absence was considered by applying the Chi-square test (if ≥5, but <10 non-zeros in both groups) or Barnard test (if <5 non-zeros in one group). The Benjamini–Hochberg false-discovery rate (FDR) was used to correct for multiple comparisons in the differential abundance tests (36) and significance was considered when FDR≤0.1 at week 8 or when FDR≤0.1 for at least two visits (i.e. 12 and 26 weeks).

Other faecal parameters (pH, FISH outcomes, SCFAs, lactic acids):

The following rule was applied to faecal parameters that were subject to limit of detection (LOD): If a value was below detection limit and the percentage of values below detection limit was at most 30%, then the value was replaced with LOD/2. For parameters with more than 30% of the values below LOD only presence-absence was considered, and P-values were based on a logistic regression model. The P-values for continuous data were based on the analysis of covariance (ANCOVA) or Van Eiteren test depending on normality of the residuals. All statistical models were corrected for baseline levels (if applicable) and stratification factor, and for statistical significance P≤0.05 was considered.

Results

Subject characteristics were well balanced between groups as reported previously (14). In total, 378 (125 saliva and 253 faecal) samples were successfully sequenced with a sequencing depth ranging from 15265 to 129780, and a median depth of 39761 sequences per sample (Table S1). Principal component analysis (PCA), which is an unconstrained ordination method (34), was used to explore the taxonomic compositions of saliva and faecal samples. A clear clustering by sample origin was observed (Figure 1A), which confirms that community composition is primarily determined by body habitat (37). A summary of the most dominant taxa identified at the bacterial family level showed that saliva compared to faecal samples were typically characterized by increased relative abundance of *Streptococcaceae* (53.3±17.8%), *Microcrococcaceae* (9.7±6.4%) and *Actinomycetaceae* (5.2±5.8%).


Faecal samples were typically characterized by *Bacteroidaceae* (20.9±18.7%), *Lachnospiraceae* (15.4±12.8%), *Enterobacteriaceae* (14.2±14.2%), *Bifidobacteriaceae* (7.6±9.2%), *Ruminococcaceae* (6.7±8.2%) and *Verrucomicrobiaceae* (5.3±10.7%) (Figure 1B).

**Figure 1:** Principal Component Analysis (PCA) of faecal and salivary microbiota composition (A) and summary of major bacterial families identified (B). The PCA sample scatterplot is displayed on the first two axes summarizing most of the species variation, which is based on the OTU count data for each sample. The distance between the sample symbols (rounds for saliva and squares for faecal) approximates the dissimilarity of their species composition as measured by their Euclidean distance. Mean relative abundances (± SD) are summarized at the family level (“_f_”) for taxa > 1% and summarized in the heat map (Red-Yellow-Green color scheme indicating high to low relative abundance). Abbreviations used for bacterial phylum levels: Acti = Actinobacteria, Bact = Bacteroidetes, Firm = Firmicutes, Prot = Proteobacteria, Verr = Verrucomicrobia.

### Bacterial diversity

The species diversity indexes (PD and Shannon) were analysed at a rarefaction depth of 16114 sequences per sample, which omitted one saliva sample (control, week 8) from comparison. Control and test group did not differ in salivary species...
diversity based on PD (Figure 2A) or Shannon index (Figure 2C). A treatment effect over time was observed for faecal species diversity, which was characterized by a more gradual increment (from baseline until 26w) in test compared to control for both PD (Figure 2B, estimated difference per week = -0.022, P=0.069) and Shannon index (Figure 2D, estimated difference per week = -0.026, P=0.005). The estimated average difference between test vs control was significantly different at week 12 (PD=-0.349, P=0.031 and Shannon=-0.236, P=0.049) and week 26 (PD=-0.653, P=0.012 and Shannon=-0.596, P=0.002). The HBR group showed the lowest average diversity (PD=4.37±1.14 and Shannon=3.63±0.80) compared to both test (PD=4.89±1.05 and Shannon=3.75±0.67) and control (PD=5.17±0.88 and Shannon=4.01±0.71) at week 8.

**Figure 2:** Least square (LS) means with 95%CI of phylogenetic diversity in saliva (A) and faecal samples (B), and Shannon diversity in saliva (C) and faecal samples (D) for treatment by time. The HBR reference values (age-matched to week 8) are plotted as well. P-values are based on a random intercept mixed model with Week 8/12/26 values as outcome, stratification factor and baseline values as covariate and treatment as fixed effect: *P≤0.05 and **P≤0.01.
Time-dependent treatment effects on oral microbiota

Redundancy analyses (RDA) were carried out to test the effect of treatment (test/control) across time on the salivary community composition. We fitted both an RDA with and without correction for timepoint (baseline and week 8) and compared the results of the MCPT on the first axis of the model. The P-value for the RDA with correction for timepoint (0.3816) was larger than our pre-set threshold of 0.05, so we used the simpler model (with P=0.003) as a basis for interpreting the time-dependent treatment effects.

Figure 3: Redundancy Analysis (RDA) plot of salivary microbiota composition based on treatment by time (week 0/8) interaction with adjustment for stratification factor (A). The top 15 OTUs are plotted based on best fit with the first two RDA axes. In bold the OTUs that were significantly different at week 8 between test vs control as confirmed with the two-part statistics. The two-part statistics combines a test to compare the proportion of zeros (plotted as bars) and one to compare the median of the non-zero values (plotted as points) and are displayed for the two taxa (B and C) identified as differentially abundant. False Discovery Rate (FDR) was used to correct the raw P-values for multiple testing with significance at 0.1. If <5 non-zeros in one group) to compare the proportion of zeros. Taxa are summarized at the OTU level with unique (but arbitrary) numbers as identifiers, genus level (“_g_”) and phylum level: Acti = Actinobacteria, Bact = Bacteroidetes, Firm = Firmicutes, Fus = Fusobacteria, Prot = Proteobacteria, Sacc = Saccharibacteria.
The top 15 OTUs with the best fit on the first two axes (explaining most of the variation) were plotted in the RDA (Figure 3A) and further evaluated for differential abundances between test and control using the two-part statistics method (35). No differences were observed at baseline (based on FDR≤0.1), but two OTUs out of this top 15 were found differentially abundant between test and control at week 8. This included a decreased relative abundance of *Peptostreptococcus* sp. (Figure 3B, FDR=0.0525) and an increased presence of *Parabacteroides* sp. (Figure 3C, FDR=0.0525).

**Time-dependent treatment effects on gut microbiota**

In order to assess the time-dependent treatment effects for the faecal community composition we used the Principal Response Curves (PRC) method (38). The PRC is based on the RDA method, in which the principal component is plotted against time (baseline, week 8, 12 and 26) to enable the assessment and visualization of time-dependent treatment effects. The MCPT applied to test the significance of the resulting PRC model was significant for the first axis (P=0.001). The top 15 OTUs with the best fit on the first axis were plotted (Figure 4A) and further evaluated with the two-part statistics method (35). No differences were observed at baseline, but a total of 13 OTUs out of the top 15 were confirmed to be differentially abundant between test and control at week 8 or at 2 or more timepoints. This included increased relative abundances in test vs control of 6 OTUs, of which 5 were assigned to *Bifidobacterium* and 1 was assigned to the *Veillonella* genus. The other 7 OTUs showed decreased relative abundances, of which 5 were assigned to 3 genera within the *Lachnospiraceae* family (*Tyzzerella, Blautia* and *Lachnoclostridium*) and 2 were assigned to the genera *Ruminococcus* and *Alistipes*, respectively.

**FISH quantification of faecal bacterial groups**

The treatment effects on gut microbiota, as revealed by 16S rRNA-gene sequencing, were mostly associated with a relative increase of several species of the genus *Bifidobacterium* and a decrease of several species of the family *Lachnospiraceae*. FISH enumeration of these two bacterial groups was used to verify the absolute differences in abundance between treatments, of which results have
been reported before (14). In summary, FISH analysis confirmed a significant enrichment of bifidobacteria in test vs control across time (Figure 4B). Moreover, the proportions for test (36.0±22.4%) as compared to that of the control group at week 8 (14.5±16.4%) were close to the levels observed for the HBR group (48.1±26.5%). The FISH probe used to quantify the ER/CC group targets the majority of *Lachnospiraceae* spp. including the differentially abundant OTUs associated with the genera *Tyzzerella*, *Blautia* and *Lachnoclostridium* as identified with 16S rRNA-gene sequencing.

![Figure 4](image)

**Figure 4:** Principal Response Curves (PRC) of faecal microbiota composition testing the treatment by time interaction with adjustment for stratification factor and time (A). The treatments are presented as a single response curves across time (on the horizontal axis) with Control as reference with zero PRC values (on the vertical axis) and so its curve lays over the horizontal axis. The top 15 OTUs are plotted on the separate vertical (one-dimensional) plot based on best fit with the first PRC axis. In bold the OTUs that were significantly increased (>0) or decreased (<0) in Test vs Control at week 8 or at two or more timepoints (week 8/12/26) as confirmed by the two-part statistics (FDR≤0.1). Percentages (means with 95%CI) of bifidobacteria (B), and ER/CC group (C) quantified by FISH at week 0/8/12/26. The HBR reference values (age-matched to week 8) are plotted as well. P-values are based on ANCOVA comparing Test vs Control with Week 8/12/26 values as outcome, stratification factor and baseline values as covariate and treatment as fixed effect: **P≤0.01; ***P≤0.001. Taxa names are given at the OTU level with unique (but arbitrary) numbers as identifiers, genus level (“_g_”), family level (“_f_”): Bact = Bacteroidaceae, Bifi = Bifidobacteriaceae, Lach = Lachnospiraceae, Rumi = Ruminococcaceae, Veil = Veillonellaceae; and phylum level: Acti = Actinobacteria, Bact = Bacteroidetes, Firm = Firmicutes.
A decreased abundance of the ER/CC group in test vs control across time confirm these findings (Figure 4C). Additionally, the levels for test (11.8±10.9%) as compared to that of the control group at week 8 (25.2±16.9%) were close to the levels observed for the HBR group (10.4±10.6%).

Faecal pH, SCFA and lactate

**Figure 5:** Mean with 95%CI of stool pH (A) for treatment by time. The box-plots summarize the amounts (in mmol/kg) of acetate (B), propionate (C), butyrate (D), iso-valerate (E), iso-butyrate (F) for treatment by time, respectively. Percentage of faecal samples (plotted as bars) with detectable levels of valerate (G), L-lactate (H), and D-lactate (I) for treatment by time, respectively. The HBR reference values (age-matched to week 8) are plotted as well. P-values for stool pH and acetate are based on ANCOVA comparing Test vs Control with Week 8/12/26 values as outcome, stratification factor and baseline values as covariate and treatment as fixed effect. P-values for the variables summarized in Figure C-E are based on Van Elteren test comparing test vs control with respect to change from baseline at Week 8/12/26, taken the stratification factor into account. P-values for the variables summarized in Figure G-I are obtained from a logistic regression model comparing test vs control at Week 8/12/26 with adjustment for baseline values. *P≤0.05.
To assess whether the observed changes in gut microbiota composition also led to changes in gut physiology and microbial metabolites produced, the faecal pH and levels of SCFA and lactate were determined. No statistically significant differences were observed for faecal pH, acetate, propionate, butyrate, iso-valerate (Figure 5A-E) and D-lactate (Figure 5I) at the different timepoints. L-lactate was detected in a greater number of samples in test vs control at week 26 (38 vs 4%, P=0.020) (Figure 5H). In contrast, valerate was detected in a smaller number of samples in test vs control at week 8 (44 vs 12%, P=0.036) and week 26 (67 vs 29%, P=0.021) (Figure 5G). Moreover, the concentration of iso-butyrate was lower in test vs control at 26 weeks (P=0.050) (Figure 5F).

