Development of *Bifidobacterium* spp. in infants

Age dependent patterns and correlating factors

Ekaterina Avershina

Høgskolen i **Hedmark**

Master's Thesis in Applied Biotechnology

HØGSKOLEN I HEDMARK

2011

Acknowledgement

The project was performed at the Hedmark University College from August 2010 until May 2011. First of all, I would like to thank my supervisor, Prof. Knut Rudi, for giving me an opportunity to work on the project, for his guidance through the work, his dedication to the project, his valuable advice and his endless enthusiasm. I would also like to thank my second supervisor, Associate Prof. Rob Wilson, for his helping hand with laboratory work and for being always open for discussions.

I would like to give my regards to Norwegian University of Science and Technology NTNU, Trondheim, and particularly to Ola Størro, Roar Johnsen and Torbjørn Øyen, for their kindly provided clinical material and information.

Further, I want to thank Monika Sekelja from Genetic Analysis AS for her help with optimization of the alignment code and Signe Marie Drømtorp from Nofima for the help with troubleshooting. I also thank Rolf Erik Hove, a co-founder of Biometric Games, for his idea on how to develop a basecaller script.

I would also like to thank all the teachers of Master's Biotech 2009-2011 class. We learned a lot during these two years and we couldn't have accomplished that without your support and encouragement.

Finally, I would like to thank all my class mates for the unforgettable time we spent together.

Tusen hjertelig takk til alle!

Table of contents

Appendices

Appendix A – Analysis of 16S rRNA mixed amplicons of IMPACT data

Appendix B – Use of PLS regression for the MCR-ALS prediction

Appendix C – Nucleotide variation between *clpC* amplicons of *B. longum longum* and *B. longum infantis*

Appendix D – MCR-ALS resolution of *clpC* amplicons with 6 components

Appendix E – Five components MCR-ALS resolved spectra of *clpC* amplicons

Appendix F – Identification of *B. longum* group clusters

Appendix G – *B. longum longum* and *B. longum infantis* competition test

Appendix H –Comparison of *Bifidobacterium* species in mother/newborn pairs

Appendix I – Cross-correlation contingency tables of tested factors

Appendix J – One time point calculations – environmental factors

Appendix K – Time change calculations – environmental factors

Appendix L – One time point calculations – IgE and eczema

Appendix M – Time change calculations – IgE and eczema

Appendix N – Basecaller script for the work in a MATLAB® environment

Abbreviations

bp – basepair(s);

clpC – caseinolytic protease C;

CLA – conjugated linoleic acid;

COW – correlation optimized warping;

EFA – evolving factor analysis;

GIT – gastrointestinal tract;

HGT – horizontal gene transfer;

HMO – human milk oligosaccharides;

IgE (IgA) – immunoglobulin E (A);

IL – interleucine;

LSA – local similarity analysis;

MCR-ALS – multivariate curve resolution with alternating least squares;

MLST – multilocus sequence typing;

PCA – principal component analysis;

PCR – polymerase chain reaction;

RMSEP – root mean square error of prediction;

RT – room temperature;

SCFA – short chain fatty acids;

SEM – standard error of the mean;

TGF – transforming growth factor;

TNF – tumor necrosis factor.

Abstract

Bifidobacteria are a major microbial component of infant gut microbiota which is believed to promote health benefits for the host and stimulate the maturation of immune system. Despite the importance we know little about the natural development of bifidobacteria in the infant gut. To address this question, we analyzed mixed *Bifidobacterium clpC* gene sequences from IMPACT (Immunology and Microbiology in Prevention of Allergy among Children in Trondheim) stool samples of 83 infants and their mothers using the novel multivariate statistical MCR-ALS approach. We also developed a novel basecaller script in MATLAB® environment which simplifies the use of the MCR-ALS method so that manual decoding is no longer required. Faecal material was sampled during the first and the second part of pregnancy, at 3 days, 10 days, 4 months, 1 year and 2 years after birth. Five dominant *Bifidobacterium* species were identified and verified by cloning. Stool samples were predicted to be rich in *B. adolescentis*, *B. bifidum*, *B. dentium*, *B. breve* and *B. longum* species. The *B. longum* group consisted of *B. longum longum* irrespective of age, however *B. longum infantis* was mostly identified in four-month-old individuals indicating a potential infant-to-infant nature of its transmittance. Local Similarity Analysis revealed self-supportive correlations between *Bifidobacterium*, *Lactobacillales*, *Clostridium*, *Proteobacteria* and other bacterial groups. The amount of *B. breve* in infants correlated with the incidence of common cold infections during pregnancy. Development of all *Bifidobacterium* and *B. longum* groups had correlation to the occurrence of vaginal fungal infection during pregnancy. Elevated amounts of *B. adolescentis* in four-month-old infants correlated to high IgE levels and eczema symptoms, while levels of the *B. longum* group in infants – to eczema. In conclusion, new patterns in bifidobacterial development and interaction are described. We believe this new knowledge will be important for the future understanding of bifidobacteria in health and disease.

Keywords: *Bifidobacterium*, *clpC*, mixed sequence, MCR-ALS, interaction, cold, fungi infection, IgE, eczema

1 Introduction

1.1 Gut microbiota

Our gastro intestinal tract (GIT) is a habitat for more than 10^{14} microbial cells, which outnumber human cells in the entire body by at least one order of magnitude. Humans are born germ-free, but right from the delivery and soon thereafter our guts become densely colonized by various bacteria. The colonization pattern is influenced greatly by various factors including manner of delivery, infant diet, antibiotic use and hygiene (Wall R. et al., 2009). During early days of life, infant stool samples are rich in facultative anaerobic *Proteobacteria*. Later on, when oxygen levels decrease, anaerobic species belonging to *Bacteroides, Clostridium* and *Bifidobacterium* colonize the gut (Marques et al., 2010). The bacterial composition in the gut becomes similar to that of adults by the age of one year, and then it remains rather stable (Palmer C. et al., 2007). Usually, adult microbiota is dominated by *Firmicutes* and *Bacteroidetes* which make up to 98 % of all 16S rRNA sequences isolated from intestinal samples. The other phyla associated with the human gut are *Proteobacteria, Actinobacteria, Fusobacteria, Verrucomicrobia* and *Cyanobacteria* (Ley R. et al., 2006). It is still not certain how many species reside in our guts and the suggested numbers range from 500 (Hattori M., Taylor T., 2009) up to 12 000 – 36 000 different species (Brugere J. et al., 2009). Such a discrepancy results from the fact that most of the intestinal microbes are strict anaerobes which have never been cultured – even with new culture-independent methods in hand, more than 80 % of the intestinal microbiota has not yet been taxonomically characterized (Fliss I. et al., 2010). All together, intestinal microbiota constitutes approximately 2 % of our biomass and its collective genome has 100 times more genes than the human genome (Brugere J. et al., 2009). The majority of human gut microbiota consists of bacterial cells; however, some fungi and archaea, mostly represented by *Methanobrevibacter smithii*, are also found in the gut (Eckburg P. et al., 2005). Generally, on a species level, microbial communities within our colons are as unique as a fingerprint (Dethlefsen L. et al., 2007), even though possibly being influenced by host genetics to a certain extent. For example, the composition of intestinal species between monozygotic twins who had not been living together for more than 5 years was found closer to each other than to their respective marital partners, but still unique for each twin (Zoetendal E. et al., 2001).

Despite the huge individual variation of intestinal species, gut microbiota seems to be more stable functionally (Stecher B. et al., 2010). At the same time, it is worth noting, that on a phylum level, the composition of human gut microbiota is also similar to that of other mammals, but most bacterial families and genera are distinct (Dethlefsen L. et al., 2007). These facts suggest co-evolution of humans with their microbial gut residents. In line with this, Sekelja et al. (2010) have discovered two core phylogroups shared among 210 individuals with various ethnic background and diet habits.

Recently, the occurrence of immune-mediated disorders such as inflammatory bowel disease (IBD), asthma, and type I diabetes, has increased so rapidly in the Western world that solely the genetic basis of these diseases is unlikely to be the only reason (Blaser M., Falkow S., 2009). Instead, the altered composition of intestinal microbiota, and thus host-microbiota interactions, due to various antimicrobial strategies might also be an explanation of this phenomenon. Several correlations between the microbial composition and the disease state have been reported so far, however it is not certain whether the reported dysbiosis is the cause or, conversely, the consequence of a disease (Round J., Mazmanian S., 2009). Certainly, in many aspects of our physiology and well-being we are truly dependant on these intestinal residents and therefore nowadays it is becoming more and more common to regard intestine - microbiota relationships as symbiotic, and not commensal as thought previously. In various studies of gnotobiotic animals, it has been demonstrated that germ-free animals show defects in the development and maturation of a range of gut-associated lymphoid tissues, which, however, could be restored shortly after introduction of gut bacteria (Shanahan F., 2009). For example, germ-free mice, infected with enteric parasite *Encephalitozoon cuniculi*, and treated with DNA isolated from intestinal microbiota, have shown upregulation of pro-inflammatory T_H17 and T_H1 cells response which resulted in a better resistance to the parasite (Round J., Mazmanian S., 2009). At the same time, enteric bacteria can not only stimulate, but also suppress the immune response – for instance *Bacteroides fragilis* produces polysaccharide A (PSA), a molecule which stimulates the production of anti-inflammatory T_{reg} cells (Shanahan F., 2009). Intestinal microbiota has also been shown to produce short chain fatty acids (SCFA), conjugated linoleic acid (CLA), essential amino acids, vitamins B and K, all contributing to the well-being of a host (Marques T. et al., 2010).

1.2 *Bifidobacterium* species in the human gut

1.2.1 Description of *Bifidobacterium*

Bifidobacteria are Gram-positive branched rod-shaped bacteria belonging to the phylum of *Actinobacteria*. They are non-spore forming non-motile anaerobic microorganisms producing acid but not gas from various carbohydrates (Tannock G., 2010). The species of this genus was first discovered in 1899 by Tissier and was referred to as *Bacillus bifidus*, but later was classified as *Bifidobacterium bifidum*. For their resemblance with lactobacilli in many morphological and physiological aspects, at first they were collectively referred to as *Lactobacillus bifidus*. In 1960s, the 'bifid shunt', a specific *Bifidobacterium* hexose fermentation pathway with help of a fructose-6-phosphate phosphoketolase (F6PK), was discovered (de Vries W., Stouthamer A., 1967). Only after this discovery, the genus was separated from lactobacilli. Thirty-two species of *Bifidobacterium* have been discovered so far, including *B. tsurumiense* and *B. mongoliense*, identified in 2008 and 2009, respectively. The species were isolated from human intestines, oral cavities, blood, food, animal and insect intestines, and sewage (Ventura M. et al., 2010a). As assessed by comparative genomic analysis of conserved genes including 16S rRNA sequences, the genus comprises 6 groups with the *B. asteroides* group being the deepest branch. Out of 32 species, complete genomic sequences of 9 bifidobacterial species and partially assembled genomes of 17 species have been generated to date. In general, bifidobacterial genomes show strong synteny both between and within the species. However, some species, e.g. *B. longum*, possess high numbers of unique genes (Ventura M. et al., 2010b).

Bifidobacteria are prototrophic organisms capable of synthesis of at least 19 amino acids, pyrimidine and purine nucleotides, nicotinate, thiamine and folate. So far, 696 orthologous proteins forming the core genome of *Bifidobacterium* species have been identified on the 2.0 – 2.8 Mb long chromosome with a high G+C content (Ventura M. et al., 2010b). All together, about 10 % of bifidobacterial proteome comprises enzymes involved in carbohydrate transport and metabolism, which is almost 30 % higher than that of other enteric bacteria such as *E. coli*, *Enterococcus faecium* and *Lactobacillus* spp. *Bifidobacterium* possess a rich set of internal glycosyl hydrolases (GH) which are often linked and clustered with genes encoding ATP-binding cassette (ABC) transporters, permeases and proton symporters used for the binding and transporting of complex sugars into the cell. It has been recently estimated that

carbohydrate transport comprises about 5 % of the bifidobacterial genome and most of the genes involved in sugar metabolism have been duplicated (Ventura M. et al., 2009). The level of horizontal gene transfer (HGT), which is responsible for the bacterial adaptation and competitiveness in their ecological niche, varies greatly among bifidobacteria. For example, less than 5 % and 1 % of chromosome sequences of *B. longum longum* NCC2705 and *B. dentium* Bd1 respectively have been acquired by HGT (Ventura M. et al., 2010b). Bifidobacterial GHs are predicted to hydrolyze starch, fructooligosaccharides (FOS), galactooligosaccharides (GOS), arabinogalactans, arabinoxylans and human milk oligosaccharides (HMO). The latter is especially evident among *B. longum infantis* which is the only *Bifidobacterium* species growing robustly on purified HMOs. LoCascio et al. (2010) have recently shown that *B. longum infantis* possesses conserved HMO utilization gene regions as opposed to the closest related species, *B. longum longum*.

Nine *Bifidobacterium* species are commonly associated with the human GIT – *B. adolescentis*, *B. breve*, *B. longum* subsp. *longum* and *B. longum* subsp. *infantis*, *B. pseudolongum*, *B. bifidum*, *B. pseudocatenulanum*, *B. dentium*, *B. animalis* subsp. *lactis*. The recent study of stool and intestinal biopsy samples have revealed that the latter two species are mostly found in stool, but not mucosal samples, suggesting that these represent allochthonous, i.e. transient, microbiota and are not mucosa-adherent species (Turroni F. et al., 2009). *B. dentium* is often found in oral cavities and is believed to be associated with dental caries, whereas *B. animalis lactis* has a dairy origin and is normally associated with food (Ventura M. et al., 2010b). It is generally believed that *Bifidobacterium* composes around 2 % of the total intestinal microbiota in adults, whereas at early days of life this genus makes up the majority of intestinal bacteria and reaches up to 90 % (Tannock G., 2010). However**,** some scientists claim the relative numbers in adults are higher, and that bifidobacteria constitute up to 15 % of the total gut microbiota (Menard O. et al., 2008), whereas others suggest much lower abundances of these species among infants (Palmer C. et al., 2007). Some bifidobacterial strains, e.g. *B. pseudolongum*, are exclusively associated with adult gut microbiota, and some, especially *B. longum infantis*, are typically isolated from infants. At the same time, other bifidobacterial strains, such as *B. adolescentis*, *B. longum longum*, *B. breve* and *B. pseudocatenulanum*, seem to be present in our guts irrespective of the age (Ventura M. et al., 2010a).

1.2.2 Role of *Bifidobacterium* in the human infant gut

High abundance of *Bifidobacterium* species in infants is believed to have a promoting effect on sustaining health of an individual as well as on the development and maturation of the immune system. Bifidobacteria can stimulate the immune system with a host-health dependent pattern of stimulation – they are capable of both strengthening immune responses in healthy individuals and down-regulating pro-inflammatory responses in pathological conditions such as inflammatory bowel disease (Rossi M., Amaretti A., 2010). *Bifidobacterium* strains suppress production of T_H2 cytokine and IgE, an abnormal response of which is regarded to be characteristic of an allergic reaction (Miyauchi E. et al., 2010). However, the immunomodulatory ability of *Bifidobacterium* species appears to be strain dependent – *B. bifidum, B. dentium* and four *B. longum* strains induce T_H1 and/or T_H2 cytokines, whereas other have a low (*B. breve*, three strains*)* or no effect (*B. adolescentis)* on the immunity of gnotobiotic mice (Menard O. et al., 2008). Even though *B. adolescentis* seemed to have no effect on the immune system in the latter study, this species is often found in higher counts among allergic children (Ouwehand A. et al, 2001; Stsepetova J. et al., 2007). *B. longum infantis* has been shown to inhibit production of pro-inflammatory IL-17 by induction of anti-inflammatory IL-10 in splenocyte and colon cultures (Miyauchi E. et al., 2010). In the other study, oral administration of *B. lactis* Bd12 increased levels of faecal IgA in preterm infants 44 % compared to placebo group (Mohan R. et al., 2008). In the same study, levels of calprotectin, a protein elevated concentrations of which in preterm infants is associated with higher risk of necrotizing enterocolitis, was found lower in a probiotic group.

By producing lactic and acetic acids in even larger amounts than lactobacilli and lactococci, bifidobacteria lower the pH in the colon and thus help to prevent colonization of the gut by pathogens. For example, it has been shown that *B. longum infantis*, isolated from stool samples of infants resistant to diarrheal infections, inhibited growth of *C. perfigens* and *E. coli* (Fliss I. et al., 2010). At the same time, the antimicrobial effect of *Bifidobacterium* species is likely to be caused by production of organic acids rather than by solely lowering pH. When lactic and acetic acid were used for lowering pH, the growth inhibition of *E. coli* and *S. typhimurium* was higher than with hydrochloric acid, indicating that there are other mechanisms involved (Makras L., de Vuyst L., 2006). Another antimicrobial ability of bifidobacteria is connected to the production of antimicrobial peptides: bacteriocins. The species of this genus are generally believed to produce true bacteriocins and bacteriocin-like

inhibitory substances (Fliss I. et al., 2010). There are many reports on substances either inhibiting some pathogens or preventing their adhesion to epithelial cells, but to date, only one bacteriocin, namely bifidocin, has been isolated and characterized (Yildirim Z., Johnson M., 1998).

The ability to adhere to intestinal cells is another important factor for the ability to persist in the colon and to compete with enteric pathogens. Bifidobacteria have proven themselves to be very effective in inhibiting the adhesion of number of enteric pathogens to Caco-2 and HT-29 adenocarcinoma-based cell lines in numerous studies (Fliss I. et al., 2010). Four strains of *B. bifidum* have been shown to be capable of displacing such pathogens as *E. coli*, *C. difficile*, *Ent. sakazakii*, *S. typhimurium* and *L. monocytogenes* attached to the mucosal epithelium. The displacement levels ranged from 15 % to 70 % depending on the strain and the pathogen tested (Guiemonde M. et al., 2007). Though the mechanism remains unknown, the authors hypothesize that various factors, some of which could be thermolabile in nature, are involved, as both treatments with proteinase K, lipase, metaperiodate and thermotreatment have reduced the adhesion of tested strains.

Mammals lack the ability to synthesize certain vitamins, and therefore these have to be obtained from other sources like diet or gut microbiota. *Bifidobacterium* species are generally believed to produce vitamins of the group B including folate, cobalamin, pyridoxine, riboflavin and thiamin. In 2007, 76 strains of 15 *Bifidobacterium* species of human and animal origin were tested for their ability to produce folate, – water-soluble vitamin B9. In total, 17 strains belonging to 9 species were capable of folate production, with two strains of *B. adolescentis* and one strain of *B. pseudocatenulanum* being the top-three producers (Pompei A. et al., 2007). In the other study, 24 strains belonging to species of *B. bifidum*, *B. breve*, *B. longum longum*, *B. longum infantis* and *B. adolescentis* were tested for their ability to produce B vitamins (Deguchi Y. et al., 1985). All tested strains were able to produce pyridoxine (vitamin B6) and small amounts of cobalamin (vitamin B12), which is very important in early days of life as it is responsible for the normal growth and development of the nervous system (Marques T. et al., 2010). Production of thiamin (vitamin B1) and nicotinic acid (vitamin B3) seems to be strain-dependant as all tested *B. bifidum*, *B. longum infantis*, and some of *B. breve* strains were high producers, while *B. adolescentis* and some strains of *B. longum longum* and *B. breve* lacked this ability (Deguchi Y. et al., 1985). Despite this ability, however, it remains unclear whether the amount of vitamins produced by

bifidobacteria makes any contribution to the vitamin status of the host (Rossi M., Amaretti A., 2010).

Certain *Bifidobacterium* strains, especially *B. breve* and *B. dentium*, are also capable of transformation of linoleic acid into bioactive conjugated linoleic acid (CLA) – the component believed to have anti-carcinogenic, anti-diabetes, anti-obesity and anti-allergy functions. Same as with vitamins, there is little information available on whether or not bifidobacteria can make a contribution to the levels of CLA in the host (Rossi M., Amaretti A., 2010). However, some scientists ascribe the change in the fatty acid composition of the liver and adipose tissue to the consumption of certain bifidobacteria strains (Shanahan F., 2009).

1.2.3 *Bifidobacterium* spp. identification and quantification

Culturing is the conventional method of *Bifidobacterium* species identification and enumeration and many selective media have been developed for that purpose (Ferraris L. et al., 2010). However, the recovery of different bifidobacterial species varies and every medium may favor some, while inhibit the growth of other, *Bifidobacterium* species (Roy D., 2001). Therefore molecular methods are widely used for the identification and quantification purposes. Quantitative real time PCR (qRT-PCR) targeting 16S rRNA gene is an example of a molecular approach, which is commonly used for *Bifidobacterium* quantification (Haarman M., Knol J., 2005). However, the number of 16S rRNA gene copies per bifidobacterial chromosome may vary from 1 to 5 depending on the strain (Candela M. et al., 2004), which may lead to over- or underestimation of some species. Furthermore, 16S rRNA provides relative low resolution for strain discrimination due to the conserved nature of this gene (Ventura M. et al., 2006).

Multilocus sequence typing (MLST) is a molecular approach of species delineation by comparing the sequences of several housekeeping genes (Maiden M. et al., 1998). MLST approach is widely used for various bacteria and the databases with sequence profiles are publically available (Jolley K., 2011). Deletoile et al. (2010) used seven genes for MLST of 119 strains belonging to four *Bifidobacterium* species. This resulted in identification of 104 sequence types (ST) based on a combination of allelles, which demonstrates high discriminatory power of the method.

The gene encoding caseinolytic protease C (*clpC*) is one of the seven genes used for MLST of *Bifidobacterium* species and it has 34 identified alleles. Caseinolytic protease C (ClpC) belongs to a heat stress resistance proteins of the Clp ATPase family, and is believed to be involved in recognition of proteins for the degradation by ClpP caseinolytic peptidase (Sanchez B. et al., 2010). The *clpC* gene was first proposed as a molecular marker for *Bifidobacterium* delineation in 2005 (Ventura M. et al., 2005). Among 32 type strains of *Bifidobacterium* species tested, *clpC* gene was successfully amplified in all strains apart from *B.minimum* LMG 11592^{T}, though in the latter species *clpC* was detected by hybridization (Ventura M. et al., 2006). The *clpC* gene allows better phylogenetic resolution of bifidobacterial taxa compared to 16S rRNA – mean sequence similarities of 16S rRNA sequences are higher and mean pairwise distances are lower than those of *clpC* sequences. For example, 16S rRNA sequences of *B. animalis animalis* and *B. animalis lactis*, and of *B. longum longum*, *B. longum infantis* and *B. longum suis* have more than 99 % sequence similarity making it impossible to discriminate them, whereas delineation of these species by means of *clpC* sequences is supported by high bootstrap values (Ventura M. et al., 2006). At the same time, in *Bifidobacterium*, the *clpC* gene is a single-copy gene (Ventura M. et al., 2005), making it more suitable for the relative quantification of *Bifidobacterium* species as opposed to the 16S rRNA gene.

In addition to its conservation, this gene is also unlikely to be affected by HGT (Ventura M. et al., 2006). However, in the other study, where 119 strains belonging to four *Bifidobacterium* species were tested, the *clpC* gene of *B. longum infantis* CIP6467 was identical to that of *B. bifidum* strains and the *clpC* gene of *B. bifidum* B526 – to *B. longum longum* isolates, indicating the possibility of HGT (Deletoile A. et al., 2010).

1.3 Application of mixed sequences to analyze bacterial diversity

Trosvik et al. (2007) used mixed sequences for the quantitative analysis of the microbial population. They showed that multivariate regression analysis of mixed sequences allows direct quantification of genetically coherent units (GCUs) in the mixture of bacterial species. However, with the statistical approach they used, the prior knowledge of the composition, i.e. what GCUs are present in the system, was required.

Multivariate curve resolution analysis with alternating least squares (MCR-ALS) is a statistical tool which allows the recovery of pure components from various mixtures when no prior knowledge about the nature and composition of these mixtures is available (Tauler R. et al., 1995). It is extensively used for resolution of spectral, pH, time and elution profiles of diverse environmental samples (Tauler R. et al., 2010) and recently it has been proposed as a promising tool for the analysis of image data from DNA microarrays (Jaumot J. et al., 2006). However, to our knowledge, apart from the study of Sekelja M., MCR-ALS has never been applied for the analysis of mixed DNA sequence spectra (Zimonja M. et al., 2008). The approach was tested on two datasets of predefined mixtures of 16S rRNA amplicons from three (*Bacteroides*, *E. coli* and *Faecalibacterium*) and seven (*Bacteroides*, *E. coli*, *Faecalibacterium*, *Clostridium*, *Eubacterium*, *Lactobacillus* and *Ruminococcus*) bacteria respectively. All in all, for both datasets tested, MCR-ALS proved itself to be a suitable tool for resolution, identification and relative quantification of mixed 16S rRNA amplicons (Zimonja M. et al., 2008).

Mathematically, the MCR-ALS method presents the initial experimental matrix of spectral data D (m×n) as the combination of concentration matrix C (m×i) and matrix of pure components S (i×n), and a residual term E (m×n):

$$D = CS + E \qquad \text{(Eq. 1.1)}$$

Prior to the MCR-ALS analysis, only the initial matrix D is known. However, the initial number of components (initial estimates i) should also be defined (Tauler R. et al., 1995). To do so, progressive or iterative local rank methods like Principal Component Analysis (PCA) or Evolving Factor Analysis (EFA) are used. PCA transforms the initial data into another coordinate system and defines the number of significant components in the system depending on how much of the whole system's variance the component explains (Smith L., 2002). EFA on the other hand, repeatedly applies PCA to sections of the dataset starting from the first two samples and subsequently adding the next one every run of the PCA. EFA can be performed in two ways – forward or backward analysis depending on the direction of moving the window, i.e. sub-datasets of mixtures (Garrido M. et al., 2008).

A simplified explanation of the MCR-ALS analysis is presented in Figure 1.1. Imagine the hypothetical dataset of mixed DNA spectra consists of two sequences. PCA and EFA predicted 3 components to be present in the given dataset. Performing MCR-ALS with a

number of components to be resolved set to 3, will result in two separate matrices, one of which will contain the information on the spectral data of these three components (S), and the other – the concentration profiles of each of the resolved components in every mixed DNA sample (C).

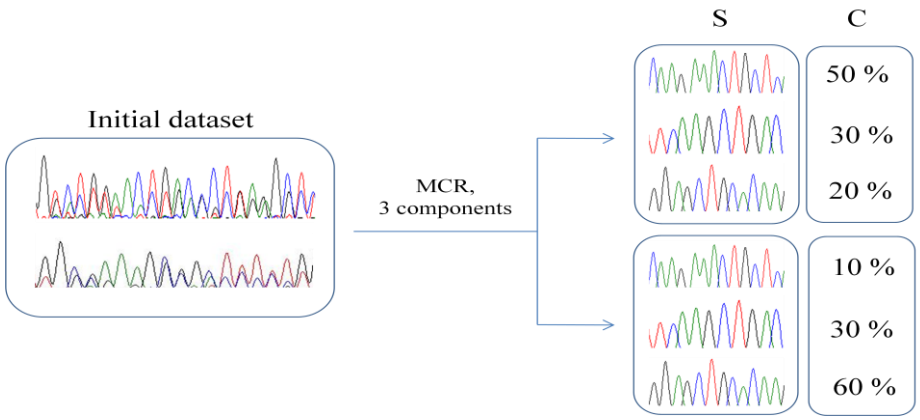

Figure 1.1 Simplified scheme of MCR analysis. The number of components to resolve is defined from PCA/EFA analysis. As a result, pure spectra and their relative amount in each of mixed spectra are defined.

1.4 Aims of the project

Despite the growing interest in *Bifidobacterium* spp., little is known about the development of these bacteria in infant guts and about the correlation of its composition in the infant gut to various environmental factors. The aim of this project was therefore to determine the factors which correlate to the development of *Bifidobacterium* species in infants. The specific goals to reach the overall aim were:

- To identify the dominant *Bifidobacterium* species and their composition at different ages relative to both the total bacterial load and to *Bifidobacterium* spp.;
- To identify *B. longum* group species composition at different ages;
- To reveal the interactions between *Bifidobacterium* species and other bacterial groups;
- To determine mother- and infant- related factors correlating to the composition of *Bifidobacterium* species;
- To determine the difference in the composition and development of *Bifidobacterium* species in healthy infants and those who suffered from allergy and/or eczema.

To reach these goals we used a data generation approach involving MCR-ALS analysis of mixed *Bifidobacterium clpC* gene sequences amplified from stool samples of mothers and their infants. Data analyses were based on non-parametric tests of statistical significance, network analysis and presence/absence test.

1.5 Frame of the project

This project was conducted using stool specimens received from the IMPACT study. The Prevention of Allergy among Children in Trondheim (PACT), a controlled non-randomized intervention study of pregnant women and their children up to two years of age, was started in 2000. (Størro O. et al., 2010). At the same time, a randomized nested-cohort study (Immunology and Microbiology in Prevention of Allergy among Children in Trondheim – IMPACT) of 720 pregnant women and their offsprings, recruited from PACT control cohort, was also initiated. The aim of the IMPACT study was to analyze the development of intestinal microbiota in infants during the first years of their life depending on various environmental factors, as well as its influence on the maturation of immune system and establishment of asthma (Øyen T. et al., 2006).

Stool samples from the IMPACT study were received from the Norwegian University of Science and Technology (NTNU) and DNA isolation was performed by de Muinck (de Muinck E. et al., 2011). PCR amplification of *clpC* gene was completed as a part of a Bachelor's degree project (Glazowska S, Czebreszuk U, 2010). Information on total bacterial composition was received from MCR-ALS based analysis of 16S rRNA amplicons (Appendix A). The MATLAB® code for pre-analytical alignment and correction of mixed spectra was designed by Sekelja M. (Zimonja M. et al., 2008).