**Correlations of faecal microbiota composition and metabolic activity across time**

A redundancy analysis was used to summarize the faecal microbiota composition over time as explained by treatment (Test or Control) and the HBR group (Figure 6A). The RDA recapitulates the results of the PRC analysis, but in addition confirmed the proximity in community composition of the test group at week 8 with the HBR. The additional faecal parameters measured (FISH, pH, SCFAs and lactic acids) were supplemented to this RDA in a separate biplot (Figure 6B). An inverse correlation was observed for the FISH quantified levels of *Bifidobacterium* spp. with the ER/CC group, which reflects the major differences observed for test (and HBR) with the control group. Moreover, the increase of *Bifidobacterium* spp. in test was positively correlated with increased levels of L-lactate. In contrast, the more abundant levels of the ER/CC group across time in control was associated with increased levels of butyrate, valerate, iso-butyrate and iso-valerate. In test, the ER/CC group gradually increased from 12 to 26 weeks, which was associated (similarly as for control) with an increment of butyrate, valerate, iso-butyrate and iso-valerate at 26 weeks.
Discussion

We previously reported the specific enhancement of bifidobacteria and decrement of the ER/CC group in the faeces of infants receiving the AAF with synbiotics consisting of a prebiotic blend of oligofructose and long-chain inulin and the probiotic strain *Bifidobacterium breve* M-16V (14). In this study, we applied a 16S rRNA-gene sequencing approach on both faecal and saliva specimens to elucidate more specifically which taxa responded to the intervention within the respective bacterial communities and what the effect was on their diversity and functionality.

We demonstrated that the effect of the synbiotic-containing AAF on infant microbiota was most pronounced for the gastro-intestinal tract and only minimally affected the oral microbiota. The AAF including synbiotics compared to the AAF...
without synbiotics showed a more gradual increment over time of bacterial diversity, which is also typically observed in longitudinal studies investigating early life gut microbiota development of breastfed infants as compared to formula-fed infants (6, 7, 39, 40). These studies showed that the lower diversity of gut microbiota in breastfed infants is not only observed during the exclusive human milk-feeding period, but also during the complementary feeding-period until full transition to family foods, which reflects the sustained effects of human milk oligosaccharides on the bifidobacterial species that effectively thrive on these compounds (6, 7). The AAF including synbiotics was found to enhance the bifidobacterial community, as several bifidobacterial species had increased, which was also reflected by an increase of the fermentation end-product L-lactate in the faeces of these infants. Interestingly, the concordant increase observed in this study for *Veillonella* sp. is most likely explained by the ability of this species to utilize and convert lactate into propionate (41). In contrast, the infants receiving the control formula showed an early adoption of adult-like bacterial taxa belonging to the ER/CC group (resembling *Lachnospiraceae* spp.), namely *Tyzzerella*, *Blautia* and *Lachnoclostridium* spp., as well as species of *Ruminococcus* and *Alistipes*. This increase of adult-like taxa was associated with an increase of valerate and the branched-chain SCFA iso-butyrate, which are fermentation products that result from the degradation of proteins and amino acids (42, 43). Overall, these results indicate that the synbiotic-supplemented AAF induced a saccharolytic fermentation profile, while infants receiving the AAF without synbiotics showed a more proteolytic fermentation activity, which is generally associated with metabolite profiles that may be less beneficial for colonic health (44, 45).

To date, several case-control studies have specifically investigated the gastrointestinal microbiota of infants and children with confirmed CMA compared to age-matched healthy controls (46-50). All of them reported altered gut microbiota in infants and children with CMA, although with mixed findings. However, the common characteristics that were identified in these studies included lower levels of bifidobacteria (47-50) and increased levels of members of the heterogenous ER/CC group (46-48). In analogy with our study, the case-control study of Thompson-Chagoyan et al. (46) in addition observed increased fecal butyrate and
branched-chain SCFA (iso-butyrate, iso-valerate) concentrations in CMA infants compared to healthy infants. Interestingly, our study demonstrated that 8-weeks use of the synbiotic-supplemented AAF approximated the composition and activity of the gut microbiota of the age-matched healthy, breastfed reference group.

Our study has several limitations as addressed before (14), which includes the challenges in making and confirming a specific and accurate diagnosis of non-IgE mediated allergy. The chance of including infants with other (food) allergy presentations were mitigated by applying a robust diagnostic work-up (14). For a number of subjects, no specimens were available due to insufficient material or not completing the study until 26 weeks (Table S1), which limited the number of evaluable samples at week 12 and 26. This limitation was however similar in test and control groups and would, therefore, not have affected the observed differences between groups. Moreover, the identified microbial signatures showed very consistent patterns across time and were, regarding the relative abundances of bifidobacteria and the ER/CC group, independently confirmed by 16S rRNA-gene sequencing and FISH. Although, we specifically studied subjects with non-IgE-mediated CMA, Burks et al. (13) showed that an AAF, including ingredients from the current synbiotic blend, was safe in patients with IgE and non-IgE-mediated CMA, and affected the microbial signatures in an equal manner. Importantly, this study was primarily designed to investigate the effects of the synbiotic-containing AAF on gut microbiota and the suitability for the use in dietary management of CMA. As reported previously (13, 14), the AAF including synbiotics showed reduction of allergic symptoms as seen for the control AAF, and in addition showed potential beneficial systemic effects based on the adverse events, which reported fewer subjects in test with infections and need for anti-infective medication, including antibiotics (13, 14). Currently investigations are ongoing to assess whether the AAF including synbiotics influences cow’s milk tolerance acquisition in a clinical trial conducted in infants with confirmed IgE-mediated CMA (Netherlands Trial Register NTR3725).

Based on the data showing pronounced effects on gut microbiota composition, diversity and metabolic activity, we conclude that the AAF including the specific
synbiotics offers an effective nutritional strategy to modulate the gut microbiota of infants with suspected non-IgE-mediated CMA closer to a healthy breastfed profile.

**Acknowledgement**

The authors would like to thank all infants and their caregivers for their participation in the ASSIGN study. The authors also would like to thank: The Analytical Sciences team and Heleen de Weerd of the Gut Biology and Microbiology team of Nutricia Research, The Netherlands for the support in metabolite analysis and bioinformatic analysis, respectively. We also thank the Clinical Study and Data sciences teams of Nutricia Research, The Netherlands, for suggestions and critical review of the manuscript.
Supporting Information

Table S1: Summary of saliva and faecal samples of subjects with CMA (ITT) and the healthy reference group (age-matched to week 8). Sequence depth, after filtering for low quality reads, is given as mean (±SD).

<table>
<thead>
<tr>
<th></th>
<th>SALIVA SAMPLES</th>
<th></th>
<th>FAECAL SAMPLES</th>
<th></th>
<th>Grand total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test (N=35)</td>
<td>Control (N=36)</td>
<td>Test (N=35)</td>
<td>Control (N=36)</td>
<td>Healthy subjects (N=51)</td>
</tr>
<tr>
<td>week 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n subjects</td>
<td>35 (100%)</td>
<td>36 (100%)</td>
<td>29 (83%)</td>
<td>33 (92%)</td>
<td></td>
</tr>
<tr>
<td>mean (SD)</td>
<td>30258 (14753)</td>
<td>28802 (6845)</td>
<td>47476 (14165)</td>
<td>47269 (22263)</td>
<td></td>
</tr>
<tr>
<td>week 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n subjects</td>
<td>24 (69%)</td>
<td>30 (83%)</td>
<td>24 (69%)</td>
<td>31 (86%)</td>
<td>48 (94%)</td>
</tr>
<tr>
<td>mean (SD)</td>
<td>27387 (6555)</td>
<td>29724 (15773)</td>
<td>47037 (13765)</td>
<td>47909 (16532)</td>
<td>51871 (19134)</td>
</tr>
<tr>
<td>week 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n subjects</td>
<td></td>
<td></td>
<td>18 (51%)</td>
<td>26 (72%)</td>
<td></td>
</tr>
<tr>
<td>mean (SD)</td>
<td></td>
<td></td>
<td>52921 (13491)</td>
<td>51391 (19953)</td>
<td></td>
</tr>
<tr>
<td>week 26</td>
<td></td>
<td></td>
<td>21 (60%)</td>
<td>23 (64%)</td>
<td></td>
</tr>
<tr>
<td>n subjects</td>
<td></td>
<td></td>
<td>51706 (22464)</td>
<td>55888 (17966)</td>
<td></td>
</tr>
<tr>
<td>Grand total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n samples</td>
<td>125</td>
<td></td>
<td>253</td>
<td></td>
<td>378</td>
</tr>
<tr>
<td>mean (SD)</td>
<td>29160 (11837)</td>
<td></td>
<td>50200 (18201)</td>
<td></td>
<td>43242 (19124)</td>
</tr>
</tbody>
</table>
References

1. Robbins KA, Wood RA, Keet C. Persistent Cow’s Milk Allergy is Associated with Decreased Growth: A Longitudinal Study. Journal of Allergy and Clinical Immunology. 2018;141(2):AB150.


CHAPTER 6

Gut microbiota from infant with cow’s milk allergy leads to atopic orientation in a murine model

Mauras A1*, Wopereis H2,3*, Yeop I4, Esber N1, Delannoy J1, Labellie C1, Reygner J1, Kapel N1, Slump R2, Van Eijndthoven T2, Rutten L2, Knol J2,3, Garssen J2,5, Harthoorn LF2, Butel MJ1, Bajaj-Elliott M4, Hartog A2,5, Waligora-Dupriet AJ1

*Joint first

1EA 4065 Ecosystème intestinal, Probiotiques, Antibiotiques, Université Paris Descartes, Faculté de Pharmacie, Paris, France,
2Danone Nutricia Research, Utrecht, The Netherlands
3Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands
4Great Ormond Street Institute of Child Health, University College of London, London, UK,
5Department of Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

Accepted for publication by The European Journal of Allergy and Clinical Immunology (Allergy) as a Letter to the Editor
Abstract

**Background:** Cow's milk allergy (CMA) is a significant health burden affecting up to 5% of children. Emerging data raises the hypothesis that an altered gut microbiota in infants with CMA contributes to disease onset and persistence. Herein, we report the effects of fecal microbiota transplantation (FMT) of a healthy control (HC) and CMA infant into germ-free mice using a murine model of sensitization and challenge.

**Methods:** Infants (5-16 months of age) were recruited and their stools extensively characterized prior to selection of a representative for FMT in germ-free mice. Using an established murine model of CMA; we investigated the establishment of the gut microbiota and its effect on clinical features, sensitization markers, humoral and T-cell immunity in response to antigenic challenge.

**Results:** Decreased levels of *Bifidobacterium* and increased levels of *Lachnospiraceae* spp. characterized the CMA microbiota, which was maintained upon transplantation. In the murine model, clinical scores (scratching, puffiness, loss of mobility) were significantly higher in CMA sensitized group (S) versus non-sensitized group (NS) and HC-S group. Interestingly, minimal differences in mMCP-1 and allergen-specific sensitization markers were observed between the groups suggesting a non-IgE mediated pathway in CMA-S group was involved. Additionally, CMA-S and CMA-NS groups showed increased total IgE-levels and IgG1/IgG2a ratio when compared to the HC-S and HC-NS groups, respectively. Further, the CMA microbiota induced significant increases in colonic *gata3* and *foxp3* mRNA expression.

**Conclusion:** The presence of a CMA-associated infant gut microbiota promoted clinical and immune parameters of an atopic orientation in a murine model of CMA.
Introduction

Cow’s milk allergy (CMA) is a significant health burden estimated to affect up to 5% of the population during the first years of life (1). CMA can be classified as IgE-mediated (immediate), non-IgE-mediated (delayed) or a less prevalent dual form (2, 3). Although most infants and children outgrow CMA before the age of 5 years, an increasing number suffer from persistent symptoms or develop other related conditions, known as the allergic march (4).