In this thesis, mixed *clpC* gene amplicons of the DNA isolated from IMPACT stool samples were sequenced and then analyzed using a novel MCR-ALS approach. As a result, the dominant *Bifidobacterium* species were identified. To verify the dominant species, *clpC* gene amplicons of selected samples were cloned into *E. coli,* PCR amplified and sequenced. The percentage of *Bifidobacterium* species, relative to both total *Bifidobacterium* and total bacterial load (identified with the same approach on 16S rRNA gene amplicons; Appendix A) was compared to various clinical and environmental factors to check for the correlations. Local Similarity Analysis was performed to reveal the correlations between the abundances of

bifidobacterial species. The *B. longum* group was resolved by sequencing with a specific *B. longum*-group primer. A growth competition test was performed on *B. longum longum* DSM20219 and *B. longum infantis* DSM20088, grown on semi-synthetic media with a single carbon source. Isolation of *B. longum longum* and *B. longum infantis* from stool samples, where they were predicted to be abundant, was attempted. A flow diagram showing the processes performed in this project is presented in Figure 1.2.

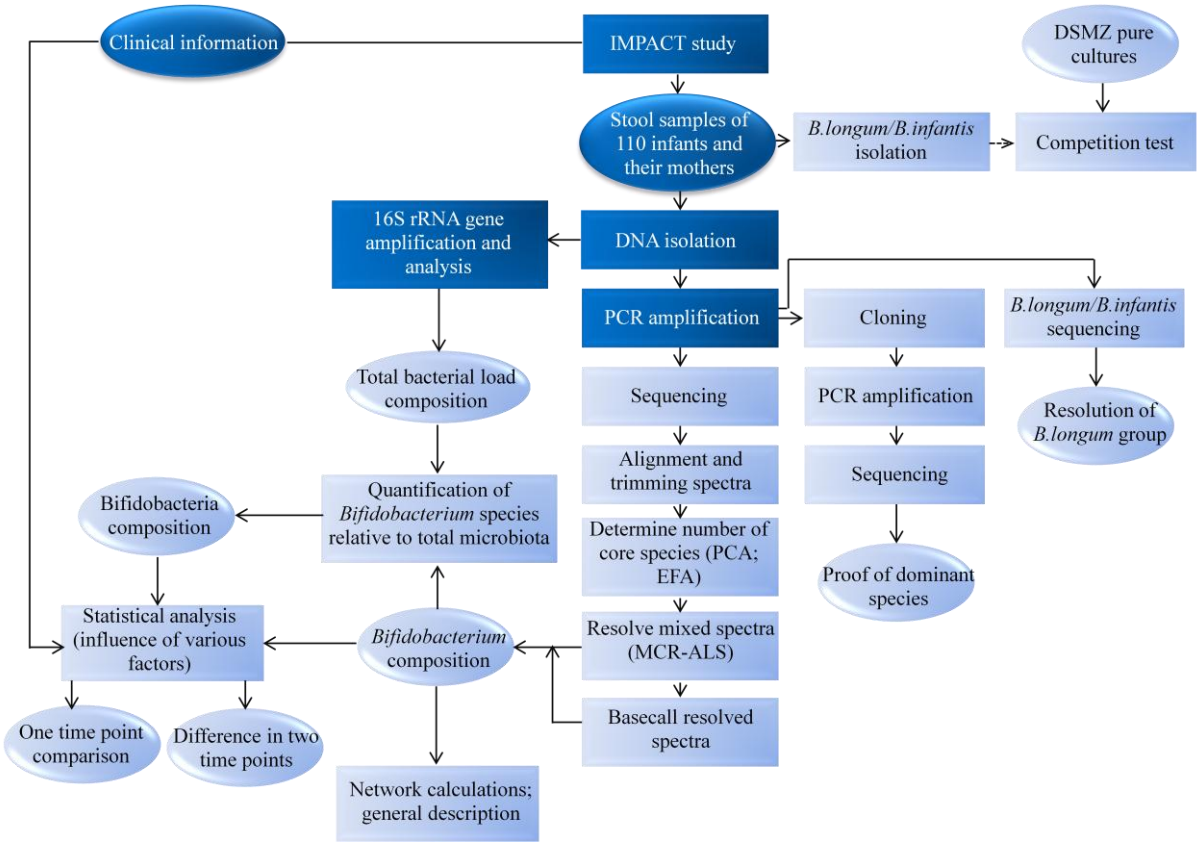

Figure 1.2 Flow diagram of the study. Boxes and ellipses represent processes and data or materials obtained from them respectively. Dark background shows work done prior to the start of the project.

2 Materials and methods

2.1 Clinical material and information

Stool samples from 110 children and their mothers, as well as the information on their diet, allergy status and diseases, were received from the IMPACT study (Øyen T. et al., 2006). Mothers' stool samples were collected at two points during pregnancy (early period – from first to the middle of the second trimester; late period – third trimester) and infant stool samples – at age 2-4 and 10 days; 4 months; 1 year and 2 years after birth. The samples (2 g) were diluted in 10 ml of sterile Cary Blair medium (BD Diagnostics, USA) and frozen at -20 °C no later than 2 hours after the collection. Within a month, samples were transferred and stored at -80 °C (Øyen T. et al., 2006).

Information on the lifestyles of pregnant women and their children was gathered through a questionnaire filled out during pregnancy, at 6 weeks, 1 year and 2 years after birth (Størro O. et al., 2010). Information on the mothers' style of living during pregnancy included pets at home; smoking; whether the mother experienced having a cold or fungal infection; whether mothers ate fatty fish and/or took fish oil; and whether they ate vegetables more than once and/or more than four times a week. Information on the infants consisted of data on their lifestyle during the first year of life (pets at home; smoking parents; whether they were given multivitamins during first 6 weeks and 1 year of life; and whether he/she was breast-fed throughout the first year of life), his/her allergy status (total specific IgE levels in blood serum more than 0.1; 0.35 and 0.7 kU/ml) and eczema status by the age of 2 years. The level of serum IgE was measured for such allergens as mites, mold, cat and dog dander, birch, timothy (grass) and mugwort pollens, cow's milk, hen's egg white, codfish, hazelnut and peanut (Størro et al., 2011). Eczema was assessed using both the UK Working Party (UKWP) diagnostic criteria for atopic dermatitis and mothers' answers to a questionnaire (whether the baby was experiencing eczema symptoms).

2.2 DNA isolation

2.2.1 DNA isolation from freeze-dried strains

Pure strains of *B. longum longum* (DSM20219), *B. longum infantis* (DSM20088), *B. breve* (DSM20213) and *B. adolescentis* (DSM20083) were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Germany). After the glass ampule with freeze-dried culture was opened according to recommendations from the manufacturer, the pellet was resuspended in 500 µl of 'solution A' [25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0; 0.22 µm sterile-filtered] for 30 min. The sample was then transferred to a tube with 0.25 g acid-washed glass-beads ≤ 106 µm (Sigma-Aldrich, Germany), boiled for 5 minutes and cooled on ice before cell homogenization. Cells were homogenized in MagNA Lyser (Roche, Switzerland) at 6500 rpm for 25 seconds 4 times (to prevent DNA degradation due to overheating, tubes were cooled down for 1 minute between each step). Proteinase K (25 µl; 100 mg/ml) was added to 180 µl of lysate and incubated at 56 °C for 30 min. After incubation, ethanol (96 %, 200 µl) was added to the tube and DNA isolation was proceeded as described in 'Purification of Total DNA from Animal Tissues (Spin-Column Protocol)' for Qiagen Stool kit starting from step 4 (Qiagen, USA).

2.2.2 DNA isolation from bacterial cultures

Bacterial cultures were aliquoted into Eppendorf tubes (1 ml), centrifuged at 13 000 rpm (Biofuge 13) for 5 min at RT and the cell pellet was resuspended in TE Buffer [10 mM Tris-HCl pH 8.0; 1mM EDTA] (100 µl). Then lysozyme (10 µl; 10 mg/ml) was added and the tubes were incubated for 30 min at RT. After the lysis, binding buffer [500 µl; 7.5 M Guanidinium thiocyanate; 150 mM Tris-HCl pH 6.8; 7.5 mM EDTA] and 10 µl of SiMAG/MP-DNA Magnetic Beads (Chemicell GmbH, Germany) were added into the tubes which subsequently were incubated at RT for 5 min. The tube was then placed on the magnetic separator for 30 sec and the bead/DNA pellet was retrieved and treated with washing solution I [1 ml; 60 % (w/v) Guanidinium hydrochloride; 1 % Triton X100; 10 % TE Buffer pH 8.0], 70 % ethanol (1 ml) and ddH_2O (1 ml). After each washing step the tube was placed into the magnetic separator and the pellet was collected. Then elution buffer [100 µl; 10 mM Tris-HCl pH 8.0; 1 mM EDTA] was added to the tubes and samples were incubated in a thermomixer (Eppendorf, Germany) for 10 min at 60 °C with mixing (1000 rpm). Finally, the beads were collected and the solution with eluted DNA was transferred to a new tube. The

concentration of isolated DNA was determined using a NanoDrop spectrophotometer (Thermo Scientific, USA).

2.2.3 DNA isolation from stool samples

Stool samples (1 ml) were first diluted in 1 ml of Solution 1 [50 mM glucose; 25 mM Tris-HCl pH 8.0; 10 mM EDTA pH 8.0]. The resulting suspension was then diluted 1:4 in 4 M guanidinium thiocyanate. The sample (500 µl) was transferred to a sterile FastPrep®-tube (Qbiogene Inc, USA) with 250 mg acid-washed glass beads (≤ 106 µm; Sigma-Aldrich, Germany) and homogenized in FastPrep® Instrument (Qbiogene Inc, USA) for 40 sec. Then 170 µl of the sample together with 10 µl of Silica particles (Merck, Germany) were transferred to 96-wells Greiner U-plate (Greiner Bio-One, Germany) which was then placed in a Biomeks 2000 Workstation (Beckman Coulter, USA). After addition of Sarkosyl (1 %), the plate was incubated at 65 °C for 10 min, followed by 10 min at RT. Then the supernatant was discarded and a bead pellet was washed twice with 50 % ethanol. The beads were suspended in Buffer C [100 µl; 1 mM EDTA pH 8.0, 10 mM Tris–HCl pH 8.0], incubated at 65 °C for 30 min and the solution with eluted DNA was collected.

2.3 PCR amplification

One microliter sample DNA was mixed with a PCR-master mix [1x Dynazyme II HotStart Buffer, 0.08 mM dNTPs, 0.2 µM forward primer, 0.2 µM reverse primer, 0.04 U/µl Dynazyme II HotStart enzyme, in DNase/RNase free water] to a final volume of 25 µl. Concentrations are given according to final concentrations in the PCR mix with added DNA. For each PCR reaction, 30 cycles were used. Initial denaturation was performed at 95 °C for 10 min. Denaturation times at 95 °C and annealing times were set for 30 sec. For the annealing temperature, elongation time and primers used see Table 2.1. Polymerization was finished at 72 ºC for 7 min. Amplicons were checked with 1.5 % Agarose gel (100 V; 30 min). All used oligonucleotides were obtained from Invitrogen (USA).

Table 2.1 Primers used for DNA amplification and sequencing. OligoCalc online calculator (Kibbe W., 2007) was used for melting temperature calculation

Primer name	Sequence of the primer in 5'→3' direction	Direction	T_m, °C	Annealing temp, °C	Elongation time, min:sec
clpC-F	GAGTACCGCAAGTACATCGAG	forward	54.4	60	1:00
clpC-R	CATCCTCATCGTCGAACAGGAAC	reverse	57.1		
clpC-L	AGAAGCTGGAAGCCGAT	forward	52.1	60	–
HU	CGCCAGGGTTTTCCCAGTCACGACG	forward	64.2	65	2:00
HR	GCTTCCGGCTCGTATGTTGTGTGG	reverse	60.8		

2.4 DNA sequencing

To degrade excess single-stranded primers, PCR-products (1 µl) were treated with Exonuclease I (New England Biolabs, UK) [BigDye Sequencing buffer 5X; ExoI 0.4 U/µl; DNase/RNase free water to 5 µl] at 37 °C for 60 min. Enzyme was inactivated at 85 °C for 15 min. ExoI-treated PCR products (1 µl) were sequenced in forward direction [BD Sequencing Buffer 5X; 0.32 µM sequencing primer 1 µl; BigDye Terminator v1.1 1 µl; DNase/RNase free water to 10 µl]. DNA was sequenced for 25 cycles at 96 °C for 15 sec, followed by 60 °C for 4 min. Sequencing reaction was then cleaned up using BigDye XTerminator (ABI, USA) following manufacturer's protocol. Capillary electrophoresis was performed on 3130xl Genetic Analyzer (ABI, USA) with 7 sec injection time. Sequencing was performed in two sets: first, 1 µl of ExoI-treated PCR product was used. After examination, those samples which had low or no signal were re-sequenced with a double amount of PCR product used. In case of still absent/low signal, the samples were removed from the analysis.

The clpC-F primer was used to sequence all *Bifidobacterium* species present in the samples; whereas for *B. longum longum*/*B. longum infantis* separation, clpC-L, a primer, specific for *B. longum* group, was designed (Chapter 3.2.2).

2.5 Cloning

To confirm whether *Bifidobacterium* spp., predicted by MCR-ALS, were truly present in the dataset, *clpC*-amplicons of DNA isolated from 14 stool samples were cloned into pCR® 2.1-TOPO® vector and then used to transform chemically competent TOP10 *E.coli* cells using TOPO TA Cloning® kit (Invitrogen, USA). Due to the fact that for successful cloning, PCR

products should be not more than 1 day old, and due to lack of DNA, the cloning was performed on PCR products of 1/10 diluted PCR products. Amplification was performed according to the protocol described above (Chapter 2.3) using 10 cycles.

Transformed *E.coli* cells were then grown on LB medium agar plates with addition of kanamycin (50 mg/l; 0.2 µm filter-sterilized) at 37 °C for 48 – 72 hours. Colony PCR (10 colonies per sample) with HU/HR primers (Table 2.1) targeting the vector, was then performed on transformants. The sequencing of PCR products was done according to the protocol described above using the clpC-F primer (Chapter 2.4).

2.6 Bacterial isolation and cultivation – competition experiment

2.6.1 Bacterial isolation

Four stool samples which were predicted to be rich in the *B. longum* group (2 in *B. longum longum* and 2 in *B. longum infantis*) were diluted in saline [0.91 % (w/v) NaCl; 0.2 µm filter-sterilized] 10, 100 and 1000 fold. Though higher dilution rates are recommended, normally experiments are performed on fresh stool samples, whereas in case of this study, stool samples were kept at -80 °C for 8 years. Diluted samples were then streaked on plates containing Bereens medium [Columbia Agar base (Merck, Germany) 42.5 g/l; glucose 2.5 g/l; lactulose 2.5 g/l; cysteine HCl 0.5 g/l; riboflavin 0.01 g/l; propionic acid 5.0 ml/l]. The medium was autoclaved at 118 °C for 15 min prior to addition of propionic acid and riboflavin. Plates were then placed in a 3.5 l Anaerobic jar (Oxoid, France) with GENbox anaerobic generator (BioMerieux, France) and kept at 37 °C for 72 hours. Anaerobic conditions were verified using a Microbiology Anaerotest indicator (Merck, Germany). The experiment was repeated using Wilkins-Chalgren for *Bifidobacterium* Mupirocin (Ferraris L. et al., 2010) [Wilkins-Chalgren agar (Merck, Germany) 43 g/l; Tween 80 0.5 %; glucose 10 g/l; cysteine-HCl 0.05 % (w/v); mupirocin 50 mg/l]. Prior to inoculation, plates were kept anaerobically at RT for 24 hours.

After 72 hours incubation, plates were examined and colonies of interest were inoculated into liquid MRS medium (Merck, Germany) supplemented with 0.1 g/l cysteine-HCl and grown anaerobically at 37 °C for 1 day. DNA isolation, PCR and sequencing reactions were performed as described previously (Chapters 2.2.2; 2.3 and 2.4 respectively).

2.6.2 Competition experiment

Pure cultures of *B. longum longum* and *B. longum infantis* were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Germany). Ampules containing a freeze-dried culture of *B. longum longum* (DSM 20219) and *B. longum infantis* (DSM 20088) were opened, 50 µl of Anaerobe Basal broth (Oxoid, France) was added to the inner tube and cells were left to rehydrate for 30 min. After the pellet was gently resuspended, cells were added to 5 ml of Anaerobe Basal broth (Oxoid, France), and incubated for 2 days at 37 °C. Pure cultures were mixed in 1:1 proportion, and 100 µl of the resulting mixture were added to 1 ml of a semi-synthetic medium (Barrangou R., 2003) with addition of (a) lactose (Sigma Aldrich, Germany) or (b) lacto-N-biose (Dextra Laboratories, UK) as a single C-source (1.5 %). Medium without a C-source and Anaerobe Basal broth medium were used as controls. Prior to mixing bacteria, pure cultures were diluted in Anaerobe Basal broth medium to reach equal OD_{600}. DNA isolation, PCR and sequencing reactions were performed both prior to, and after, incubation at 37 °C for 1 day with a semi-synthetic medium.

2.7 Bioinformatics and statistics

All the analyses of sequence spectra were performed using MATLAB® R2010a software (The MathWorks Inc., Natick MA, USA), Statistical and Bioinformatics toolboxes for MATLAB®. For EFA, PCA and MCR analyses, PLS Toolbox v5.8 for MATLAB® (Eigenvector Research Inc., USA) was used.

Following sequences were used as references: *clpC* allelic profiles downloaded from the *Bifidobacterium* MLST Database (Brisse S. et al, 2009); *clpC* gene of *B. adolescentis* (DQ238016); *B. breve* (AB437352); *B. dentium* (AY722387); *B. bifidum* (DQ206821); *B. longum longum* (AB437353) and *B. longum infantis* (AY722381).

2.7.1 Alignment and trimming of sequence spectra

The original script for aligning and trimming 16S rRNA sequence spectra for further analysis using MCR-ALS was written by Monika Sekelja (Zimonja M. et al., 2008). For this project, the script was optimized for its use with *clpC* gene amplicons. All query sequences are aligned to a predefined start fragment (20 nucleotides long), trimmed, then aligned to the end fragment (17 nucleotides long) and trimmed again (Figure 2.1). Quality of spectra alignment

is crucial for the performance of MCR-ALS – any shifts or deviations in peaks would mask actual similarity between the data and thus hinder the prediction of the common spectra. To correct for retention shift differences and to increase the linearity of the data (Zimonja M. et al., 2008), after the alignment the spectra were first preprocessed using correlation optimized warping (COW) and then normalized. COW allows aligning the given spectrum towards the reference one by moving segments of the spectrum within the defined limit (Tomasi G. et al, 2004).

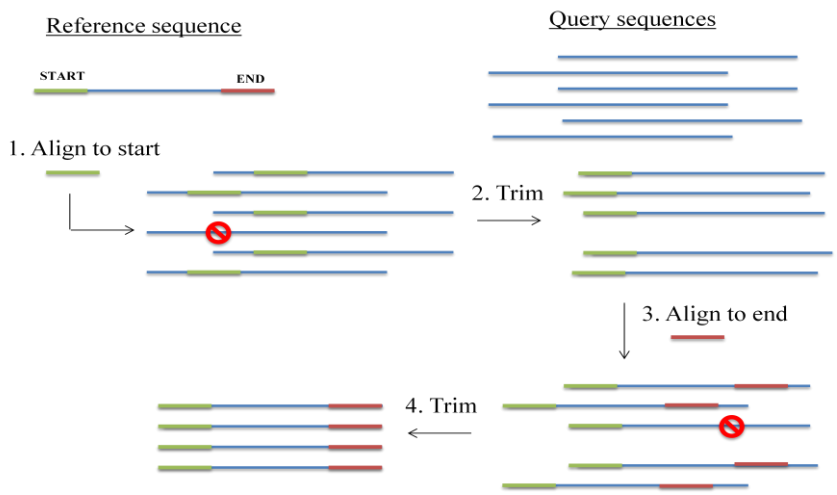

Figure 2.1 The scheme of sorting and trimming of sequences. Query sequences are screened for the reference 'start' and 'end' fragment (each approx. 20 bp long). Those which fail to have either of them are removed. All the others are trimmed up to the 'start' and beyond the 'end.' The consensus sequence from the multiple alignment of Bifidobacterium clpC alleles was used as the reference.

2.7.2 Resolving spectra

A multivariate curve resolution analysis (MCR-ALS) with a non-negativity constraint on contribution (true least squares solution) was conducted to resolve the spectra. Initially, the number of components, i.e. the number of bacterial groups present in the system, was determined both by principal component analysis (PCA) and evolving factor analysis (EFA).

2.7.3 Basecalling spectra

For decoding resolved pure spectra, we developed a basecaller script for use in the MATLAB® environment (Appendix N). Each spectrum is represented by an array containing fluorescence intensity measurements of each of four nucleotides at any given time. Plotting

these intensities for each of the nucleotides in one figure allows us to visualize the spectrum and see the corresponding nucleotide sequence of it. Nucleotides are plotted one after another, meaning that firstly all the measurements for one nucleotide are visualized, then they are overlaid by the intensities of the second nucleotide and so on, until all four are plotted. The basecaller script is based on a similar idea – it has four cycles, during which it takes all the intensity measurements for one nucleotide and screens them for peaks (Figure 2.2). Due to the fact that background fluorescence intensity can also form noise which has no biological meaning, a custom intensity threshold can be set by the user. Each time the peak with an intensity higher than a threshold is determined, its coordinate and a part of the initial array where it belongs to (i.e., which nucleotide it is), is stored in a separate array. After all nucleotides were screened, the resulting array is sorted by peak coordinates, which in turn results in nucleotide sequence of the spectra. In case of mixed spectra or bad resolution of them, sometimes peaks might be on top of, or too close to each other. Therefore, the possibility for setting a threshold for mixed peaks was developed. If the difference between the coordinates of two subsequent peaks is lower than the given threshold, the script calls an IUPAC nucleotide ambiguity code to re-designate doubly-called peaks.

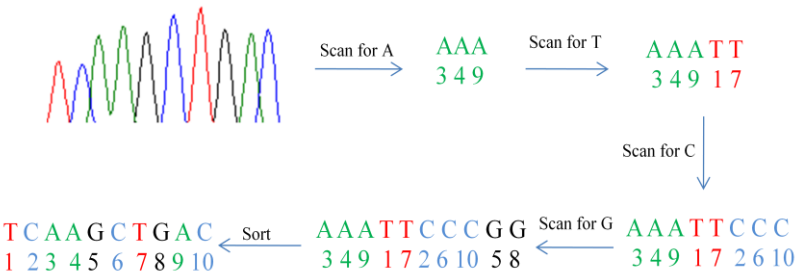

Figure 2.2 Simplified scheme of a basecaller algorithm. In each cycle it takes fluorescence intensity measurements for one nucleotide, screens them for peaks and saves the coordinate (retention time) of it in a special array. After all the nucleotides were screened, the program sorts coordinates, which results in a sequence of a given spectrum

2.7.4 Correlation analysis

All factors were checked for the cross-correlation to each other using Pearson's chi-square test – an approximate test which uses a contingency table to calculate whether the two observations develop independently (Statistics solutions, 2006).

The correlation of both the abundance of dominant *Bifidobacterium* species present in the samples at a given time point and of its change over two subsequent time points to various factors was calculated using two-sided permutation test with 10^6 permutation replications. The rationale of the permutation test is that, if both groups (e.g., sensitized and non-sensitized) belong to one distribution, then randomly rearranging the labels to the groups of observed values does not affect the absolute difference between means of the 'new' reassigned groups compared to the original ones (Edgington E., 1995).

The percentage and development in time were calculated both relative to the *Bifidobacterium* group (MCR predicted percentages) and to the total bacterial load (recalculated from percentage of *Bifidobacterium* group relative to total bacterial load). Data on the total faecal microbiota were taken from the previous study of 16S rRNA amplicons of IMPACT samples (Appendix A). The significance of the change in the abundance between two subsequent time points was calculated with Kruskal-Wallis test, a non-parametric version of one way ANOVA. This test makes an assumption that both groups have the same distribution shape and scale, except for a difference in medians, rather than that they come from same normal distributions as in ANOVA (MATLAB® documentation, 2010).

In all cases the null hypothesis was rejected at the level of 5 %.

2.7.5 Presence/absence test

Individuals, having measurements for their first 4 days of life, as well as for their mothers during pregnancy, were included in the analysis. The samples, where the *Bifidobacterium* group was not detected by sequencing, were also added to the dataset. This resulted in a set of 37 infants and their mothers. Due to the error of MCR prediction, the threshold of 7 % was used for binarizing the data – samples with the percentage of *Bifidobacterium* species lower than a threshold were marked as absent, whereas those with the percentage higher or equal to 7 % were marked as present. The correlation was tested using Fisher's Exact test calculator designed by Øyvind Langsrud (Langsrud Ø., 2004). This independence test determines whether associations between two categorical variables and two factors of comparison are received by chance by operating on a 2x2 contingency table of these variables and factors and calculating a probability value of the relation (Agresti A., 1992).

2.7.6 Network calculation

To analyze the correlation between the given species with regards to the development of other species or bacterial groups, a Local Similarity Analysis (LSA) was conducted. This test allows identification of spatial and temporal correlations among species, even if those correlations are delayed in time and/or are present during a limited period (Xia L. et al., 2008). This analysis is based on calculating pairwise local similarity scores between time series and then checking for the statistical significance of these scores using a permutation method (Ruan Q. et al., 2006).

The resulting network was visualized with Cytoscape software (Cline M. et al., 2007). Calculations were performed relative to both *Bifidobacterium* and total bacterial loads.

3 Results

3.1 Characteristics of raw sequencing data

Mixed sequences of *Bifidobacterium clpC* gene from stool samples of mothers and their children were obtained in parallel with designing the method. Most of the sequences had abnormally shaped fluorescence peaks from the start and up to 30 – 60 bp. Also, the resolution of the spectra decreased drastically after 400 bp (Figure 3.1).

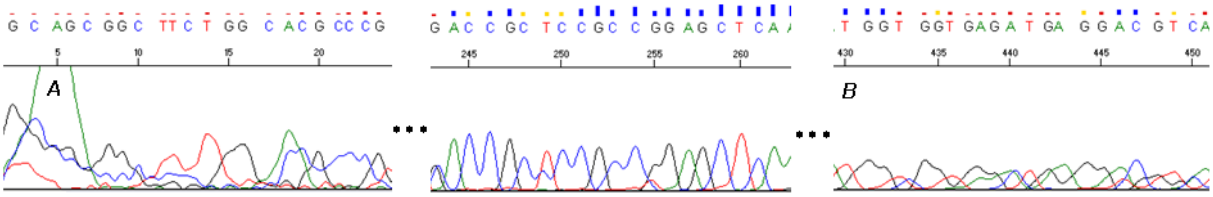

Figure 3.1 Example of both (A) abnormally shaped peaks in the beginning of the sequence and (B) spectra resolution loss after 400 bp of Bifidobacterium clpC amplicons from stool samples of mothers and their children.

Another common problem we encountered with the sequences obtained from *clpC* amplicons was the formation of high fluorescence intensity signals in several regions of the sequences. Mostly, these abnormal peaks were coming from the T-nucleotide (Figure 3.2).

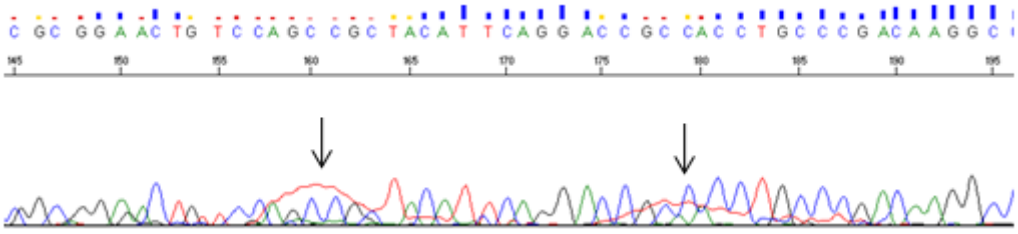

Figure 3.2 Abnormally shaped T-nucleotide high intensity fluorescence peaks that may have caused improper correction of spectra alignment

3.2 Design and validation of mixed *clpC* sequence analysis

3.2.1 MCR-ALS resolution of mixed spectra

Applicability of MCR-ALS analysis for the prediction of *Bifidobacterium* species present in the data set was assessed on predefined mixtures of four species (*B. adolescentis*, *B. breve*, *B. longum longum* and *B. longum infantis*), which are commonly associated with human gut microbiota (Tannock, 2010). Mixtures were designed using simplex lattice design resulting in

35 mixtures where the PCR product of each strain was present in each of five proportions (1:0:0:0; 0.25:0.25:0.25:0.25; 0.75:0.25:0:0; 0.5:0.5:0:0; 0.5:0.25:0.25:0). Initial concentrations of PCR products were quantified using Molecular Imaging software (Kodak, USA). The experiment was performed in duplicate.

Profile sequences of 34 known *clpC* gene alleles were downloaded from the *Bifidobacterium* MLST Database (Brisse S. et al, 2009) and were aligned collectively. Two conserved regions which could serve as a start (19-29 bp and 37-59 bp) and two regions which could serve as an end (265-288 bp; 580-602 bp) fragments for trimming query spectra were identified (Figure 3.3).

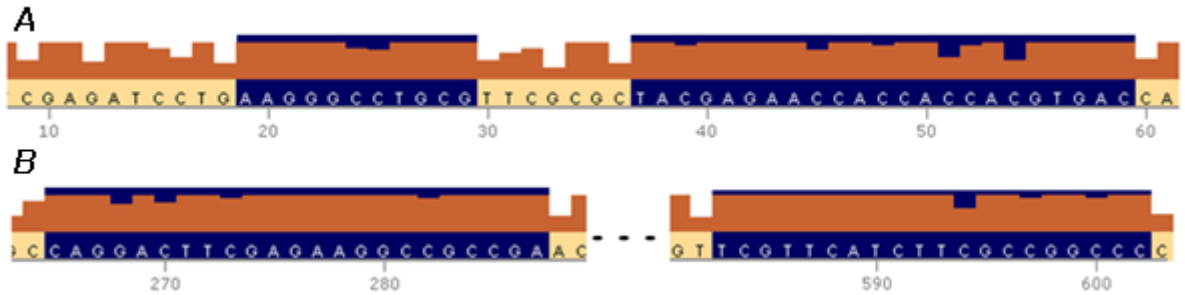

Figure 3.3 Consensus sequence of 34 clpC allelic profiles progressive multiple alignment. Profile sequences were downloaded from the Bifidobacterium MLST Database (Brisse S. et al, 2009). Conserved regions are highlighted in blue. A: two potential start regions for spectra trimming; B: two potential end regions.