Clinical and epidemiological studies suggest an association between abnormal development of the gut microbiota in early life to clinical manifestations of allergy (5, 6). At present however, there is no consensus on a clear signature of a CMA-microbiota and, whether the observed microbial alterations are a cause or consequence of allergy (6). In case-control studies of infants with CMA, a decrease in bifidobacteria and an increase in *Eubacterium rectale-Clostridium coccoides* (ER/CC) group was reported (7-10). Moreover, a fecal transplant of healthy infant microbiota dominated by *Bifidobacterium* and *Bacteroides* spp. showed a protective effect on sensitization and food allergy in a murine model of CMA (11, 12). However, none of the studies investigated the impact of a CMA-associated infant gut microbiota
compared to a healthy infant gut microbiota in triggering sensitization and development of allergic responses.

In this study, we recruited infants who were either healthy or with paediatrician-diagnosed CMA. After detailed analyses of their fecal microbiota, samples from a representative CMA-donor and HC-donor were transferred into germ-free mice to assess the role of microbiota in the development of CMA.

Methods

Subjects, allergic workup

Allergic infants under the care of Great Ormond Street Hospital, London, UK, were recruited alongside healthy infants from the community (REC No 14/LO/0364). The CMA was diagnosed based on clinical presentation (immediate reaction with urticaria) and this was confirmed subsequently following oral challenge. Only term, healthy infants who were fully or partially breastfed, with no family history of atopy, no exposure to antibiotics (in the antenatal or postnatal period), nor any other medications were recruited for comparison with the allergic cohort. Stool from 1 nappy were separated into three containers (1 container with 5 ml 10% glycerol in 0.9% NaCl), then stored and transported at 4°C until storage at -80°C within 24 hours.

Colonization of germ-free mice

The protocol was approved by the Regional Council of Ethics for animal experimentation (Île-de-France, Paris Descartes, CEEA34.AJWD.062.12).

Germ-free C3H/HeN mice (Anaxem, INRA, Jouy-en-Josas, France) were housed in sterile isolators in the animal care facilities of CRP2-UMS 3612 CNRS-US25 INSERM-IRD at the Faculté de Pharmacie de Paris, Université Paris Descartes, France. Mice were allowed ad libitum intake of autoclaved water and pellet AIN93G based chow sterilized by γ-irradiation at 50 kGy (ssnifspezialdiäten, Soest, Germany). They received the two selected fecal microorganisms by oral gavage at weaning age (21±3 days of life) producing 2 groups, i.e. healthy infant microbiota-associated mice (HC group) and CMA infant microbiota-associated mice (CMA group). This transfer
was performed on 3 consecutive days. Stool was cultured in Tryptone-Glucose-Yeast-Hemin medium at 37°C for 24h under aerobic conditions or 48h under anaerobic conditions (MACS anaerobic cabinet; AES-Chemunex, Bruz, France; N₂/H₂/CO₂; 80:10:10). On day 3 (D3), 100µl of the aerobic culture was administered to mice; on D4, 100µl mix of anaerobic/aerobic culture (2:1, v/v) was administered and on D5, the mice received 100µl of stool diluted in PBS at 1:10 (Figure 1). In total, 2 independent mice experiments were performed.

**Oral sensitization and immune challenge**

Each FMT group was divided into 2 subgroups of 6 to 12 mice each: the non-sensitized group included 5 to 6 mice and sensitized groups included 12 to 15 mice. The first subgroup (S group) was sensitized with whey proteins (WP, lacprodan 80, Arla, Lyon, France; 15mg per mouse) and adjuvant cholera toxin (CT) (List Biological Laboratories, Campbell, CA; 10µg mouse⁻¹) in PBS. The second subgroup received only CT in PBS as a control (non-sensitized group, NS). Sensitizations were performed by oral gavage once a week for 5 weeks (Figure 1). One week after the last sensitization (D50), all mice received an oral challenge with 60mg of β-lactoglobulin (BLG, Sigma Aldrich, Saint-Quentin-Fallavier, France).

**Evaluation of allergic response**

Clinical scoring was performed on D50, 30min after the BLG challenge, by two investigators blinded to the sensitization protocol and the mouse groups for 15 min. As previously described (11), allergic symptoms were evaluated based on three criteria: scratching behavior, loss of mobility, and puffiness (details in Supporting Information). Rectal temperature was taken before the challenge and after the observation period.

**Sampling**

Fecal pellets were collected after the 1ˢᵗ, 3ʳᵈ and 5ᵗʰ sensitization and the day after (Figure 1) and scored as follows: normal=0, wet=1, glairy and very soft=2, diarrhea=3, no feces or anal inflammation = 1. The fecal score was defined as the sum of the two scores (consistency and inflammation) and therefore, ranged from 0 to 4.
On day 50, blood was recovered in K3-EDTA tubes, centrifuged immediately, and plasma was stored at -80°C for immunoglobulin and mast cell protease-1 (mMCP-1) measurements. Spleens and mesenteric lymph nodes (MLN) were used for lymphocyte cultures followed by cytokine dosages. Two-cm of duodenum, jejunum, ileum and colon were collected and stored in RNA-later (Sigma-Aldrich, France) for relative expression gene measurements. Cecum contents were stored in microtubes. Samples were stored at -80°C until analysis. All methods are further detailed in Supporting Information.

**Figure 1: Experimental design.** Three-week old mice were orally inoculated with selected healthy infant microbiota (HC group) or CMA infant microbiota (CMA group) cultured aerobic and anaerobic at D3, D4 and with fresh diluted stool at D5. From days 15 to 43, mice were orally sensitized with whey protein and cholera toxin once a week (S group). Control mice were treated with cholera toxin alone (C group). At D15, 16, D29, 30, D43 and 44, feces were collected and scored (4h and 24H after the sensitization). All mice were orally challenged with BLG one week after the last sensitization, followed by an allergic response score and sacrifice.

**Gut microbiota and other parameters in infant and mice samples**

The gut microbiota composition and diversity were determined by 16S rRNA-gene sequencing and, in the case of infant stools, complemented by fecal pH-analysis plus the quantification of *Bifidobacterium* spp. and *Eubacterium rectale/Clostridium coccoides* group (ER/CC) by fluorescent in situ hybridization (FISH) (13), as well as the analysis of the immune markers eosinophil-derived neurotoxin (EDN) (14), calprotectin (15), and secretory IgA (16) as previously described. Bacterial metabolic
activity in both infant and mice samples was assessed through the analysis of short-chain fatty acids (SCFA) (i.e., acetate, propionate, butyrate, iso-butyrate, valerate, and iso-valerate) and lactic acids (D- and L-lactate) as described previously (17).

Detailed information on the extraction of fecal and cecal DNA, and the 16S rRNA-gene sequencing analysis are available in the Supporting Information.

Statistical analysis

For statistical analysis and visualization of microbiota community data, Graphpad Prism software version 7.00 for Windows (La Jolla California USA) was adopted for comparisons of species diversity and Canoco 5 software was used for Principal Component Analysis (PCA) (18) with Aitchison log-ratio transformation of bacterial genus compositional data (19). Differential abundance analysis was performed at the bacterial family and genus level using the R-package ANCOM (20) and the iTOL version 4 for visualization of the discriminant taxa (21). Corrections for multiple comparisons were controlled using the Benjamini-Hochberg false discovery rate (FDR) method with significance below 0.05 (22). Other fecal and cecal parameters, mice clinical response data and immune data were analyzed with Graphpad Prism software applying Mann-Whitney test for two-group comparisons and Kruskal-Wallis to compare more groups, with Dunn's correction for multiple testing. A P value of less than 0.05 was considered significant.

Results

Infant characteristics

Infants were recruited over a period of 4 months. All CMA infants (n=5, 7-16 months of age, all girls) had a family history of atopy and had previous exposure to medications with 2 infants having regular medications during sample collection (Table S1 and S2). Two CMA infants were born by vaginal delivery and all were initially breastfed. All CMA-infants were on amino acid-based formula (4 on Neocate and 1 on Nutramigen AA) and 3 infants were diagnosed with multiple food allergies. Six healthy infants (5.5-9 months of age, 3 girls) were recruited with 4 infants born by vaginal
delivery. All healthy infants were breastfed with 2 infants also receiving standard infant formula (Aptamil). All infants were weaned onto solids when fecal samples were collected.

**Gut microbiota from CMA infants showed increased diversity and decreased Bifidobacterium /Lachnospiraceae ratio**

All samples were successfully sequenced with a median sequencing depth of 61282(41953-76653) reads per sample (Table S3), except one sample with insufficient fecal material available for DNA extraction (nr. 10, Table S1). Microbial richness (P<0.05, Figure 2A) and Shannon diversity (P=0.067, Figure 2B) were both higher in CMA infants compared to HC. *Bifidobacteriaceae* spp. and its genus *Bifidobacterium* were less abundant in CMA-infants than in HC-infants, while *Lachnospiraceae* spp. and one of its genera, i.e. *Eisenbergiella* (FDR<0.05, Figure 2C) were observed to be more abundant in CMA-infants. FISH quantification of *Bifidobacterium* spp. and the ER/CC group (which includes most of the *Lachnospiraceae* family) confirmed the patterns observed by 16S rRNA-gene sequencing (P<0.05, Figure S1A and S1B). Significant increases of bacterial SCFAs - butyrate, iso-valerate and iso-butyrate - were observed in CMA-infants compared to healthy controls (P<0.05, Figure S1F-H). No significant differences were observed in stool pH, levels of acetate, D-lactate, L-lactate, slgA, calprotectin and EDN (Figure S1B, S1D, S1I-M, respectively).

Principal component analysis (PCA) revealed distinct microbial compositions for CMA versus HC (Figure 2D). The other fecal parameters analyzed were incorporated into this PCA and projected as biplot (Figure 2D), which clearly visualizes the aberrances observed in microbiota composition (driven by differential abundances of *Bifidobacterium* and *Lachnospiraceae* spp.) and associated bacterial metabolites (SCFAs and lactic acid) and immune parameters (SIgA, calprotectin and EDN).
Figure 2: Donor screening and selection for FMT. Box-whisker plots of richness (A) and shannon diversity (B) of fecal samples collected from infants with cow’s milk allergy (CMA) or without (HC). Samples selected for FMT indicated as grey symbols. Cladogram with discriminant taxa identified when comparing CMA-infants with HC using the ANCOM method (C). The cladogram visualizes the phylogenetic relatedness at the different taxonomic levels with from the outside to the inside: genus, family, order, class and phylum level, respectively. Discriminant taxa identified at the genus level (darkest colour) and/or at the family level (lighter colour) are highlighted. Green colours indicate increased relative abundances in HC, and red colours indicate increased relative abundances in CMA (FDR at 0.05). PCA of gut microbiota compositions of CMA and HC-infants at the taxonomic genus level (D) supplemented all fecal parameters measured. Donor samples selected for fecal transfer to germ-free mice are shown in bold (F03 and F04). AA=acetic acid, BA=butyric acid, PA=propionic acid, VA=valeric acid, iso-BA=iso-butyric acid, iso-VA=iso-valeric acid, DLA=D-lactic acid, LLA=L-lactic acid, ER/CC=Eubacterium rectale-Clostridium coccoides group, EDN=Eosinophil-derived neurotoxin.
Selection for FMT

We selected representative infant fecal microbiota for transplantation to obtain healthy infant microbiota-associated mice (HC group) and cow’s milk allergic infant microbiota-associated mice (CMA group). The selected infants were matched for age, gender and delivery-mode. Based on these criteria and the results of the fecal workup, we selected a healthy infant of 9 months old (Infant 3, Table S1) and a 10-month-old CMA-infant (Infant 4, Table S1), both female and born by caesarean section.