Due to high fluorescence regions and loss of resolution (Figure 3.1) it was decided to use conserved regions positioned at 37-59 bp and 265-288 bp as start and end fragments respectively. Prior to the MCR-ALS resolution, alignments of trimmed spectra were corrected using COW.

To correct for abnormal T-nucleotide peaks (Figure 3.2), we evaluated the possibility of spectra resolution without the information on T-nucleotide. We removed fluorescence intensity measurements of the T-nucleotide from predefined mixtures, analyzed the spectra and compared the results with those, when information on all four nucleotides was available. For the spectra identification, T-information was re-substituted on the basecalling stage.

Classification of pure resolved spectra was performed both by searching the sequence against the nucleotide database with BLAST (Table 3.1) and by calculating p-distances between resolved components with reference *clpC* gene sequences of *B. adolescentis*, *B. breve*, *B. longum longum* and *B. longum infantis* (Table 3.2).

Table 3.1 Lowest E-value BLAST search hit of MCR-ALS resolved spectra sequences against nucleotide database.

MCR-ALS resolved spectrum	Closest BLAST search hit	GenBank accession number	E - value	Identity[1)], %
Component 1	*B. adolescentis*	DQ238016.1	$3*10^{-107}$	99
Component 2	*B. breve*	AB437352.1	$3*10^{-107}$	99
Component 3	*B. longum*	AP010889.1	$4*10^{-111}$	100

1) Query coverage – 100 %.

Table 3.2 Pairwise p-distances between MCR-ALS resolved pure spectra from predefined mixtures and corresponding reference *clpC* gene sequences

Spectra resolved by MCR	Reference spectra[1)]			
	B. longum infantis	*B. longum longum*	*B. breve*	*B. adolescentis*
Component 1	0.2330	0.2375	0.2417	**0.1029**
Component 2	0.1674	0.1632	**0.0879**	0.2541
Component 3	**0.0795**	**0.0837**	0.1715	0.2459

1) Prior to distance calculation, reference sequences were aligned to start and end fragments and trimmed

Correlation coefficient and root mean square error of prediction (RMSEP) were used to evaluate the accuracy of MCR-ALS prediction.

Table 3.3 Correlation coefficients and RMSEP values of MCR-ALS predicted values with regard to actual concentration profiles

		B. adolescentis		*B. breve*		*B. longum* spp.	
		Corr. coef.	RMSEP	Corr. coef.	RMSEP	Corr. coef.	RMSEP
General prediction	Duplicate1	0.971	7.677	0.962	7.399	0.985	5.933
	Duplicate2	0.981	5.957	0.970	5.499	0.981	6.506
T- nucl. information removed	Duplicate1	0.972	7.435	0.963	7.166	0.985	5.905
	Duplicate2	0.987	3.813	0.974	5.169	0.967	6.903

There was 96 % - 98 % correlation between actual and predicted values regardless of whether information on all, or just three, nucleotides was present in the dataset (Table 3.3). The removal of T-information reduced the error of prediction.

The possibility of further correction of predicted concentrations using PLS regression was tested. This, however, improved the MCR-prediction only to a minor extent by slightly decreasing the error (Appendix B).

3.2.2 Resolution of the *B. longum* group

Due to the fact that most of the nucleotide variation between *B. longum longum* and *B. longum infantis* starts from 441 bp with regards to the profile consensus (Appendix C), we combined the *B. longum* group for all species identification and resolved it separately. To do so, we designed a specific primer which binds closer to a variation site between *B. longum longum* and *B. longum infantis.* All allelic profiles were collectively aligned with reference sequences of the four species used in the mixture design, as well as with those of *B. bifidum* (DQ206821.1), *B. dentium* (AY722387.1), *B. ruminantium* (DQ238030.1) and *B. animalis lactis* (AY722379.1). Two positions with a nucleotide specific only for *clpC* alleles and reference sequences of *B. longum* group were identified in a 200-440 bp region (Table 3.4). As long as a variation site between *B. longum longum* and *B. longum infantis* starts at 441bp, it was decided to use a T-nucleotide positioned at 321 bp as a 3'-end nucleotide in a primer for *B. longum*-group sequencing (Table 2.1).

Table 3.4. Nucleotide variations in *B. longum longum* and *B. longum infantis* profile alleles and reference sequences compared to other *clpC* alleles and reference sequences of *B. adolescentis, B. breve, B. ruminantium, B. dentium, B. bifidum* and *B. animalis lactis*[1)]

Position number[2)]	Nucleotide in *B. longum* group	Nucleotide in other alleles and reference sequence	Specificity for the *B. longum* group
198	T	C(G)	**Yes**
213	C	T(G)(C)	No
300	G	C(G)	No
321	T	G	**Yes**
328	C	G(A)(C)	No
423	C	T(C)	No

1) Only those positions for which there was a 100 % likelihood of the given nucleotide in the *B. longum* species and *clpC* alleles corresponding to the *B. longum* group species, were compared.
2) Position number is given according to allelic profiles (603 bp).

To check for the specificity of the primer, sequencing reactions were performed on PCR products from pure *B. adolescentis, B. breve, B. longum longum* and *B. longum infantis*, as well as on mixtures of PCR products containing equal parts of two of these *Bifidobacterium* species. PCR products of *B. longum longum* and *B. longum infantis* were mixed in 50/50 and 70/30 proportions. There were no sequences obtained from pure *B. adolescentis* or *B. breve* PCR amplicons, whereas in case of *B. longum longum* and *B. longum infantis* mixtures, mixed sequences corresponded well with the predefined proportions of each species in the mixture (Figure 3.4).

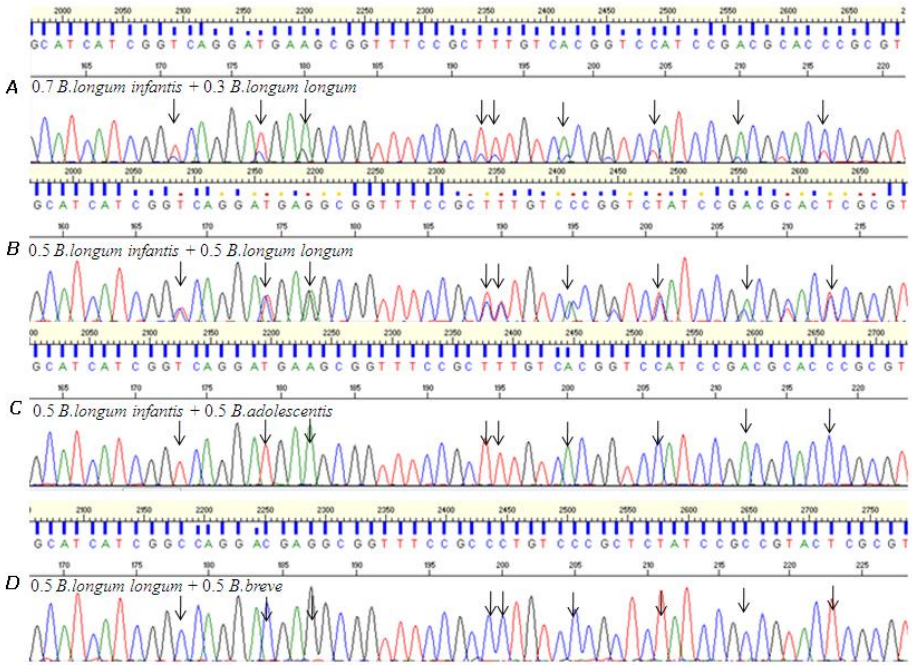

Figure 3.4 Sequencing of predefined mixtures of B. adolescentis; B. breve; B. longum longum and B. longum infantis with clpC-L primer. 100 % likely variations between B. longum longum and B. longum infantis are indicated. A, B: 70/30 and 50/50 mixtures of B. longum infantis and B. longum longum respectively; C, D: mixtures of the B. longum group species with B. adolescentis and B. breve respectively.

3.3 Identification of dominant species

3.3.1 Characteristics of clinical material

In total, DNA isolated from 554 samples was PCR amplified and sequenced. In 113 samples, *Bifidobacterium* was not detected and samples were removed from the analysis. Spectra of *clpC* sequences of the remaining samples were aligned to a common start and end followed by trimming. However, some spectra had many abnormal peaks not only caused by T-, but also C-nucleotides, or were missing start or end regions. These sequences had to be removed as well. Also, after MCR-ALS was performed, those samples that corresponded to the same individual at the same age, were removed from the analysis. Overall, out of 110 children and their mothers, *Bifidobacterium* species were identified in stool samples of 83 individuals at seven different time points (330 samples in total). At early pregnancy, bifidobacteria were identified in 53 % of individuals; at late pregnancy – 41 %; 2-4 days – 41 %; 10 days – 51 %; 4 months – 71 %; 1 year – 66 % and at 2 years – in 75 % of individuals (Figure 3.5).

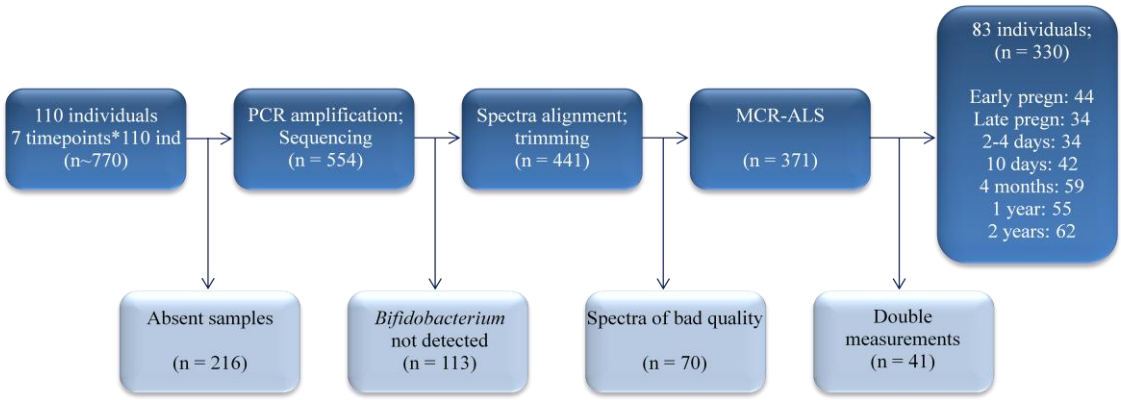

Figure 3.5 Scheme of number of stool samples (n) on each step of analysis. Loss of information is depicted in pale blue rectangles. Spectra of bad quality comprise sequences which either had low signal or missed start or end fragment of the alignment. Some individuals had two or more stool samples taken at the same age, these duplicates were also removed from the final analysis.

3.3.2 *Bifidobacterium* group resolution

PCA and EFA predicted 3 and 6 components in the dataset, respectively. Because iterative methods are commonly advised for MCR-ALS analysis (Tauler et al., 1995), we decided to perform MCR with 6 components (Appendix D). However, one of the components was poorly resolved, resulting in a longer nucleotide sequence spectrum with many mixed peaks which were hard to interpret (Figure 3.6). Therefore we repeated the MCR-ALS analysis with five components. All five components were well resolved (Appendix E). The basecalled components' spectra were identified by searching the sequence with BLAST against the NCBI nucleotide database (Table 3.5). These 5 components explained 57.3 % of the variation in bifidobacterial composition of stool samples.

Table 3.5 Lowest E-value BLAST search hits of MCR-ALS-resolved spectra sequences against the NCBI nucleotide database. Five components were used for resolution. Query coverage – 100 %.

MCR-ALS resolved spectrum (sequence length)	Closest BLAST search hit	GenBank accession number	E - value	Identity, %
Component 1 (221 bp)	*B. bifidum*	CP002220.1	$1*10^{-86}$	93
Component 2 (215 bp)	*B. dentium*	CP001750.1	$3*10^{-92}$	95
Component 3 (219 bp)	*B. adolescentis*	DQ238016.1	$3*10^{-107}$	99
Component 4 (216 bp)	*B. breve*	AB437352.1	$4*10^{-101}$	97
Component 5 (221 bp)	*B. longum*	AP010889.1	$4*10^{-111}$	100

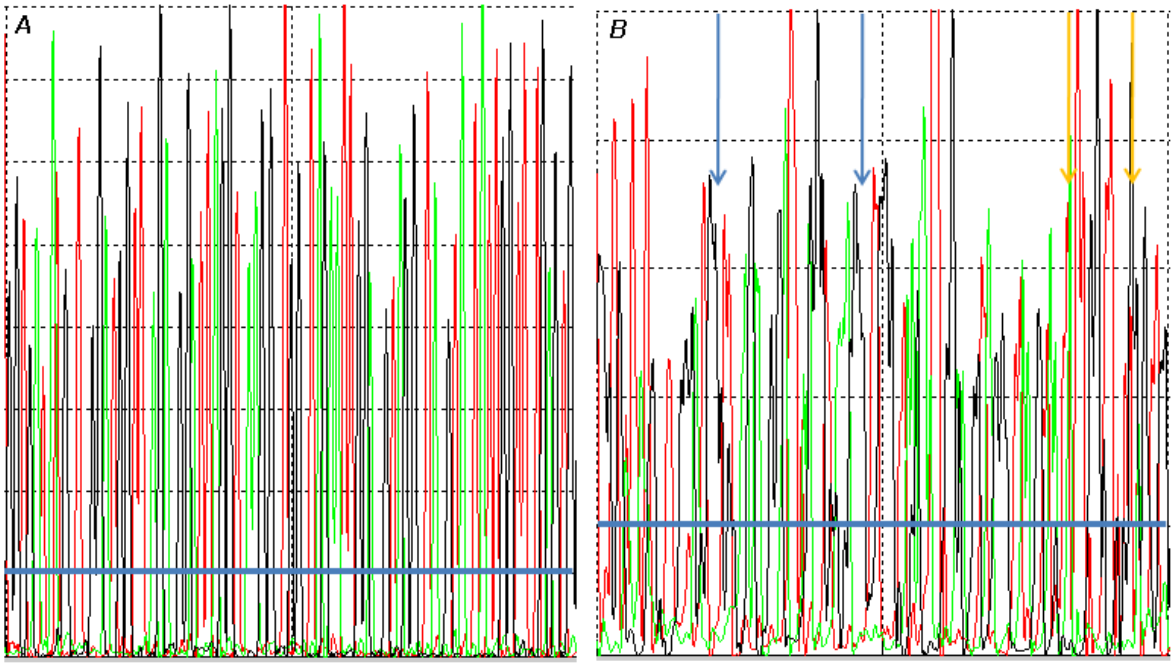

Figure 3.6 Example of well-resolved (A, MCR with 5 components) and poorly resolved (B, MCR with 6 components) component spectrum corresponding to the B. longum group. Information on the T nucleotide was removed - blank spaces correspond to T. In the case of poorly resolved spectra, two problems arise: peaks are sometimes vague and it is hard to decide whether there is information regarding one or two nucleotides (blue arrows), and peaks are mixed (yellow arrows).

To validate identification of dominant species, *clpC* amplicons of fourteen stool samples were selected for cloning. Stool samples were chosen on the basis of the percentage of each component, predicted by MCR-ALS. In total, there were 6 samples regarded as pure (the abundance of one of the components was more than 80 %), and 8 samples that had a combination of two/or more species. Sequencing of cloned inserts confirmed the presence of predicted species in corresponding stool samples (Table 3.6). Cloned inserts shared 98 % - 99 % identity with the corresponding reference sequences deposited in GenBank. In one sample, *B. animalis* was detected.