Sustained microbial signatures of CMA and HC upon establishment in mice

To determine if the microbiota patterns were replicated in the recipient gnotobiotic mice, we sequenced the fecal microbiota at the start of the 5-week sensitization period (D15), and the cecum microbiota at the end of the experiment (D50).

**Figure 3: Microbiota diversity in murine model of CMA.** Box-whisker plots of richness and shannon diversity analysed in fecal pellets collected at D15 (A and B) and in cecum content collected at D50 (C and D). Comparisons were performed by Kruskal-Wallis test with Dunn's correction for multiple testing. **P≤0.01; ***P≤0.001, ****P≤0.0001.
The increased richness and diversity in CMA vs HC was preserved in the recipient mice at D15 (Figure 3A and 3B) and sustained to the end of the experiment (Figure 3C and 3D). No differences were observed between sensitized and non-sensitized mice receiving the same fecal transfer.

Figure 4: Establishment of donor microbiota in mouse recipients. PCA of gut microbiota compositions of selected donors with mice fecal pellets collected at D15 (A), and with mice cecal samples at D50 (B) at the taxonomic genus level. Cladogram with discriminant taxa identified in mice at D15 (C) and at D50 (D) using the ANCOM method. Taxa that were differentially abundant when comparing HC-mice with CMA-mice are highlighted at the genus (darkest colour) and family level (lighter colour). Green colours indicate an increase of relative abundances in HC, and red colours indicate increased relative abundances in CMA (FDR at 0.05).
PCA showed distinct profiles for the CMA-transfer compared to the HC-transfer at D15 and D50, which closely resembled the two donor compositions (Figure 4A and 4B). Moreover, the differential abundance of families and genera between the two transfers included the taxa associated with the recruited infants and replicated the decreased levels of *Bifidobacteriaceae* spp. (and *Bifidobacterium*) and increased levels of *Lachnospiraceae* spp. (and *Eisenbergiella*) in CMA vs HC at both time points (FDR<0.05, Figure 4C and 4D). Additionally, differential abundances observed at both time points were consistent for other genera within the family of *Lachnospiraceae*, most of which were higher in CMA (*Eubacterium oxidoreducens* group, *Robinsoniella* and an unknown genus), with the exception of *Anaerostipes* and *Lachnoclostridium 5*, which were higher in HC. Other bacterial families that were consistently increased in CMA were *Porphyromonadaceae* (associated with *Dysgonomonas* spp.) and *Peptostreptococcaceae* (associated with increases of *Intestinibacter* and *Peptoclostridium*), while an unknown family of Clostridiales was consistently higher in HC-mice (FDR<0.05, Figure 4C and 4D).

Cecal levels of SCFAs and lactic acids revealed significantly increased concentrations of acetate, butyrate and iso-valerate in CMA-S compared to HC-S (Figure S2A, S2C and S2D), as well as increased prevalence of valerate and iso-butyrate, with the latter also more prevalent in CMA-NS vs HC-NS (Figure S2G and S2H). No differences were observed for propionate, D-lactate and L-lactate (Figure S2B, S2E and S2F).

**CMA microbiota, but not HC microbiota, was associated with diarrhea-related symptoms following the oral administration of allergen**

The fecal scoring, recording consistency and anal inflammation, showed different responses between the two fecal transfers during the five-week period of oral sensitization. After the first sensitization, with presentation of antigen, no significant differences in fecal scores between allergen-exposed (S) and non-sensitized (NS) mice were observed; however, the CMA-groups showed higher fecal scores (reflecting softer stools and signs of inflammation) compared to the HC-groups. Four hours after the third and the fifth sensitization, the fecal scores were higher in CMA-S group compared to CMA-NS group (p<0.001). Approximately 24h after the third but
not after the fifth sensitization, the enhanced fecal scores normalized (P<0.05 at day 44, comparison CMA-S vs CMA-NS). In contrast, no significant differences were noticed between HC-S as compared to HC-NS at the third and the fifth sensitization, which may indicate a protective effect of healthy microbiota upon allergen exposure (Figure 5A).

Figure 5: (A) Fecal scores on a scale from 0-4, with 0 indicating normal pellets and increased scores indicating softer to diarrheic pellets and/or anal inflammation. Fecal pellets were scored after the first sensitization at D15 then D16, after the third sensitization at D29 then D30 and after the fifth sensitization at D43 then D44. (B) Clinical score on a scale from 0 to 6 with line at median. Increased scores indicate increased allergic symptoms based on monitoring (i) scratching behavior, (ii) loss of mobility, and (iii) puffiness scored during a 15 min interval (30 min after BLG-challenge). (C) Concentrations of mMCP-1. (D) BLG specific IgE. (E) total IgE, (F) BLG-specific IgG1/IgG2a ratio and (G) total IgG1/IgG2a ratio are shown in dot plots with line at median. P-values were calculated using Mann-Whitney test (*P≤0.05, **P≤0.001, ***P≤0.0001).
CMA-associated microbiota induces increased susceptibility to develop an allergic reaction, without allergen-specific sensitization

The clinical scores were significantly higher in the CMA-S group compared to non-sensitized CMA-mice (p<0.001). Additionally, the clinical score of the CMA-S group was significantly higher than that of the HC-S group (p<0.05) (Figure 5B). The rectal temperature was also measured but only two mice (one in both sensitized groups) developed an anaphylactic reaction (data not shown). The increased clinical score in CMA-S versus HC-S was not correlated with differences in mast cell degranulation (Figure 5C), or specific markers of sensitization (IgE and IgG1/IgG2a ratio, Figure 5D and 5F, respectively). Independent of FMT, mMCP-1 levels were higher in sensitized mice compared to non-sensitized controls (Figure 5C). However, levels of total IgE in CMA-mice, independent of sensitization status, were increased in comparison with HC-mice (p<0.001) (Figure 5E). Similarly, the ratio of total IgG1/IgG2a representing the Th2/Th1 balance was increased in CMA-mice compared to HC-mice (Figure 5G). The ratio of total IgG1/IgG2a in HC-S group was also increased compared to the HC-NS group (p<0.001).

Colonic gata3 and foxp3 gene expression are impacted by the CMA-microbiota

The mRNA gene expression of tbet, gata3, roryt and Foxp3 were measured in duodenum, jejunum, ileum and colon tissue and relative expression levels of the CMA-S groups were compared to HC-S group (Figure 6). In small intestinal tissue only roryt expression was decreased in CMA-S group compared to HC-S group (p=0.05, Figure 6D). In contrast, colonic tissue showed significantly increased expression of gata3 and foxp3 in CMA-S group compared to HC-S (p<0.05 and p<0.001, Figure 6B and 6C, respectively) and non-significant increased expression of fcyRIII (p=0.07, Figure 6H). No differential expressions were observed for tbet, fcer2a, tgfβ and il17.
Figure 6: Gene mRNA expression in the small intestine (duodenum, jejunum, ileum) and colon of HC and CMA sensitized groups. The mRNA gene expression was normalized against the reference gene of HC sensitized (HC-S) mice. The relative gene expression of (A) tbet, (B) gata3, (C) foxp3, (D) rorγt, (E) IL17, (F) tgf-β, (G) fcer2a and (H) fcγrIII genes are shown as bar plots (mean with SEM). P-values were calculated using Mann-Whitney test (*P≤0.05, **P≤0.001).
CMA-microbiota and its activity induce an enhanced atopic orientation

We evaluated the systemic response with the concentration of cytokine produced by splenocytes and MLN cells following in vitro BLG stimulation. No differences between CMA-S group and HC-S group in cytokine levels were observed (Figure S3). We performed a PCA analysis of mice microbiota compositions at D50 incorporating immune (serum and cytokine data) and bacterial metabolites (SCFAs and lactate) for samples with data available (Figure 7). CMA-associated microbiota was associated with increased levels of SCFAs, which was most pronounced for iso-butyric acid and iso-valeric acid, which corroborates with the differences observed in infant stool samples. The CMA-associated microbiota composition and activity positively correlated with increased serum IgE, IgG1, and IL-5 levels in the MLN (Figure 7).

Discussion

Herein, we investigated whether an altered gut microbiota in CMA-infants as compared to healthy infants may contribute to allergic disease. Fecal microbiota signatures associated with CMA, which were characterised by decreased levels of bifidobacteria and increased levels of the ER/CC-group (resembling Lachnospiraceae spp.), were associated with enhanced allergic response in a gnotobiotic murine model.
of CMA. Importantly, these specific microbiota signatures corroborate findings of paediatric studies comparing CMA-infants with healthy breastfed infants (7-10, 13). These data strengthen the importance of the early life acquisition and establishment of bifidobacteria for healthy development (23, 24), and builds upon our previous findings with the same model, in which we demonstrated protective effects against CMA-development upon colonization of germ-free with healthy breastfed infant gut microbiota (12).

The enhanced allergic responses observed in the CMA-colonized mice compared to HC-colonized mice were associated with diarrhea-like symptoms and signs of colonic inflammation following sensitization and challenge with the allergen. Moreover, we observed increased total IgE and increased total IgG1/IgG2a ratio, which reflects an enhanced Th2/Th1 status. These observations were consistent with increased gata3 mRNA expression in the colon, which is a marker of Th2 lymphocytes. In addition, IL-5 production by mesenteric lymph nodes was correlated with the microbiota of allergic children (Figure 7). Interestingly, the expression of fcγRIII gene was increased in CMA-S group in the colon. FCγRIII is an activating receptor binding mouse IgG1 (25), which is expressed on mastocytes and basophils. These results suggest a pathway linked to IgG1 and basophils, which has been implicated in anaphylaxis (26, 27). In our model, BLG-IgG1 levels were not increased in CMA-S group, showing the complexity of allergic mechanisms. All these elements confirm a Th2 orientation of the immune system following transplantation of the CMA-associated microbiota. However, this Th2 profile was not associated with differences in mMCP-1 levels or allergen-specific immunoglobulin levels between the CMA-S and HC-S groups, which possibly points to a non-IgE mediated immune response in mice transplanted with CMA-microbiota. Moreover, human patients with non-IgE dependent allergy pre-dominantly present gastro-intestinal tract symptoms (2). The diarrhoea-like symptoms and signs of colonic inflammation in mice transplanted with the CMA-associated microbiota support this hypothesis.

Several studies showed that germ-free mice have increased total IgE levels compared to conventional mice, and that levels can be normalized when germ-free mice are colonized with commensal microbiota until 8 weeks of age (28). It has been
argued that high total IgE is a poor predictor for food allergy (29), while other studies linked increased IgE levels to poor long-term outcome in atopic dermatitis (30) or an increased risk of developing other allergic manifestations (31). We show that colonization with a HC-microbiota but not colonization with a CMA-microbiota maintains low total IgE levels like the non-sensitized control mice. This may be due to the observed enrichment of bifidobacteria and *Anaerostipes* spp. in HC-associated mice (Figure 4). We previously showed that *Anaerostipes* spp. can form a trophic chain in the developing infant gut by converting the bifidobacterial breakdown products of complex carbohydrate degradation and the resulting fermentation products lactate and acetate into butyrate, which was associated with protection from eczema in infants at increased risk for allergy (17, 32). Interestingly, Feehley et al. recently also identified *Anaerostipes* spp. as having a key role in allergy protection using a similar murine model for CMA (33).

Unexpectedly, we observed in the colon an increase in *foxp3* mRNA gene expression in CMA-S group compared to HC-S group. Foxp3 has been associated with the production of Th2 cytokines in several cell lines, including Foxp3+Gata3+ cells (34-36), as well as with regulatory T-cells (Treg) (37). If associated with Treg cells, this increase could be the hallmark of the immune response towards Th2 induction. Another explanation of the increased *foxp3* expression could be linked to the increased ER/CC bacterial group and the associated increase of cecal butyrate, which have been implicated in the induction of *foxp3* Treg cells (38, 39). In our study, CMA-infants as well as the recipient mice showed increased levels of butyrate, but also increased levels of the branched-chain SCFAs iso-butyrate and iso-valerate, which corroborate the observations of Thompson-Chagoyan et al. in paediatric patients (7). The increase of branched-chain SCFAs suggest an increased proteolytic activity in CMA-infants versus healthy infants, while these can only be derived from branched-chain amino acids and therefore provide an indicator of protein fermentation (40). Butyrate, which can be derived both from carbohydrates and protein fermentation (40), may therefore have partly resulted from the increased proteolytic activity of the CMA-associated versus HC-associated gut microbiota. Several Clostridia and members of *Peptostreptococcaceae* spp. are involved in protein and
amino acid fermentation (41) and were consistently increased in mice receiving the CMA-FMT. Protein degradation and its associated metabolites, as well as specific members of Peptostreptococcaceae, most notably Peptoclostridium (Clostridium) difficile, are generally associated with compromised health and may be implicated in the increased severity of allergic responses and symptoms observed in CMA-mice (42). In contrast, the increased saccharolytic activity as suggested by the increased abundances of bifidobacteria and Anaerostipes spp. in the HC-mice may form a more beneficial supply of butyrate to the host (40).

This study has several limitations. The CMA-infant selected was on concomitant medication and had clinically improved on an elimination diet, which are both factors known to influence the gut microbiota (43), but are also inherent to the medical condition. Moreover, only two representative microbiotas were chosen to avoid antagonistic effects that might arise by mixing different microbiota. Additionally, the FMT to germ-free mice results in adaptation to the new host (44). In our study however, despite this adaptation, the microbiota preserved its characteristic signatures. Therefore, we consider that our transfer led to a valid model to study the impact of microbiota on allergy development.

In conclusion, we demonstrated that infant microbiota with a low bifidobacteria: Lachnospiraceae ratio oriented the mice immune system towards a Th2 atopic profile resulting in enhanced allergic symptoms. Although the exact mechanism warrants further research, our data suggests that strategies to enrich the gut microbiota of infants with bifidobacteria may aid in the prevention and treatment of food allergy.

**Acknowledgements**

We would like to thank all infants and their caregivers for their participation in this study. We also would like to thank Iharilalao Dubail and her team from Animal Platform CRP2, Faculté de Pharmacie de Paris, for technical support in animal experiments. The authors also would like to thank: The Analytical Sciences team and Heleen de Weerd of the Gut Biology and Microbiology team of Nutricia Research, The Netherlands for the support in metabolite analysis and bioinformatic analysis, respectively.
Supporting Information

Supplementary Materials and methods

DNA extraction, 16S rRNA-gene sequencing and bioinformatics

DNA was extracted from fecal and cecal samples using a phenol-chloroform based method combined with bead-beating as described previously (17). Infant and mice microbiota compositions were determined by sequencing and bioinformatic analysis as described in detail before (13). Briefly, extracted DNA samples were profiled by sequencing the PCR-amplified V3-V4 regions of the 16S rRNA gene on an Illumina MiSeq instrument (San Diego, USA). Illumina reads were preprocessed, quality filtered, merged and analyzed with an adapted version of the ‘Quantitative Insights Into Microbial Ecology’ (QIIME) v1.9.0 pipeline (45). Sequences were clustered into Operational Taxonomic Units (OTUs) based on 97% sequence identity using VSEARCHv2.4.1 with exclusion of chimeric sequences identified against the RDP gold database (46, 47). Taxonomic assignment was performed using the RDP classifier (48) against the SILVA123 database (49). Singleton OTUs, OTUs with eukaryotic assignments, and OTUs with a low relative abundance up to 0.005% were excluded from further downstream analysis (50). The species diversity (α-diversity) metrics for richness (observed OTUs) and the Shannon index for diversity (51) were calculated using the R-package phyloseq with correction for the differences in sequencing depths by rarefaction (52).

Clinical Scoring

Scoring was adapted from Perrier et al (53). Allergic symptoms were evaluated based on three criteria: scratching behavior, loss of mobility, and puffiness (including bristled fur, oedema around nose and eyes, laborious breathing). Scratching was defined as the number of scratching episodes per 15-min interval as follows: 1-3 episodes=0, 4-5 episodes=1, and >6 episodes =2. Loss of mobility was graded in terms of the duration of absence of any movement as follows: <10 min=0; >10 min = 1, during the 15 min =2. Puffiness was graded as none=0 and puffiness=2. The
clinical score was defined as the sum of the three individual scores, and therefore, ranged from 0 to 6.

Measurement of plasma mouse mast cell protease-1 and sensitization markers

Plasma mMCP-1, total IgE, IgG1 and IgG2a were measured by ELISA according to manufacturer's recommendation (Ready-SET-Go!, Ebioscience, San Diego, USA). Measurements of BLG-specific IgE levels were performed by capturing with rat anti-mouse IgE (Pharmingen, BD Biosciences, Le Pont-de-Claix, France) antibody and by detecting with biotinylated BLG (Pierce, Rockford, USA) and streptavidin-horseradish peroxidase (HRP) (Clinisciences, Nanterre, France) as previously described (11). Data were expressed in terms of OD at 450nm. Levels of anti-BLG IgG1 and IgG2a were determined using BLG as the capture antigen, and goat anti-mouse IgG1 and IgG2a-HRP (Southern Biotech, Birmingham, USA) were labeled as detection antibodies as previously described (54).

Cytokine production by BLG-stimulated splenocytes and MLN lymphocytes

Spleens and MLN were crushed, filtered and treated to obtain 2x10^6 cells per well (54). Cells were cultured in 24-well plates with and without 2.5mg/ml BLG at 37°C in a 5% CO2, 95% air atmosphere for 48h. Culture supernatant levels of tumor necrosis factor-α (TNF-α), interferon-γ (IFN-γ) interleukin-4 (IL-4), IL-5, IL10, IL13, IL17, IL22 and IL33 were quantified using a ProcartaPlex (Fisher, Hampton, USA), according to the manufacturer's instruction.

Relative expression of T-helper cytokines and immunoglobulin receptor genes in the gut

Total RNA was isolated from 2-cm segment of duodenum, jejunum, ileum and colon, devoid of Peyer's patches, using a RNeasy Plus universal kit (Qiagen, Courtaboeuf, France). Extracted RNA was treated with DNase I and first-strand cDNA was synthesized using Invitrogen reagents (Thermo Scientific, Illkirch, France) (54).

Quantitative real-time PCR (qRT-PCR) was performed on an ABI Prism 7900HT sequence detection system (Applied Biosystems, Thermo Scientific, Illkirch, France). QuantiTect SYBR green and QuantiTect primer assays (Qiagen) were used to
quantify Transforming Growth Factor β (TGF-β), FcγR3 and Fcer2a. TaqMan gene expression assays with TaqMan universal master mix II (Applied Biosystems) were used to quantify IL17-A, Foxp3, T-bet, Gata3 and RorγT. Measurement were performed in duplicate, and gene expression levels were calculated using the $2^{-\Delta\Delta^{CT}}$ method (55), where $CT$ is the threshold cycle, with the TATA box (TaqMan) assay as reference gene.
**Supplementary Tables**

**Table S1: Subject characteristics.**

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Sex</th>
<th>Age</th>
<th>Status</th>
<th>Delivery</th>
<th>Family History or atopy</th>
<th>Antibiotics</th>
<th>Previous medicine*</th>
<th>Current medicine</th>
<th>Previous milk-feeding</th>
<th>Current milk-feeding</th>
<th>Formula</th>
<th>Age when started formula</th>
<th>Age weaned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>7m</td>
<td>HC</td>
<td>NVD</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>BF</td>
<td>BF</td>
<td>NA</td>
<td>NA</td>
<td>5.5m</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>5.5m</td>
<td>HC</td>
<td>NVD</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>mixed</td>
<td>mixed</td>
<td>SF (1/day)</td>
<td>birth</td>
<td>5m</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>9m</td>
<td>HC</td>
<td>LSCS</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>mixed</td>
<td>mixed</td>
<td>SF (2/day)</td>
<td>1w</td>
<td>5.5m</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>10m</td>
<td>All</td>
<td>LSCS</td>
<td>Yes</td>
<td>No</td>
<td>Ranitidine, Lansoprazole</td>
<td>BF</td>
<td>FF</td>
<td>AAF</td>
<td>4.5m</td>
<td>4.5m</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>7m</td>
<td>HC</td>
<td>NVD</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>BF</td>
<td>BF</td>
<td>NA</td>
<td>birth</td>
<td>5m</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>9m</td>
<td>HC</td>
<td>NVD</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>BF</td>
<td>BF</td>
<td>NA</td>
<td>NA</td>
<td>4.5m</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>16m</td>
<td>All</td>
<td>LSCS</td>
<td>Yes</td>
<td>No</td>
<td>Infacol, Colic ease</td>
<td>No</td>
<td>BF</td>
<td>FF</td>
<td>AAF</td>
<td>5w</td>
<td>6m</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>9m</td>
<td>HC</td>
<td>LSCS</td>
<td>No</td>
<td>No</td>
<td>Paracetamol</td>
<td>No</td>
<td>BF</td>
<td>FF</td>
<td>AAF</td>
<td>6m</td>
<td>6m</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>12m</td>
<td>All</td>
<td>NVD</td>
<td>Yes</td>
<td>No</td>
<td>Paracetamol</td>
<td>No</td>
<td>BF</td>
<td>FF</td>
<td>AAF</td>
<td>4m</td>
<td>6m</td>
</tr>
</tbody>
</table>

Sex: M=Male, F=Female; Status: HC=healthy control, All=allergic; Delivery: NVD = normal vaginal delivery, LSCS = lower segment caesarean section, El= Elective; *Previous medicine including antibiotics. Previous/Current milk-feeding: BF=breastfed, FF=formula-fed or receiving both breast- and formula-feeding (mixed), SF=standard cow’s milk-based formula, AAF=Amino acid-based formula. Ages are indicated in months (m) or in weeks (w). In bold the infants selected for fecal microbiota transfer to murine model.
Table S2: Characteristics of CMA patients.

<table>
<thead>
<tr>
<th></th>
<th>Patient 4*</th>
<th>Patient 7</th>
<th>Patient 9</th>
<th>Patient 10</th>
<th>Patient 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (months)</strong></td>
<td>10</td>
<td>16</td>
<td>16</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td><strong>Clinical manifestations</strong></td>
<td>Urticarial rash, vomiting/GOR</td>
<td>Immediate vomiting, rash, abdominal pain, blood per rectum</td>
<td>Urticarial rash, vomiting, failling growth, loose stools (with blood and mucus), episodes of chestiness</td>
<td>Urticarial rash, vomiting</td>
<td>Immediate swelling and rash, vomiting/GOR, failling growth, constipation</td>
</tr>
<tr>
<td><strong>Personal atopic past history</strong></td>
<td>CMPA, allergies to wheat, egg and soya; Reactive airway disease</td>
<td>CMPA, allergies to egg, wheat, soya, seafood and beef; Eczema</td>
<td>CMPA, allergies to egg, wheat, soya, shellfish and nut; Eczema</td>
<td>CMPA</td>
<td>CMPA, allergies to egg, wheat, soya and rice; Eczema</td>
</tr>
<tr>
<td><strong>Familial atopic past history</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Final diagnosis</strong></td>
<td>Multiple Food Protein Allergies</td>
<td>Food protein-induced enteropathy syndrome</td>
<td>Multiple Food Protein Allergies</td>
<td>CMPA</td>
<td>Multiple Food Protein Allergies</td>
</tr>
<tr>
<td><strong>Amino acid-based formula</strong></td>
<td>Nutramigen AA®</td>
<td>Neocate®</td>
<td>Neocate®</td>
<td>Neocate®</td>
<td>Neocate®</td>
</tr>
</tbody>
</table>

*Post-exclusive breast feeding, infant #4 developed urticarial rash and classic CMA symptoms when challenged with dairy on repeat occasions; with symptom resolution on exclusion. Infant 4 was stable (amino acid-based formula with dairy/egg/wheat/soya exclusion) when FMT sample was obtained. CMPA=cow’s milk protein allergy. GOR=gastroesophageal reflux.