Table 3.6 Number of transformed *E.coli* colonies that contained *clpC* gene amplicons from specified species.

ID	age	MCR-predicted composition, %					Number of colonies where the species was detected[1)]						Correspond with the prediction[3)]
		Comp 1- *B. bifidum*	Comp 2 - *B. dentium*	Comp 3 - *B.adolesc.*	Comp 4 - *B. breve*	Comp 5 - *B. longum*	*B. bifidum*	*B. dentium*	*B.adolesc.*	*B. breve*	*B. longum*	*B. animalis*	
1384	1 y	4.3	3.3	0	4.2	88.2	2	–	3	3	1	–	++/–
1892	1 y	0	100	0	0	0	–	7	–	–	–	–	+
2650	1 y	46.0	17.2	0	8.7	28.1	10	–	–	–	–	–	+
138	4 m	4.4	0	2.4	93.1	0	–	–	–	8	–	–	+
1353	1 y	0.4	3.7	89.6	0.3	6.0	–	–	9	–	1	–	++
1870	1 y	0	0	0	1.0	99.0	–	–	–	–	10	–	+
1138	11 d	29.2	0	0	3.6	67.2	–	–	–	–	7	–	+
1807	L pr[2)]	17.2	2.0	52.8	2.0	26.1	–	–	–	–	–	10	–
1370	4 m	22.0	0	30.9	2.1	45.0	–	–	1	–	7	–	++
1391	2 y	42.9	4.9	44.4	7.8	0	–	–	–	9	–	–	+
1384	4 m	22.3	36.9	0.2	0	40.6	–	3	–	–	–	–	+
195	4 d	0	1.4	0	0	98.6	–	–	–	–	8	–	+
1230	4 m	1.1	0.3	0	65.0	33.6	–	–	–	–	10	–	+
1870	4 d	15.0	81.8	3.3	0	0	–	10	–	–	–	–	+

1) All positively transformed colonies per each sample were analyzed
2) L pr – late pregnancy
3) ++: more than 1 of the predicted components detected; +: one of the predicted components is detected; –: the component, predicted to be absent, was detected

3.3.3 Average composition of *Bifidobacterium* species in the stool samples

A graphical representation of average composition of *Bifidobacterium* species in stool samples of infants and their mothers is presented in Figure 3.7. Stool samples from pregnant women were predicted to be rich in *B. adolescentis*, *B. bifidum* and *B. longum*, while *B. breve* and *B. dentium* were present in smaller amounts. Right after birth and up to 4 months of age, *Bifidobacterium* species were dominated by *B. longum* and *B. breve*. The *B. longum* group had its peak at 10 days of age comprising 57 % of bifidobacterial microbiota at this age. Interestingly, *B. breve* separated all 10-day-old infants in two distinct groups. In one group, it was accounting for less than 15 %, whereas in the other for more than 75 % of the bifidobacterial load. There was no single infant who had an intermediate abundance of *B. breve* between the two values at this age. By the age of 4 months, *B. breve* increased in its abundance in many infants, and, in average, it became predominant, whereas *B. longum* remained on the level of 38 % from this time on. After 4 months, *B. breve* and *B. adolescentis*

became negatively correlated, as one decreased, reaching the level of 7.0 % by the age of 2 years, while the other increased up to 28.3 %. Highest fluctuations in the percentages were exhibited by *B. longum*, *B. breve* and *B. adolescentis*. *B. dentium* seemed to be rather stable in time, whereas *B. bifidum* showed a decrease from birth up to 10 days and then increased its abundance, reaching levels of around 20 % by 1-2 years (Appendix E).

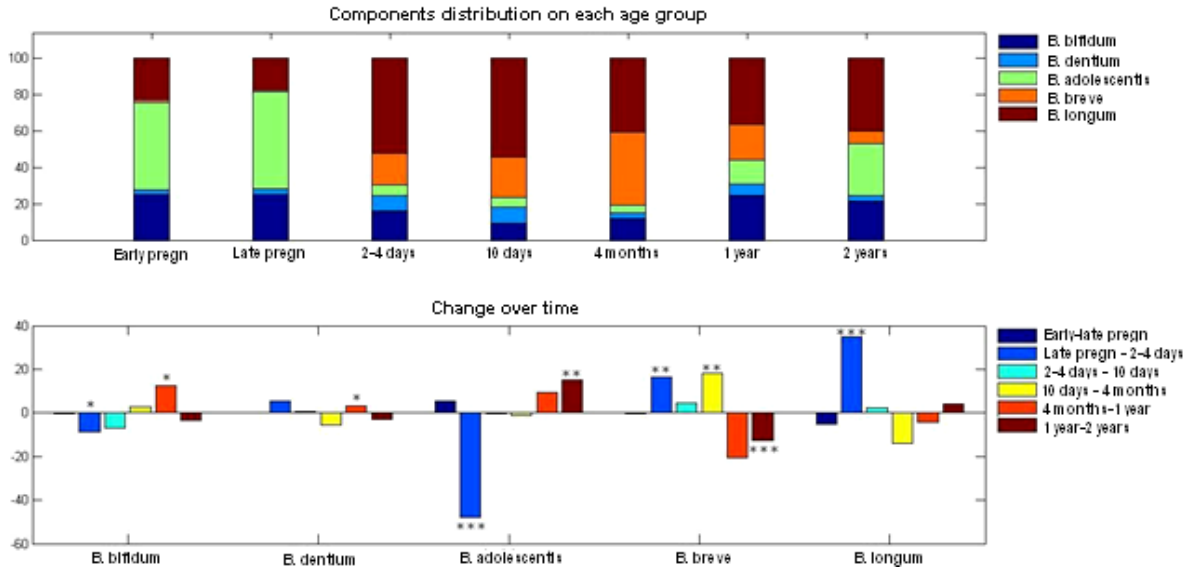

*Figure 3.7 Relative Bifidobacterium species composition and their change over time in stool samples of infants (from 3 days to 2 years of age) and their mothers during pregnancy. Significance of the changes were tested using the Kruskal-Wallis test. *p<0.05; **p<0.01; ***p<0.001.*

Relative to the total microbial load, fluctuations in *B. adolescentis* were masked by the overall development of the whole *Bifidobacterium* group, which showed a remarkable decrease from 59.3 % at the age of four months to 10.8 % and further to 6.7 % at 1 and 2 years, respectively (Appendix E). Here, *B. adolescentis* levels also remained quite stable over time, leaving most of fluctuations to *B. longum*, *B. breve* and *B. bifidum* (Figure 3.8).

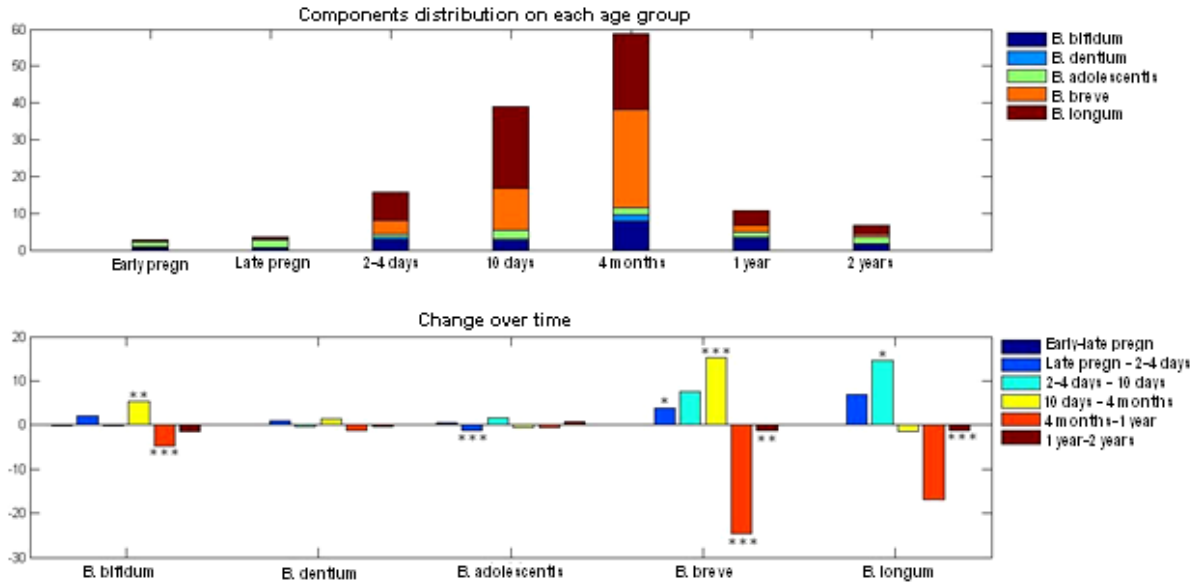

*Figure 3.8 Bifidobacterium species composition and their change over time relative to the total gut microbiota in stool samples of infants (from 3 days to 2 years of age) and their mothers during pregnancy. Significance of the changes were tested using the Kruskal-Wallis test. *p<0.05; **p<0.01; ***p<0.001.*

3.3.4 Local Similarity Analysis

Relative to the remaining *Bifidobacterium* species, abundance of *B. adolescentis* seemed to be in synch with that of *B. bifidum* and opposite to that of *B. breve*. At the same time, the development of *B. longum* and *B. breve* appeared to be counter-indicative of *B. bifidum*, although there was a two time point delay between *B. bifidum* and *B. breve*. Relative to *Bifidobacterium*, *B. dentium* supposedly developed independently of other dominant species (Figure 3.9).

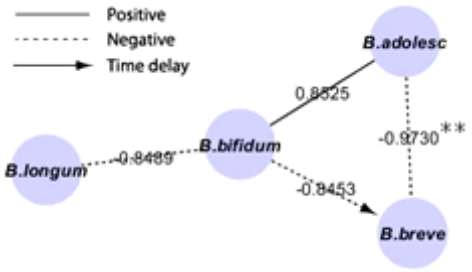

*Figure 3.9 Network interactions and local similarity scores between dominant Bifidobacterium species. Only interactions with a p-value of less than 0.05 are presented. **p<0.01.*

For the LSA relative to the total bacterial load, outer bacterial groups were also taken into consideration. Here, *B. dentium*, *B. breve* and *B. bifidum* seemed to show a positive

correlation with each other, whereas the abundance of *B. longum* was only found to correlate with that of *B. breve*, compared to the remaining *Bifidobacterium* species. However, with respect to other residents of the gut, *B. longum* showed the correlation with *Lactobacillales*, other *Bacilli*, *Clostridia* and *Proteobacteria* groups (Figure 3.10). Interestingly, relatively to *Bifidobacterium*, *B. bifidum* seemed to exhibit a negative time-delayed correlation to *B. breve*, whereas relative to the total microbial load, abundances of these bacteria developed in the same direction.

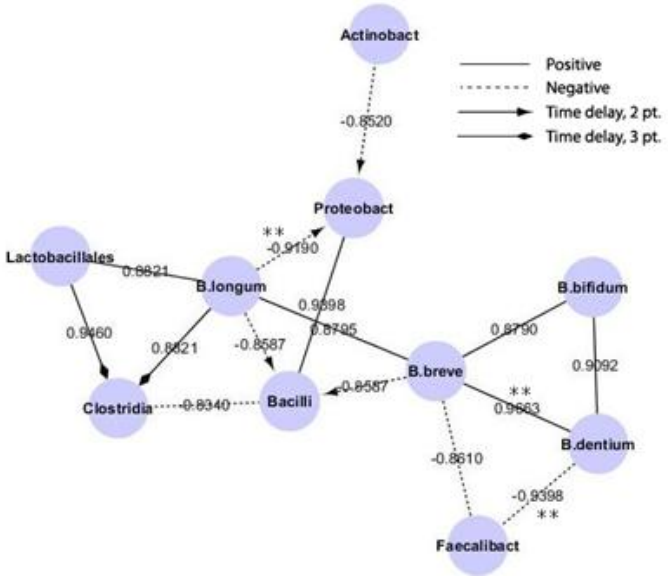

*Figure 3.10 Network interactions and local similarity scores between dominant Bifidobacterium species and other bacterial groups. Percentages shown are relative to the total microbial load. Only interactions with a p-value of less than 0.05 are presented.**p<0.01.*

3.3.5 Presence/absence test

The abundance of *Bifidobacterium* load in mothers was not correlated to that of their newborn infants as revealed by Fisher's Exact test performed on binarized data. Plotting the MCR predicted abundances of dominant species in stool samples from infants against that from the mothers also suggested independence (Appendix H).

3.3.6 *B. longum* group composition

There was a high concordance between MCR-ALS predictions of the *B. longum* group and whether or not the sequence was obtained with clpC-L primer – the *B. longum* group was detected in 214 samples out of 280 predicted by MCR-ALS, and was not detected in 49 out of

92 where it was predicted to be absent ($p = 8.99*10^{-8}$). For this calculation, the sequence was regarded as predicted to be present by MCR-ALS if the percentage of the *B. longum* group was predicted to be equal to or more than 6.5 % (mean RMSEP for the *B. longum* group, Table 3.3). In total, the *B. longum* group was detected in 80 % of all samples, with the smallest number of sequences belonging to the 3 day-old infants (73 % of all individuals in whom *Bifidobacterium* spp. were identified) and highest number to 2-year-old children (98 % of all individuals in whom *Bifidobacterium* spp. were identified). The sequences obtained were pure and thus MCR-ALS analysis was not performed. Instead, all the sequences were collectively aligned and inspected manually. Most of them belonged to the *B. longum longum* species. In total, there were 39 sequences with more than 3 nucleotide variations compared to the multiple alignment consensus, out of which only 13 had a nucleotide variation pattern resembling that of *B. longum infantis*, and 5 sequences lacked one or two of nucleotide variations between *B. longum longum* and *B. longum infantis* (for the list of nucleotide variations, see Appendix C). Sequences with a similar nucleotide variation pattern were aligned together and the consensus sequence of the alignment was searched against the NCBI nucleotide database using BLAST (Appendix F). One of the clusters could not be assigned to any of *B. longum* group species, therefore, henceforth in this thesis, these sequences will be referred to as Unclassified. Sequences which were identified as *B. longum longum*, but had 3 or more nucleotide variations compared to the reference *B. longum longum* sequence will be referred to as *B. longum* unclassified. Those that lacked one or two nucleotide variations between *B. longum longum* and *B. longum infantis*, will be denoted *B. longum infantis* – like *B. longum*; and sequences possessing all nucleotide variations will be referred to as *B. longum infantis*. Two sequences belonging to *B. longum infantis*–like *B. longum* were identified as *B. longum suis*.

The maximum likelihood phylogenetic tree of 39 extracted sequences together with reference sequences of *B. adolescentis*, *B. dentium*, *B. bifidum*, *B. breve*, *B. longum longum* and *B. longum infantis* is presented in Figure 3.11. Most of these 39 sequences originated from the stool samples of 10-day- and 4-month-old children, and only one came from a pregnant woman.