Table S3: Summary of sequence depths obtained with 16S rRNA-gene sequencing across the different samples.

<table>
<thead>
<tr>
<th>Sample description</th>
<th>n samples</th>
<th>Average sequence depth</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor samples (fecal samples)</td>
<td>10</td>
<td>60285</td>
<td>10743</td>
</tr>
<tr>
<td>Mice D15 (fecal pellets)</td>
<td>84</td>
<td>47865</td>
<td>12625</td>
</tr>
<tr>
<td>Mice D50 (cecal content)</td>
<td>89</td>
<td>57996</td>
<td>11587</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>183</strong></td>
<td><strong>53471</strong></td>
<td><strong>13053</strong></td>
</tr>
</tbody>
</table>
**Supplementary Figures**

**Figure S1: Fecal markers analyzed in infant donor samples.** Box-whisker plots with FISH quantified levels of *Bifidobacterium* spp. (A), and *Eubacterium rectale – Clostridium cocoides* (ER/CC) group (B), and stool pH (C). Concentrations of fecal SCFAs (not valeric acid, due to majority of samples with undetectable levels) and lactic acids are summarized (D-J), as well as the Log-10 concentrations of secretory IgA (sIgA) (K), Calprotectin (L) and eosinophil-derived neurotoxin (EDN). Final selected donors for fecal transfer to mice model are shown as grey symbols. Statistics comparing healthy control (HC) with cow’s milk allergic (CMA) infants were performed using Mann-Whitney test. *P<0.05, **P<0.001.
Figure S2: Levels of SCFAs and lactic acids in cecal content mice (D50). Box-whisker plots (Tukey method) with the concentrations of acetate (A), propionate (B), butyrate (C), iso-valeric acid (D), D-lactic acid (E) and L-lactic acid (F). The detection of iso-butyrate (G) and valerate (H) are presented as percentage of samples with detectable levels (P) and undetectable levels (A). Statistics comparing the HC-groups with the CMA-groups were performed using Kruskal-Wallis test with Dunn's correction for multiple comparisons and applying fisher's exact test for presence (P)-absence (A) data. *P≤0.05, **P≤0.001, *** P≤0.0001, **** P≤0.00001.
Figure S3: Cytokines produced after BLG stimulation. (A) Splenocytes, (B) MLN lymphocytes shown as bar plots (mean with SEM). Statistics comparing HC-S with CMA-S were performed using Mann-Whitney test.
References


26. Miyajima I, Dombrowicz D, Martin TR, Ravetch JV, Kinet JP, Galli SJ. Systemic anaphylaxis in the mouse can be mediated largely through IgG1 and Fc gammaRIII. Assessment of the cardiopulmonary changes, mast cell degranulation, and death associated with active or IgE- or IgG1-dependent passive anaphylaxis. J Clin Invest. 1997;99(5):901-14.


General discussion

The aim of this thesis was to investigate normal gut microbial trajectories in early life and to identify deviating patterns that are associated with allergic manifestations, such as atopic dermatitis and food allergy. Generally, it is estimated that the gut microbiota reaches its adult-like configuration around 3 years of age (Chapter 2), although some taxa are even not yet fully established before 5 years of age (1). We hypothesized that a suboptimal implementation of the gut microbiome may contribute to the onset of allergy and/or enhancement of allergic symptoms. Perturbation of gut microbiome at critical time windows during early life development may be particularly important, because that is when the immune system, with most of its cells residing in the gut, matures (2). Modern lifestyle and associated factors, such as antibiotic-exposure, caesarean-section delivery and low breastfeeding rates, that are associated with depletion of key gut microbial taxa, may have disrupted the mechanisms of mucosal immunologic tolerance against innocuous antigens, such as food proteins. These factors in addition to genetic risk factors have been implicated in the rising trend of allergies observed in affluent countries (Chapter 2).

The sequencing era

The number of observational paediatric studies, both prospective and case-controlled, that report on the association between altered gut microbiota and allergic disease have rapidly increased since our 2014 review of existing literature (Chapter 2). The decline in costs for DNA sequencing technologies that allow for rapid and accurate characterization of microbial communities greatly contributed to the expansion of this field of research. The plethora of 16S and metagenomic surveys of human samples have vastly improved our understanding of the microbiota composition and functional potential in both healthy and diseased states. However, the accurate description of these surveys critically depends on sample collection, storage, processing, and on the downstream bioinformatic and statistical analyses applied (3-7). In this thesis the importance of PCR primer choice in 16S rRNA-gene sequencing was shown by curating one of the recommended primer-pairs from the
NIH Human Microbiome Project (8). The primer-set was found to contain a one-base mismatch in the reverse primer for bifidobacteria, which may explain the relatively low abundance of this genus reported in several US birth cohorts (9). We showed that application of the ‘bifido-optimised’ primer-pair successfully detected and quantified bifidobacteria without distorting proportions of other taxa (Chapter 3).

**The development of gut microbiota**

**Early life nutrition**

We applied the 16S rRNA-gene sequencing technology to investigate the gut microbiota development and the pivotal role of early life nutrition in two clinical trials, which included infants with a family history of allergic disease (and therefore deemed ‘high-risk for allergy’) (Chapter 4), as well as infants with cow’s milk allergy (CMA) (Chapter 5). Since more than 100 years it has been known that breastfed babies have a different gut microbiota compared to infants receiving a formula based on cow’s milk (10-13), which is explained by a wide array of non-digestible oligosaccharides present in human milk, known as human-milk oligosaccharides (HMO) (14). A growing number of studies indicate that the addition of specific prebiotic or synbiotic ingredients to infant formula can have similar functionalities as HMOs on gut microbiota composition and activity (15). Hence, such microbiota modulators are of interest in nutritional strategies for allergy prevention and treatment in infants for whom breastfeeding is not possible or insufficient. Based on HMO-research a specific mixture of short-chain galacto-oligosaccharides and long-chain fructo-oligosaccharides (scGOS/lcFOS) (9:1) was developed to mimic the molecular weight profile and partly building blocks of HMOs in breast milk (15), and its supplementation to infant formula has been shown to reduce the risk of allergy in not-exclusively breastfed infants, both at high- (16, 17) and at low-risk for developing allergy (18), which contributed to new guidelines from the World Allergy Organization that recommend prebiotics for allergy prevention under certain conditions (19). Current international guidelines for the dietary management of allergic diseases for not exclusively breastfed infants also include the recommendation of partially
hydrolysed cow’s milk protein-based formulas (pHP) for high-risk infants (20). However, the preventive effects of reduced allergen exposure by such formulas is currently being debated and challenged (21). Infants suffering from cow’s milk allergy (CMA) rely on the dietary avoidance of cow’s milk to resolve symptoms, which include specialized infant formula for those who cannot be breastfed (22). While most infants respond to extensively hydrolysed cow’s milk-based formulas (eHP), some require amino acid-based formulas (AAF) due to an increased risk to an anaphylactic reaction, or more complex symptoms, or when symptoms do not resolve on eHP (22-24). In the cohort with infants that were high-risk for developing allergy (Chapter 4), we investigated the modulatory effects of a partially-hydrolyzed cow’s milk protein formula (pHP) including the scGOS/lcFOS (9:1) mixture on gut microbiota in the first 6 months of life. In chapter 5 the modulatory effects of an amino acid-based formula (AAF) with short-chain and long-chain fructo-oligosaccharides (scFOS/lcFOS) (9:1) and the probiotic strain Bifidobacterium breve M-16V was investigated in infants suffering from cow’s milk allergy (CMA). This synbiotic concept was based on the scGOS/lcFOS-mixture, however in which scGOS was replaced by scFOS, since the former is produced from cow’s milk-derived lactose, which may pose risks in infants with severe CMA that rely on an elimination diet for the resolution of allergic symptoms (25). Importantly, the combination of the prebiotic with the probiotic Bifidobacterium breve M-16V provided preclinical proof to be more effective in treating the allergic response in a murine model of food allergy than the sole administration of the scFOS/lcFOS-mixture (26).

The healthy reference in gut microbiota development

In both trials, we reaffirmed that the microbiota of breastfed infants was characterized by a low diversity and Bifidobacterium domination. Moreover, we showed that the addition of prebiotics to pH (Chapter 4) and the addition of synbiotics to AAF (Chapter 5) led to a taxonomic profile and diversity closer to that of the breastfed reference. Infants who received infant formula without prebiotics or synbiotics, however, adopted a more diverse adult-type of microbiota, which was characterized by an increase of microbes belonging to the order of clostridia and its main bacterial family of Lachnospiraceae. Importantly, we observed these
closer-to-breastfed effects during the exclusive milk-feeding period (typically before 4 months of age), but also during the complementary feeding period, namely at 6 months of age (Chapter 4) and 8 months of age (Chapter 5), respectively. This period, when the first solid foods are introduced and gradually replace the milk-based diet (i.e. human milk and/or infant formula when breastfeeding is not possible), marks an important shift of the microbial ecosystem to the diversifying infant diet (27). The WHO guidelines recommend to phase in solids at six months of age with continued breastfeeding up to 24 months of age (28, 29). European guidelines recommend to phase in complementary feeding between 4-6 months of age with continued breastfeeding (30). The increased \textit{Bifidobacterium}/\textit{Lachnospiraceae} ratio (B/L-ratio) indicate that the addition of these specific prebiotic or synbiotic ingredients may also normalise the rate of gut microbiota establishment during weaning towards that observed in breastfed infants. The establishment of gut microbiota in breastfed infants is namely characterized by a more gradual transition from infant- to adult-like composition compared to non-breastfed infants, which was confirmed in several recent longitudinal 16S and metagenomic surveys of infant samples (31-33), as well as in a number of meta-analyses performed on publicly available data (9, 34-36). The timely establishment of gut microbiota as observed in breastfed infants may therefore be essential in forming a stable ecosystem (34-36), and may contribute to the association of breastfeeding with several health benefits, such as a lower risk of diarrhoeal diseases and infections in early childhood (37), a lower risk of metabolic diseases later in life (38), and possibly also a lower risk of developing allergy and asthma (20, 39-41).

Diversity and ecosystem development

The decreased microbial diversity observed in breastfed infants, but also observed in infants receiving prebiotic- and synbiotic-supplemented formula, may seem counterintuitive. Especially since low gut microbiome diversity in early infancy has been reported to precede allergic manifestations, such as eczema (42-44), which we however did not confirm in our study with high-risk infants (Chapter 4). Arguably, an increased diversity does not necessarily reflect improved ecosystem function, stability or resilience to perturbation, but should be viewed in its ecological context.
(45, 46). In that respect, the lower diversity of breastfed versus formula-fed infants versus the lower diversity associated with later-life allergic manifestations likely reflects different underlying ecologic causes and consequences. For instance, Abrahamsson et al. (43) linked the lower diversity at 1 month of age in infants developing atopic eczema by 2 years of age to a decreased diversity of the bacterial phylum/genus Bacteroidetes / Bacteroides, also when limiting their analysis to exclusively breastfed infants (47). This may indicate decreased maternal transmission of these pioneering infant gut colonizers upon birth (48), which seemed unrelated to delivery by caesarean section or exposure to perinatal antibiotics, both well-known risk-factors for decreased vertical transmission and allergic disease (49). In contrast, the lower microbial diversity that we observed in breastfed, prebiotic-fed and synbiotic-fed infants as compared to control-fed infants, reflected the increased Bifidobacterium/Lachnospiraceae ratio (B/L-ratio). These compositional differences resulted in specific changes in gut eco-physiology with a slightly acidic pH, increased concentrations of lactate, and a specific short-chain fatty acid (SCFA)-profile, which was high in acetate and low in butyrate, propionate and branched-SCFAs (iso-valerate and iso-butyrate) (Chapter 4 and Chapter 5). The enhanced levels of colonic lactate and acetate, both fermentation end-products of bifidobacteria, are typically associated with increased colonization resistance to potential pathogens (50-52) and may contribute to the protection against infections associated with breastfeeding (37). Interestingly, lower incidence of infections was also reported for infants receiving a formula with scGOS/lcFOS (16, 53), as well as for infants receiving the AAF formulation with synbiotics (54, 55) that was reported on in this thesis (Chapter 5). The increased levels of branched-SCFAs observed in the guts of control-fed infants possibly reflect an increased proteolytic microbial metabolism (56, 57), due to the deprivation of fermentable carbohydrates (HMOs or prebiotics) that may lead to a shift from saccharolytic processes to lower yield proteolytic processes (46). It was recently observed in adult populations that a firm stool consistency, which is a proxy for long colonic transit time, correlates with high gut microbial diversity and increased proteolytic activity (58, 59). These observations in adults show interesting parallels to observations in infants, in which lower stool consistency and increased
stool frequency are observed for infants being breastfed or receiving formula supplemented with scGOS/lcFOS when compared to formula without prebiotics (60, 61). The softer stools and increased stool frequency were also confirmed for the prebiotic and synbiotic formulations investigated in this thesis (62, 63). Together, this suggests that a high gut microbial diversity does not per se imply a healthy gut microbial ecosystem and points at the sachharolytic/proteolytic ratio as an important factor to consider in microbiome and metabolomics studies, since several end-products of microbial proteolysis are considered detrimental to host health (46, 58, 64).

Gut microbiota and allergic disease

In the cohort of infants that were high-risk for developing allergy (Chapter 4), we found that infants who went on to develop atopic dermatitis (AD) showed discordant temporal development of bacterial taxa and the metabolites lactate, propionate and butyrate around the time that complementary feeding started. This nested case-control comparison was controlled for the factors that were identified as influencing the microbiota of these infants including feeding group, ethnicity and household exposure to siblings. At first, infants developing AD showed decreased levels of lactate and increased levels of propionate and butyrate at 12 weeks of age (before weaning); in contrast, they showed increased levels of lactate and decreased levels of propionate and butyrate at 26 weeks of age (during weaning). Microbially produced lactate is only intermediately present in the healthy adult, as it is converted into propionate or butyrate by a subset of lactate-utilizing bacteria (65). Indeed, we found that both Eubacterium and Anaerostipes spp., which are known to utilize lactate and acetate to produce butyrate, were enhanced in healthy infants compared to infants with AD. Importantly, we also found that infants developing AD showed decreasing levels of bifidobacteria over the first 6 months of life, which was in contrast with the increasing levels observed for healthy infants. This possibly indicates that the bifidogenic environment with high levels of lactate and acetate and their consumption by such specialist microbes may have a key role in establishing a stable community in the gradually diversifying infant gut and may protect from developing allergy. The start of weaning may form a critical step in the development of oral tolerance, while
experiments with germ-free mice have shown that the presence of both diet- and microbe-induced populations of regulatory T-cells (Treg) were required for induction of complete tolerance to food antigens (66, 67). Moreover, there is accumulating evidence that many of the microbially-derived colonic Tregs are induced via SCFAs that act as histone-deacetylase (HDAC) inhibitors and ligands for G-protein cell surface receptors (GPRs) (68-70). Indeed, several murine studies have shown that high-fiber diets effectively modulate gut microbiota and increase SCFAs, particularly acetate and butyrate, which exert regulatory effects on host immunity, including anti-inflammatory and anti-allergic effects (71-74). Interestingly, infants receiving the pHp-prebiotics demonstrated increased systemic Treg-numbers compared to control-formula at 6 months of age, although no reduction of AD incidence was observed at 12 or 18 months (75). However, the incidence of AD seemed to be affected by the age the infant was introduced to solid foods, which deviated in most infants (55%) from the standard recommendation to start after 18 weeks of age (71).

In Chapter 6 we investigated whether the increased B/L-ratio that differentiated healthy breastfed infants from CMA-infants (Chapter 5) impacted immune and allergic response in a murine model of sensitization and challenge with cow’s milk protein. In a small case-control study with infants with suspected CMA and healthy breastfed infants, we could confirm the shift in B/L-ratio. Faecal microbiota transplantation (FMT) into germ-free mice from representative infants (matched for age, gender and birth-mode) largely replicated the characteristic compositional differences in microbiota, including the decreased diversity and proteolytic activity of the healthy breastfed infant (9 months of age) compared to the CMA-infant (10 months of age). Interestingly, the CMA-associated infant gut microbiota induced systemic Th-2 immunity with increased total IgE-levels and showed more severe clinical symptoms upon sensitization with cow’s milk allergen and cholera toxin as
adjuvant. Whether similar results would have been obtained with microbiota showing a decreased B/L-ratio but obtained from an otherwise healthy infant or when treated with prebiotics or synbiotics warrant further investigation. However, these results do suggest that perturbation of gut microbiome in infants with CMA contributes, at least in part, to disease onset and severity of symptoms.

**Overall conclusion and future perspectives**

Clearly the development of the gut microbiome from birth through to childhood is critical to establish a healthy symbiosis, since this is the period in which infants encounter external stimuli for the first time and the body is trained to respond to these stimuli. Human milk is the reference, effectively shaping the infant microbiome, which is associated with optimal immune maturation and protection against infections and potentially also allergy. Specific prebiotic and synbiotic ingredients can modulate the microbiota closer to that of breastfed infants leading to clinical benefits related to infection and allergy. However, much is yet to be learnt about how and when microbes or their functional triggers optimally impact the host and protect from allergic disease. Moreover, it is crucial to understand how the many covariates, such as birth mode, antibiotics and household exposures, interfere with the colonisation pattern. Future studies in allergy need to carefully consider the effect of complementary feeding (type and timing) and treat it as an integral part of the dietary strategy for prevention and treatment. Longitudinal studies are needed to better comprehend the timely establishment of key species to develop a gut eco-physiology to form a stable ecosystem, which supports optimal gut and immune development. The omics tools currently available to perform characterization of microbiome functionality and bioactive compounds like proteomics or metabolomics integrated with 16S surveys or metagenomics will certainly bring our understanding to the next level. These tools and the identification of different endotypes of allergic disease based on clinical symptoms and immunomes, will improve our understanding of the structure and function of the microbiome in both diseased and healthy states. This integrated approach may lead to novel personalized preventive and therapeutic nutritional strategies in allergic disease.
Summary

The development of the gut microbiome from birth through to childhood is thought to be important for establishing a healthy symbiosis, but much is yet to be learnt about this phase of microbiome development. Perturbations of gut microbiome during this development have been associated with the pathology of diseases, such as allergy. In **Chapter 2**, we describe this dynamic period of gut microbiome development, the various factors involved in shaping its composition and the importance for the concurrent maturation of the immune system. Environmental factors including birth mode, exposure to antibiotics and household exposures (such as siblings and furry pets) represent important factors impacting its development, and which also have been epidemiologically implicated in the risk to develop allergic disease. Several paediatric studies indeed associated altered gut microbiota with development of allergic disease. Breastfeeding represents the most significant factor in shaping early life microbiome and is associated with several short-term and long-term health benefits, including a lower risk of developing allergic disease. Specific prebiotic or synbiotic (when combined with probiotics) ingredients added to infant formula may exert similar effects on gut microbiota composition and activity, which may benefit infants for whom breastfeeding is not possible.

**Bifidobacteria** are the most abundant bacteria in breastfed infants, but often under-represented in 16S rRNA surveys of diversity, due to poor DNA extraction techniques, poor PCR primer choice or a combination of both. In **Chapter 3**, we optimized a commonly used “universal” PCR primer set and demonstrated the effective recovery of this genus without compromising the detection of other genera. In **Chapter 4** we applied the optimized pyrosequencing method described in **Chapter 3** to analyse the gut microbiota of infants at high-risk of developing allergy, who participated in a clinical trial that investigated the effects of a partially hydrolysed protein formula supplemented with prebiotics on the prevention of eczema. We showed that the taxonomic composition of infants receiving the prebiotic-supplemented formula was closer to that of breastfed infants when compared to infants receiving an infant formula based on intact protein without prebiotics, which was driven by increased abundance of *Bifidobacterium* spp. and
decreased abundances of *Clostridium* spp. and *Lachnospiraceae* spp. Importantly this also led to specific changes in gut eco-physiology characterized by a more acidic pH and increased levels of lactate and acetate as also characteristic for breastfed infants. In a nested case-control, we found that infants who developed eczema by 18 months of age showed an altered development of bacterial taxa and metabolites around the time that complementary feeding was started. The patterns identified suggested that the establishment of specific bacteria that utilize lactate and acetate to produce butyrate may have a role in protecting from the development of eczema.

Infants who suffer from severe CMA often rely on cow's milk protein avoidance and, when breastfeeding is not possible, on specialised infant formulas such as amino-acid based formulas (AAF) to meet their nutritional needs and in order to resolve the allergic symptoms they suffer from. In Chapter 5, we investigated the modulatory effects of an AAF supplemented with synbiotics on the gut microbiota and showed that both composition and activity approximated that of an age-matched breastfed reference group as opposed to infants receiving the AAF without synbiotics. Similar as observed in Chapter 4 this was driven by an increase of the *Bifidobacterium* spp./*Lachnospiraceae* spp. ratio (B/L-ratio).

In Chapter 6 we screened the intestinal microbiota of a small set CMA-infants and healthy controls to select donor samples for faecal microbiota transplantation into germ-free mice. We confirmed the decreased B/L-ratio in CMA versus healthy infants as observed in Chapter 5, which was maintained upon transplantation into the germ-free mice. Herein, we showed that CMA-associated infant microbiota resulted in an atopic orientation, with increased immunoglobulin E levels and an enhanced responsiveness to cow's milk allergen upon sensitization, which suggested that the pathobiology of allergic disease is mediated at least in part by gut microbiome perturbation.
Chapter 7

References


Appendices

Acknowledgement – Dankwoord ................................................................. 176
About the author ....................................................................................... 179
List of publications ................................................................................... 180
Overview of completed training activities ............................................... 182
Acknowledgement – Dankwoord

De eindstreep is in zicht! Om mij heen waarschuwen verschillende mensen voor het ‘zwarte gat’ waar ik straks geheid in zal donderen. Hoogste tijd dus voor het laatste deel van mijn ‘boekie’: het dankwoord.