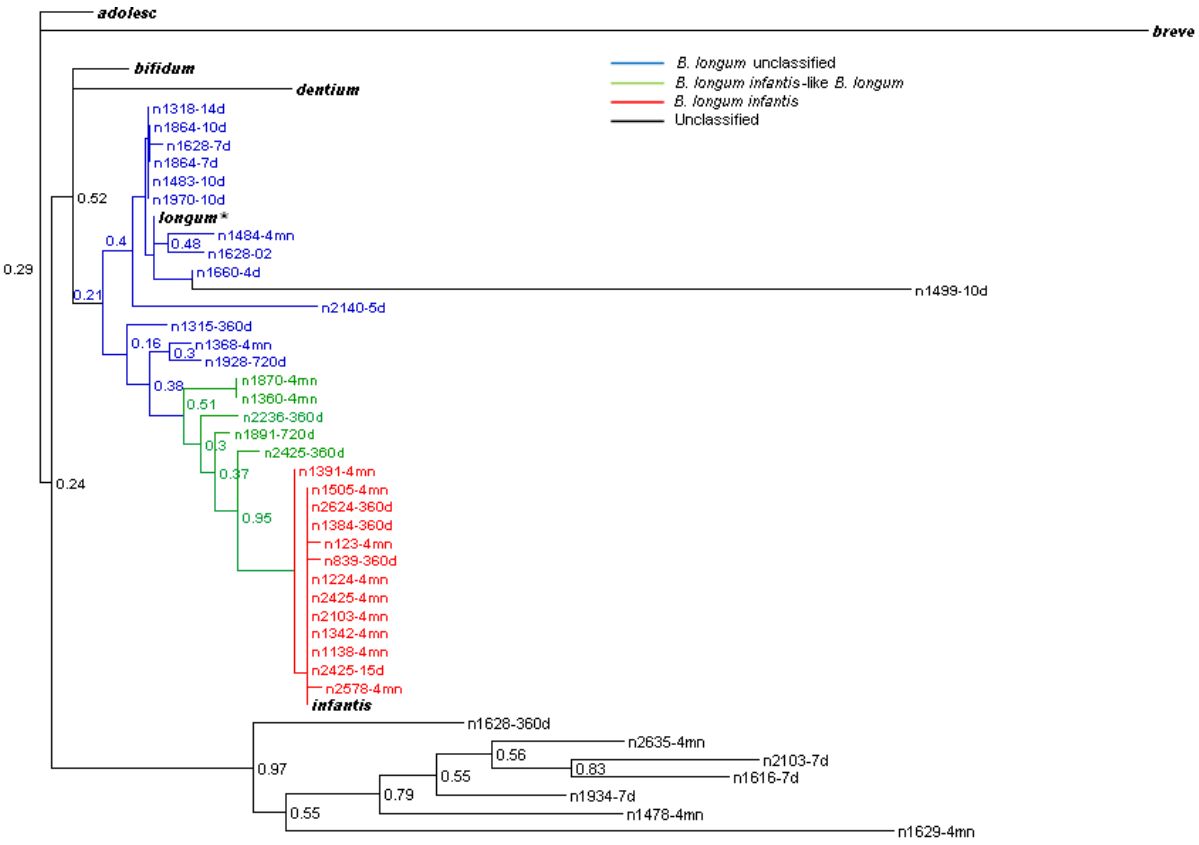

*Figure 3.11 Maximum likelihood phylogenetic tree (phyML) of B. longum sequences that had more than 3 nucleotide variations compared to the consensus of all aligned B. longum sequences. Multiple alignment was performed using the MUSCLE algorithm. Prior to phylogenetic reconstruction, all sequences were pairwise aligned to predefined start and end regions and trimmed. Reference sequences are highlighted in bold. The tree was constructed using the Phylogeny.fr online tool (Dereeper A., Guignon V. et al., 2008) and visualized with Dendroscope (Huson D. et al., 2007).*The reference sequence for B.longum longum also represents 238 sequences which had no nucleotide variations compared to the reference sequence.*

Only *B. longum longum* was detected in all pregnant women, except for one mother, who had *B. longum* unclassified (Figure 3.12). Right after birth, one baby harbored an unclassified sequence and one, the offspring of the mother with *B. longum* unclassified, had a sequence belonging to the same subgroup as the mother. From 10 days of age, though still being the most common subgroup, *B. longum longum* exhibited a decreasing prevalence, being found in 71 % and 64 % of children at 10 days and 4 months, respectively. At the same time, the unclassified subgroup was found in 11 % (10 days) and 7 % (4 months) of children. Most of *B. longum infantis* sequences were characteristic for children aged 4 months – there was only one baby, in which it was detected earlier, and only two, where it persisted up to 1 year of age. By the age of 2 years, *B. longum longum* seemed to be regaining its prevalence as it was found in 59 out of 61 children.

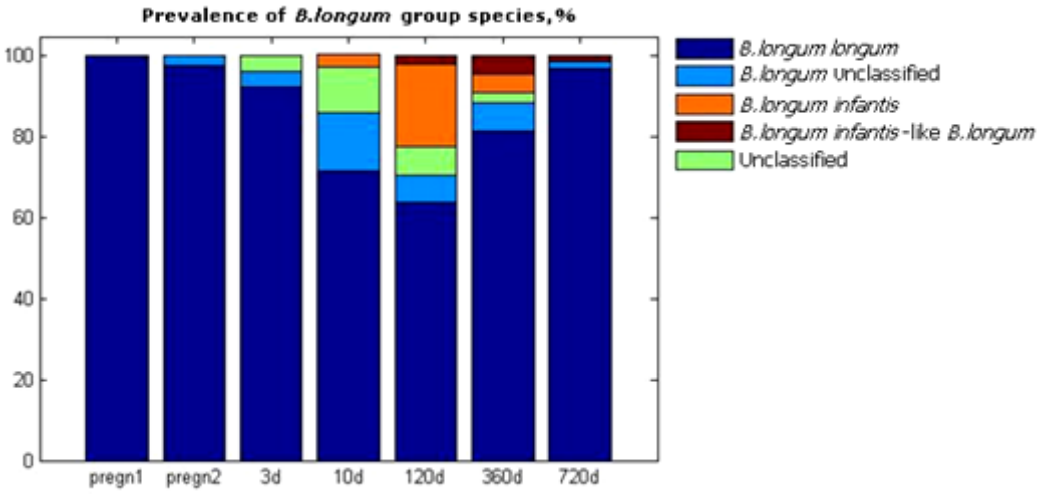

Figure 3.12 Prevalence of B. longum group species at different ages. Numbers of samples where bacterial species were detected is given as a percentage (y – axis). Pregn1 and pregn2 – early and late pregnancy periods; 3d, 10d, 120d, 360d and 720d – 2 – 4 days, 10 days, 4 months, 1 year and 2 years of age, respectively.

To check whether there is a concordance between the *B. longum* sequence of the mother and that of the child during the first days of his/her life, sequences of mother/infant were pairwise aligned and compared. There were only nine such pairs, out of which only four had 100 % or 99 % identity (Appendix H).

3.3.7 Competition test

To investigate whether *B. longum longum* and *B. longum infantis* were outcompeted by one another during infancy, we decided to perform a competition test with a one carbon-source semi-synthetic medium. Due to the fact that human milk oligosaccharides are extremely difficult to purify and are commercially unavailable, we chose to use lactose and lacto-N-biose, which is broadly known as a proxy for HMOs (Nishimoto M. et al., 2007), as single carbon sources. Unfortunately, isolation of *Bifidobacterium* spp. strains from IMPACT stool samples was unsuccessful. The experiment was performed twice with Bereens medium and Wilkins-Chalgren for *Bifidobacterium* Mupirocin medium (Chapter 2.7.1). There were no colonies observed on plates, inoculated with 1:100 and 1:1000 dilutions of stool samples, whereas colonies on 1:10 dilution plates belonged to *Lactobacillus paracasei* and *Lactobacillus casei*, as shown by sequencing of 16S rRNA gene amplicons. Therefore pure cultures of *B. longum longum* (DSM20219) and *B. longum infantis* (DSM20088) were obtained from the German Collection of Microorganisms and Cell Cultures. The competition

experiment was performed twice. However, the results were inconclusive (Appendix G). Due to the time constraints the experiment was not repeated.

3.4 Correlation between environmental factors and Bifidobacteria composition

3.4.1 Description of the factors

Seventeen factors which might potentially have an impact on the development of the *Bifidobacterium* spp. were selected. Correlations of *Bifidobacterium* species' abundance measured at one time point, and those of the changes in their abundance between two subsequent time points to these factors were calculated. To minimize the risk of statistical error, only those factors which had no less than 4 individuals in each of the factor-separated groups at any given time point were regarded. Moreover, we omitted factors in which the number of individuals in one factor-separated group was more than twice higher than in the other. There were only six factors which met both of these criteria. Number of samples in each of the time point in various groups tested is presented in Appendix J and K.

Four of the factors with sufficient numbers of samples relate to the mother's health and diet during pregnancy. As for the children's lifestyles, only multivitamin intake and breast-feeding had enough of samples to minimize the risk of statistical error. Information on infant lifestyles was gathered from the questionnaires filled out by parents during pregnancy, and when their children were 6 weeks, 1 year and 2 years old. Both for breast-feeding and multivitamin intake, we only knew that the number of samples in each group was not enough for the groups separation at 6 weeks, but sufficient enough for the later separation (1 year). Taking into account that stool samples were taken at 10 days, 4 months 1 and 2 years age, we assumed that the change was introduced somewhere after 4 months and therefore only the time points of 1 and 2 years were considered.

The cross-correlation between every pair of factors was checked using the Pearson's chi-square test (Appendix I). Generally, the factors were not cross-correlated. However, the positive correlation between vegetable consumption and common-cold-infections during pregnancy were detected: 15 out of 17 women who reported common-cold-infections ate vegetables more than 4 times a week, whereas 18 out of 36 which didn't report such infections, ate vegetables less regularly ($p = 0.007$).

3.4.2 Fish oil supplement during pregnancy

Relative both to *Bifidobacterium spp.* and to total bacterial load, mothers receiving a fish oil supplement during pregnancy did not show any significant correlation to the composition of *Bifidobacterium* species at any given time point. However, the change in the abundance of *B. longum* and *B. breve* over time was significantly different in infants whose mothers ingested fish oil during their pregnancy.

Relative both to *Bifidobacterium* spp. and to the total bacterial load, *B. breve* showed a higher increase from 10 days to 4 months of age in a group of children whose mothers consumed a fish oil supplement during pregnancy. Relative to the total bacterial load, the change in the abundance of *B. longum* during the same period of time was negative in the group of children whose mothers ingested a fish oil supplement, and positive in the group of children whose mothers did not. Relative to *Bifidobacterium* spp., however, *B. longum* levels showed a reduction in both groups, but the decrease was higher in infants of mothers with fish oil-supplemented diets (Table 3.7).

Table 3.7 Average change in the abundance of MCR-predicted *Bifidobacterium* species in infants from 10 days to 4 months of age

MCR-predicted bacterial species, relative to[1)]		Average change , % (SEM) [2)]	
		Taking fish oil supplement during the pregnancy	
		Yes	No
B. breve	B*[3)]	42.79 (14.28)	11.40 (12.39)
	TL*	36.24 (11.19)	11.47 (7.43)
B. longum	B*	-38.85 (11.66)	-5.66 (13.08)
	TL*	-15.37 (9.35)	9.37 (9.66)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) * $0.01 \leq p \leq 0.05$

3.4.3 Consumption of vegetables during pregnancy

There were no observed differences in the relative levels of *Bifidobacterium* spp. in the samples that could be correlated to whether pregnant mothers ate vegetables on a nearly daily basis (four times a week or more), or not. However, relative to the total bacterial load, the abundance of *B. breve* was higher in a group of newborns (2-4 days old) whose mothers did not eat vegetables regularly during their pregnancy (Table 3.8). The same was true for the *B. bifidum* percentage levels in 10 day-old individuals.

As for the development over time, there were many samples belonging to infants of mothers who ate vegetables regularly, but only few of those, who were not. Therefore no information on significant difference in change over time between two groups of individuals could be deduced.

Table 3.8 Average abundance of MCR-predicted *Bifidobacterium* species relative to the total bacterial load, in infants whose mothers did or did not eat vegetables more than 4 times a week, respectively

MCR-predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)]	
		Mothers ate vegetables more than 4 times a week	
		Yes	No
2 – 4 days			
*B. breve**	TL*[3)]	0.78 (0.40)	7.87 (4.27)
10 days			
*B. bifidum**	TL*	1.72 (0.57)	4.11 (1.26)

1) TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) * $0.01 \leq p \leq 0.05$

3.4.4 Reported common cold infection during pregnancy

The reporting of common cold infection during pregnancy exhibited correlation to both mothers' and children's composition of *Bifidobacterium* species. The abundance of *B. adolescentis* relative both to *Bifidobacterium* spp. and to the total bacterial load was significantly lower in mothers who reported common cold infection during their pregnancy (Table 3.9). Also, relative to the total bacterial load, mothers who reported having had a cold during pregnancy showed correspondence with their newborn infants regarding the abundance of *B. adolescentis* (Table 3.10).

B. breve was less abundant in the group of 10-day-old children whose mothers reported having had a cold. However, at four months, *B. breve* levels were higher in the same group (Table 3.9). Interestingly, the change in the relative abundance of *B. breve* was significantly different between the two groups of children (whose mothers did or did not reportedly have a cold during their pregnancy) from birth up to four months. Shortly after birth, from 2 – 4 to 10 days of age, *B. breve* levels decreased in the cold group, while showing an increase in the healthy group both relative to *Bifidobacterium* spp. and total bacterial loads. As for the period from 10 days to 4 months of age, the relative abundance of *B. breve* increased in both groups, though the extent was higher in the group of children whose mothers reportedly suffered from

a cold during pregnancy (Table 3.10). Relative both to total bacterial and *Bifidobacterium* spp. load, even at 2 years of age, the abundance of *B. breve* was significantly lower in the group of children, whose mothers reportedly had a cold during pregnancy (Table 3.9).

Table 3.9 Average abundance of MCR-predicted *Bifidobacterium* species in infants whose mothers did/did not report having had a cold during pregnancy.

MCR-predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)]	
		Mothers reported a cold during pregnancy	
		Yes	No
Early pregnancy			
B. adolescentis	B**[3)]	31.38 (12.09)	55.65 (5.55)
	TL*	0.83 (0.40)	2.16 (0.42)
Late pregnancy			
B. adolescentis	TL*	0.67 (0.27)	2.94 (1.32)
2 – 4 days			
B. breve	B*	3.87 (2.01)	26.67(8.28)
B. longum	B*	76.61(10.84)	41.71(10.58)
10 days			
B. bifidum	B**	15.25 (3.73)	4.94 (1.59)
	TL*	3.96 (1.34)	1.43 (0.60)
B. breve	B**	9.06 (6.35)	40.34 (11.64)
	TL**	4.46 (2.90)	22.84 (7.14)
B. longum	B*	62.77 (8.64)	35.84 (10.38)
4 months			
B. breve	TL*	44.97 (9.08)	22.87 (5.38)
2 years			
B. breve	B*	2.16 (0.57)	9.23 (2.64)
	TL**	0.1 (0.06)	1.06 (0.48)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

Mothers reporting having had a cold during pregnancy correlated to elevated levels of the *B. longum* group relative to *Bifidobacterium* spp., both during late stages of pregnancy and during early days of infants' lives (Table 3.9). *B. bifidum* was also elevated in 10 day-old babies whose mothers reportedly had suffered from a cold. Relative to the total bacterial load, the change in the relative abundance of *B. dentium* from 10 days to 4 months of age was negative in infants whose mothers had reported a cold and positive in those who had not (Table 3.10).