Allereerst wil ik mijn mentoren van het eerste uur bedanken: Thank you Kaouther, Raish and Rocio for supporting me to ‘grow within Danone’ since I started my career at Nutricia Research in 2006. Under your supervision I developed from technician to scientist and finally felt sufficiently confident to pursue a PhD in microbiology. Guus, dank voor de steun tijdens ‘de laatste loodjes’ van mijn promotietraject. Ik heb zeer veel profijt gehad van jouw scherpe en kritische blik op veel van de hoofdstukken in dit proefschrift. Bovenal wil ik natuurlijk Jan bedanken, die als n+2 en promotor mij heeft geïnspireerd, gestimuleerd en ondersteund in mijn wetenschappelijke loopbaan. Dank dat je me de mogelijkheid hebt geboden om in 2012 de stap voor een promotieonderzoek te gaan wagen. Hoewel, voordat ondergetekende ook daadwerkelijk de stap zette, duurde toch nog wel even (daarom ook dank voor je geduld). Maar zo rond jouw inauguratie in Mei 2013 als bijzonder hoogleraar bij de leerstoelgroep van Willem de Vos had ik toch eindelijk een promotievoorstel en applicatie voor toelating bij de Wageningen Universiteit ingediend. Dank Willem! Veel respect voor jouw inhoudelijke kennis en manager-capaciteiten. Nooit eerder zulke snelle en efficiënte meetings meegemaakt waarin ik zoveel bruikbare en intensieve feedback kreeg! De belangrijkste hobbel voor mijn toelating (ik moest even terug de schoolbanken in gezien mijn HBO-achtergrond) was dan ook snel genomen, zodat ik in November 2013 dan ook echt officieel van start kon als gast promovendus bij de WUR.

Veel dank ook aan mijn co-promotor Clara Belzer, die door haar praktische instelling en hartelijke welkom op het Laboratorium voor Microbiologie het makkelijk maakte om mijn weg te vinden binnen de Wageningen Universiteit. A big thank you to all staff and fellow PhD candidates at the Laboratory of Microbiology; it was always great to interact with you during the PhD-meetings, symposia and teaching-activities! Special thanks to the Early Life Microbiology-club: Romy, Lennart, Petra and LooWee
as fellow-pioneers, and to the new colonizers of the club: Ioannis, Emmy, Jannie, Kevin and Patrick.

Of course, I’d like to express my gratitude to all my Nutricia teammates of the Gut Biology and Microbiology platform in Utrecht and Singapore. It is a pleasure to work with you and I am looking forward to all that is yet to come. Speciale dank voor Sebastiaan en Jolanda; voor alle inhoudelijke en persoonlijke steun tijdens mijn promotietraject. Ik heb jullie beiden zonder jullie medeweten als ‘other supervisors’ opgeschreven in mijn PhD projectvoorstel, maar kon dan ook altijd bij jullie terecht om even te sparren. Daarnaast heb ik een belangrijk deel van de resultaten in dit proefschrift te danken aan Rob, Tiemen en ‘stand-in’ Kees. Heel veel dank voor de ‘niet lullen - maar scheppen’ mentaliteit, de vele FISH en qPCR analyses en de oneindige stroom aan goede en slechte grappen. En ook zeker niet te vergeten: Heleen en Martin, de wizards die alle bio-informatica hobbels wegnamen. Jullie hebben mijn leven zoveel makkelijker gemaakt. Dank daarvoor!

I also want to acknowledge all co-authors and all internal and external members of the PATCH, ASSIGN and HEALMI study teams. I felt privileged to collaborate with you in this exciting multidisciplinary field of nutrition and health! This would not have been possible without your valuable contributions and support during the past years. Also, thanks to many more Nutricia-colleagues that were not directly involved in the work presented in this thesis: Thank you for your moral support and interest and in making Nutricia Research a great place to work.

Dan mijn paranimfen: Jolanda en Roger. Beiden belangrijke aanjagers voor de nodige afleiding naast het werk, zoals Roger’s fameuze bierproeverijen, de Ierse-pub avonden bij Mick O’Connels en de vuur-en-voer sessies bij Jolanda. Ook zorgden jullie ervoor dat ik niet de hele dag achter de computer bleef zitten d.m.v. de (bijna verplichte) dagelijkse lunch-wandelingen om even de kop leeg te maken. Ik ben dan ook zeer blij dat jullie mij bij willen staan op de grote dag en de weg ernaartoe!

Lieve familie, schoonfamilie, vrienden en de kameraden van d’n Achterhoek Tsjoek: het is klaar! Hoogste tijd weer voor een borrel, dank voor alle steun, afleiding en mooie momenten. Meer van dat graag! Deze keer geef ik een rondje.
Lieve pa, ma en zus. Hier is die dan, alles op zijn tijd. Met pa’s precisie en doorzettingsvermogen, ma’s zorgzaamheid en creativiteit, en Suzan’s onderzoeksgeest (helaas soms met mij als onwetend proefkonijn, want hoe scherp is een spaanse peper nu eigenlijk?). Heel veel dank voor alle steun en vertrouwen!

Lieve Leyla en Jonne, ik kreeg jullie er ‘gewoon’ bij tijdens mijn promotietraject – iets beters had ik niet kunnen wensen. Niet als studieobjecten (hoewel ik natuurlijk wel een paar poepluiers van jullie ingezet heb voor wetenschappelijk onderzoek), maar voor al het moois dat jullie brengen. Elke dag geniet ik van jullie!

Liefste Tetske, voor wie jij bent, voor wat wij samen hebben, voor dat ik gewoon mijzelf kan zijn en voor alle liefde. Dank voor alles!

Harm
January 2019
About the author

Harm Wopereis was born in Groenlo, the Netherlands, on 22 April 1981. After completing his secondary education at Marianum Scholengemeenschap in Groenlo in 1999, he continued his education in Biotechnology at Noordelijke Hogeschool Leeuwarden & Van Hall Larenstein (1999-2003). After obtaining his BSc degree, he started working for the knowledge institute of Van Hall Larenstein (2003-2006) on innovation projects for small and medium-sized enterprises in the field of environmental microbiology. Due to his interest in gut microbiology and applied research, he joined Danone Nutricia Research (at that time called Numico Research) to study the effects of nutrition on the gastrointestinal microbiota and health in early life. In the Gut Biology & Microbiology department headed by Professor Jan Knol, Harm developed from Assistant Scientist to Senior Assistant Scientist (2009) to Scientist (2012). In 2013, he was admitted as a guest PhD at the Laboratory of Microbiology of the Wageningen University. In his PhD project, he focused on the development of early life gut microbiota in health and allergic disease. The work was done under the supervision of Dr Clara Belzer, Prof. Dr Willem M. de Vos and Prof. Dr Jan Knol, and the results are presented in this thesis. Harm will continue his research as Senior Scientist at Danone Nutricia Research on bringing health through nutrition by supporting the gut microbiota to positively influence the gastrointestinal, immune and metabolic systems.
List of publications

Reviews

Research articles


*Joint first
# Overview of completed training activities

## Discipline specific activities

### Courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Institute</th>
<th>Location</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Light in the Intestinal Tract Tunnel</td>
<td>ABS / VLAG</td>
<td>Helsinki, FI</td>
<td>2009</td>
</tr>
<tr>
<td>Systems Biology: Statistical Analysis of ~Omics Data</td>
<td>EPS / VLAG</td>
<td>Wageningen, NL</td>
<td>2014</td>
</tr>
</tbody>
</table>

### Meetings

<table>
<thead>
<tr>
<th>Meeting</th>
<th>Organiser</th>
<th>Location</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gut Day Symposium</td>
<td>Gut Flora Foundation</td>
<td>Amsterdam, NL</td>
<td>2014</td>
</tr>
<tr>
<td>International Human Microbiome Congress</td>
<td>MetaHIT / IHMC</td>
<td>Paris, FR</td>
<td>2012</td>
</tr>
<tr>
<td>The European Academy of Allergy and Clinical Immunology (EAACI) - Annual congress (poster presentation)</td>
<td>EAACI</td>
<td>Copenhagen, DK</td>
<td>2014</td>
</tr>
<tr>
<td>KNVM/NVMM Scientific Spring Meeting (oral presentation)</td>
<td>KNVM / NVMM</td>
<td>Arnhem, NL</td>
<td>2016</td>
</tr>
<tr>
<td>EAACI - Annual congress (poster pitch presentation)</td>
<td>EAACI</td>
<td>Vienna, AT</td>
<td>2016</td>
</tr>
<tr>
<td>Microbiology Centennial Symposium (poster presentation)</td>
<td>Laboratory of Microbiology, WUR</td>
<td>Wageningen, NL</td>
<td>2017</td>
</tr>
<tr>
<td>Pediatric Allergy and Asthma Meeting (poster pitch presentation)</td>
<td>EAACI</td>
<td>London, UK</td>
<td>2017</td>
</tr>
<tr>
<td>Mini-symposium 'Intestinal Microbiology of Early Life' (oral presentation)</td>
<td>Laboratory of Microbiology, WUR / Nutricia Research</td>
<td>Utrecht, NL</td>
<td>2017</td>
</tr>
</tbody>
</table>

## General courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Institute</th>
<th>Location</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;D presentation skills</td>
<td>Nutricia research / Kenhardt</td>
<td>Wageningen, NL</td>
<td>2010</td>
</tr>
<tr>
<td>Statistics for clinical studies</td>
<td>Nutricia research</td>
<td>Wageningen, NL</td>
<td>2010</td>
</tr>
<tr>
<td>Convince in 15 minutes</td>
<td>Nutricia Research / Agent Majeure</td>
<td>Wageningen, NL</td>
<td>2011</td>
</tr>
</tbody>
</table>
General courses (continued)

Key Message Writing  Nutricia Research / Schuttelaars & Partners  Wageningen, NL, 2011
R&D Project Management foundations  Nutricia Research  Wageningen, NL, 2011
Danone Leadership course  Nutricia Research  Wageningen, NL, 2012
Getting Things Done Workshop  Nutricia Research / Tijdwinst.com  Wageningen, NL, 2012
Powerworkshop  Nutricia Research / Qidos  Utrecht, NL, 2013
Patents Basic Training  Nutricia Research / Nederlands Octrooibureae  Utrecht, NL, 2013
R-workshop  Patrick Schloss (University of Michigan)  Detroit, USA, 2015
The Essentials of Scientific Writing & Presenting  WGS  Wageningen, NL, 2016
Scientific Writing  Nutricia Research / Reinschrift Science Communication  Utrecht, NL, 2018

Optionals

Preparation of PhD project proposal  VLAG  2013
PhD/postdoc meetings  Laboratory of Microbiology, WUR  2013-17
Early Life Microbiology team meetings  Laboratory of Microbiology, WUR  2013-17
Gut Microbiology and Probiotics team meetings  Nutricia Research  2013-17
The research described in this thesis was financially supported by Danone Nutricia Research.
Printing of this thesis was financially supported by Danone Nutricia Research.

Cover design by Rene Seure | Kreatuur

Printed by GVO drukkers & vormgevers B.V.
Maat van de rug = zie rekenvoorbeeld hieronder
Maat van voorkant + achterkan +rug
De totale omvang van de PDF is de maat van de omslag samen

ACHTERKANT OMSLAG  170 X 240 mm
VOORKANT OMSLAG  170 X 240 mm

RUG  (de titel moet met de bovenkant naar rechts staan en moet in het midden staan)

AFLOOP:  3 mm aan iedere kant, dit wordt weggesneden wanneer de boeken op maat gesneden worden, hier geen belangrijke dingen neerzetten.

Berekening voor de rug:
Deel het aantal pagina's in het boek door:
19  (voor 100 grams G-Print papier)
15,5  (voor 90 grams ongebleekt Biotop papier)

Bijvoorbeeld, als het boek 208 pagina's bevat en je wilt het 100 grams papier gebruiken, dan deel je 208 door 19 = 10,947 mm
Je mag dit afronden naar 11 mm

NB:
Als de boeken genaaid gebonden moeten worden moet 1mm hierbij optellen (in het voorbeeld krijg je dan een rug van 12 mm).

= de platril: dit is waar de omslag open vouwt, je ziet een lijn waar de 'ril' loopt.