Table 3.10 Average change in the abundance of MCR-predicted *Bifidobacterium* species in infants whose mothers did/did not report having had a cold during pregnancy

MCR-predicted bacterial species, relative to[1)]		Average change, % (SEM)[2)]	
		Mothers reported a cold during pregnancy	
		Yes	No
From late pregnancy to a newborn			
B. adolescentis	TL*[3)]	-0.77 (0.28)	-2.92 (1.52)
From 2 – 4 days to 10 days			
B. breve	TL**	-0.66 (0.66)	26.00 (7.87)
	B*	-3.46 (2.57)	23.40 (10.10)
From 10 days to 4 months			
B. breve	TL**	40.92 (11.28)	9.82 (8.58)
B. dentium	TL*	-0.34 (0.44)	2.21 (1.34)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

A summary of the correlation between mothers reporting having had a cold during pregnancy to the relative composition and change in time of MCR-predicted *Bifidobacterium* species of mothers and their infants is presented in Figure 3.13.

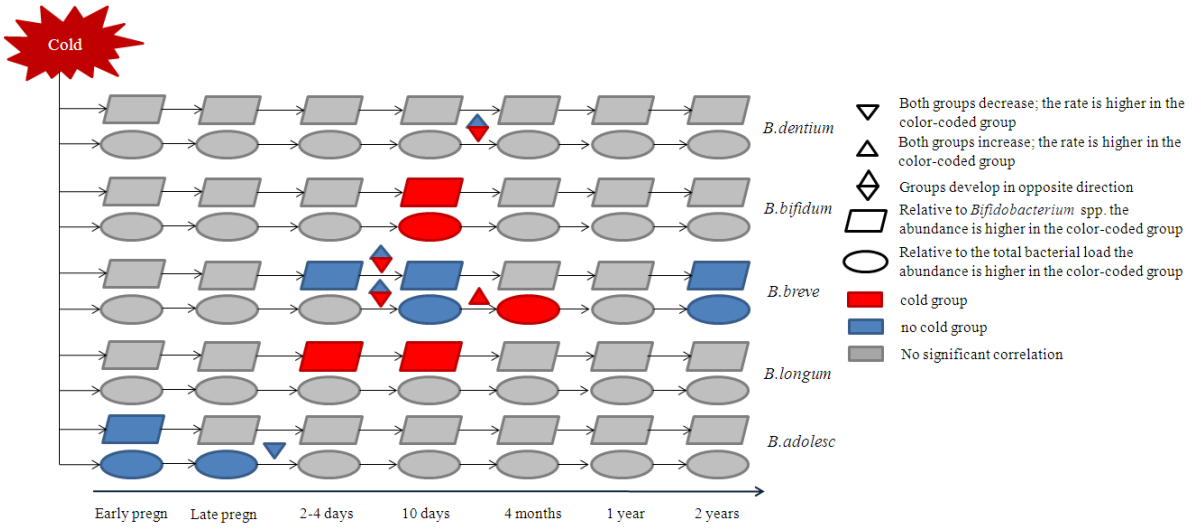

Figure 3.13 Correlation of mothers reporting having had cold during pregnancy to the abundance and the development of Bifidobacterium species in stool samples of mothers and their infants. Two groups of infants and their mothers are color-coded as: red – reported a cold during pregnancy; blue – did not report a cold during pregnancy. Time points, where no significant difference was observed, are colored grey. Colored parallelograms represent groups with significantly higher abundance relative to Bifidobacterium spp.; ellipses – relative to the total bacterial load. Triangles depict the situation, when the observed difference in the change between the two groups was significant. Triangle pointing to the top (or bottom) means that the abundance increased (or decreased) in both groups, but the rate was higher in a colored group. The combination of two triangles – two groups developed in opposite directions. The orientation of the triangle shows the direction of the change in both groups.

3.4.5 Reported vaginal fungal infection during pregnancy

For at least one time point, all MCR-predicted *Bifidobacterium* species in mothers and their infants seemed to correlate to whether mothers reported having had a vaginal fungal infection during pregnancy. The correlation was strongest relative to the total bacterial load, whereas relative to *Bifidobacterium* spp., there was only one case when a significant difference between the two groups was detected; 2-year-old children, whose mothers had reported a vaginal fungal infection, had a lower abundance of *B. bifidum* compared to the group of children whose mothers were reportedly healthy during pregnancy (Table 3.11). The change in the relative abundance of these species also developed in opposite directions in the two groups of infants from 1 to 2 years age. Relative to the total bacterial load, the difference in the change of *B. bifidum* from 10 days to 4 months of age was observed (Table 3.12). During the same period, the abundance of *B. breve* increased in all children, however, the rate of increase was higher in infants whose mothers had reported a fungal infection (Table 3.12).

Table 3.11 Average abundance of MCR-predicted *Bifidobacterium* species in infants whose mothers did/did not reportedly suffer from a vaginal fungal infection during pregnancy.

MCR-predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)] Mother reported a vaginal fungal infection during pregnancy	
		Yes	No
		Late pregnancy	
B. adolescentis	TL**[3)]	3.44 (0.60)	0.67 (0.35)
		2 – 4 days	
B. dentium	TL*	1.89 (0.90)	0.23 (0.16)
		10 days	
B. longum	TL**	39.11 (11.64)	12.68 (4.27)
		1 year	
B. longum	TL**	1.61 (0.63)	5.69 (1.78)
All *Bifidobacterium* spp.	TL*	4.18 (0.99)	10.52 (2.24)
		2 years	
B. bifidum	B*	16.90 (4.77)	30.15(3.49)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

Fungal infections during pregnancy also correlated with elevated levels of *B. adolescentis* during the late pregnancy stage and with *B. dentium* levels in newborn babies (Table 3.11). During pregnancy, the abundance of *B. dentium* relative to *Bifidobacterium* spp. decreased in mothers who had reportedly suffered from infection, and increased in those, who had not

(Table 3.12). Relative both to *Bifidobacterium* spp. and to the total bacterial load, the abundance of *B. adolescentis* in newborns seemed to resemble that of mothers more in the reportedly healthy group compared to the reportedly infected group.

Table 3.12 Average change in the abundance of MCR-predicted *Bifidobacterium* species in infants whose mothers did/did not reportedly suffer from a vaginal fungal infection during their pregnancy.

MCR predicted bacterial species, relative to[1)]		Average change, % (SEM)[2)] Mother was reportedly suffering from vaginal fungal infection during pregnancy	
		Yes	No
From early to late pregnancy			
B. longum	TL*[3)]	0.50 (0.37)	-0.24 (0.11)
All *Bifidobacterium* spp.	TL*	2.34 (2.09)	-0.66 (0.71)
B. dentium	B*	-2.81 (1.15)	1.20 (1.76)
From late pregnancy to a newborn			
B. adolescentis	TL*	-2.99 (1.45)	-0.70 (0.43)
	B*	-62.79 (2.87)	-33.94 (18.31)
All *Bifidobacterium* spp.	TL*	19.70 (9.03)	2.99 (1.34)
From 2 – 4 days to 10 days			
B. longum	TL**	38.04 (13.93)	2.31 (6.60)
All *Bifidobacterium* spp.	TL*	42.82 (10.06)	20.81 (7.26)
From 10 days to 4 months			
B. breve	TL*	38.71 (10.57)	8.70 (9.07)
B. bifidum	TL*	1.60 (1.86)	10.04 (4.55)
B. longum	TL**	-26.63 (11.68)	21.05 (9.33)
	B**	-45.95 (12.35)	5.03 (14.57)
From 1 to 2 years			
B. longum	TL**	3.84 (2.65)	-3.69 (1.81)
All *Bifidobacterium* spp.	TL*	5.86 (2.60)	-3.74 (2.93)
B. bifidum	B*	-12.67 (8.04)	3.08 (5.02)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

Noticeably, reports of vaginal fungal infection mostly correlated to the change in the abundance of both all *Bifidobacterium* and *B. longum* species. All together, significant differences were observed in every period over time except for the change from 4 months to one year (Table 3.12). Changes in the abundance of *Bifidobacterium* spp. in general, and *B. longum* in particular, in mothers who had reported a fungal infection during pregnancy were opposite to those observed in the reportedly healthy group. The same was true for the development of these bacteria in infants from 1 to 2 years of age. In one-year-old children, the relative levels of all *Bifidobacterium* spp. strains and *B. longum* were lower in the fungal

group (Table 3.11). During early infancy (from 3 to 10 days), *B. longum* increased in its abundance in both groups of children, but the rate was higher in the fungal group. Also, 10-day-old children, whose mothers reported having had a fungal infection during pregnancy, exhibited higher levels of *B. longum* relative to the total bacterial load. Moreover, the development of *B. longum* from 10 days to 4 months showed opposite trends in the two groups of children.

The correlation of having reported a vaginal fungal infection during pregnancy to the abundance and development of MCR-predicted *Bifidobacterium* species is summarized in Figure 3.14.

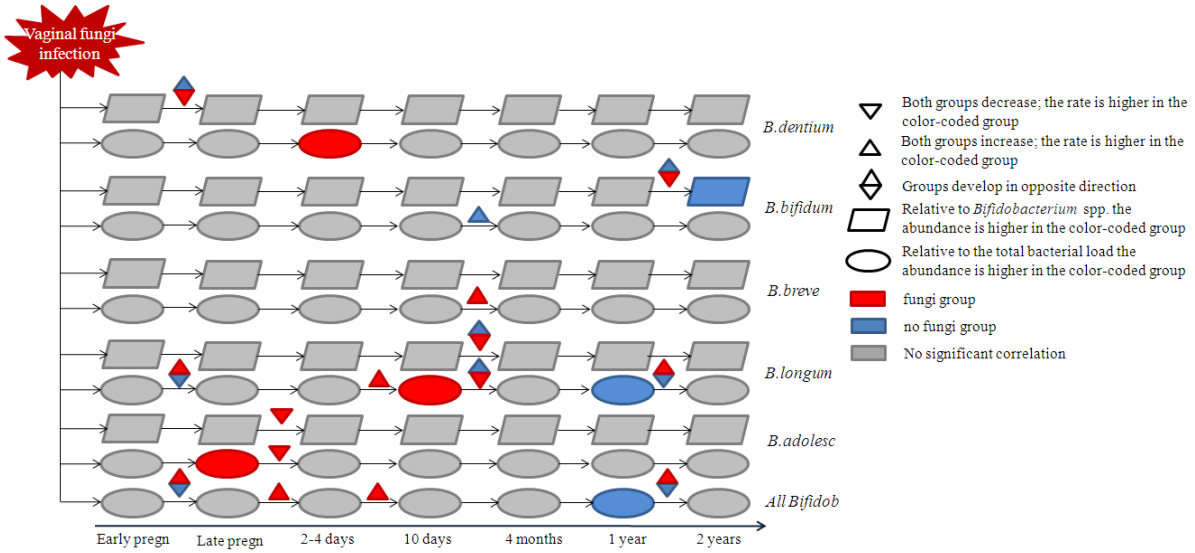

Figure 3.14 Correlation of reportedly having had a vaginal fungal infection during pregnancy to the abundance and the development of MCR-predicted Bifidobacterium species in stool samples of infants and their mothers. Two groups of infants and their mothers are color-coded as: red – reported a fungal infection during pregnancy; blue – did not report a fungal infection during pregnancy. Time points, where no significant difference was observed, are colored grey. Colored parallelograms represent groups with significantly higher abundance relative to Bifidobacterium spp.; ellipses – relative to the total bacterial load. Triangles depict the situation, when the observed difference in the change between the two groups was significant. Triangle pointing to the top (or bottom) means that the abundance increased (or decreased) in both groups, but the rate was higher in a colored group. The combination of two triangles – two groups developed in opposite directions. The orientation of the triangle shows the direction of the change in both groups.

3.4.6 Breast-feeding throughout the first year of life

Relative to the total bacterial load, levels of *B. bifidum* were elevated in those 1- and 2-year-old children, who were breast-fed through the whole first year of life (Table 3.13). However, relative to *Bifidobacterium* spp., the only significant difference was observed in 1-year-old children. *B. breve* and *B. dentium* also had higher counts in the breast-fed group of 2-year-old

children with regards to total microbial load. The *Bifidobacterium* spp. group was more abundant in 1-year-old breast-fed children.

No significant difference in the change of *Bifidobacterium* species' abundance from 1 to 2 years depending on the breast-feeding status was detected. A summary of the correlation of breast-feeding to the relative abundance of bifidobacterial species is presented in Figure 3.15.

Table 3.13 Average abundance of MCR-predicted *Bifidobacterium* species in infants who were/were not breast-fed throughout the first year of life

MCR predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)] Breast-fed through the whole first year of life	
		Yes	No
		1 year	
B. bifidum	TL**[3)]	4.79 (2.34)	0.58 (0.20)
	B*	28.86 (7.12)	15.75 (4.33)
All *Bifidobacterium* spp.	TL*	13.18 (4.21)	5.23 (0.95)
		2 years	
B. bifidum	TL**	3.70 (2.15)	0.83 (0.25)
B. breve	TL*	1.22 (0.70)	0.16 (0.08)
B. dentium	TL*	0.33 (0.18)	0.10 (0.04)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

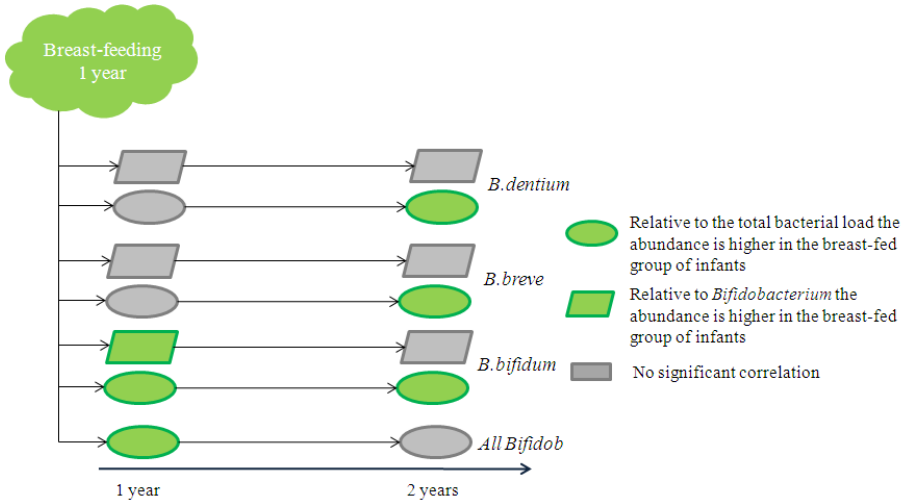

Figure 3.15 Correlation of breast-feeding throughout the first year of life to the abundance of MCR-predicted Bifidobacterium species in stool samples of infants. The group of infants who were breast-fed throughout the first year of life is color-coded green. Time points, where no significant difference was observed, are colored grey. Colored parallelograms represent time points where significantly higher abundance was observed in individuals belonging to the breast-fed group relative to Bifidobacterium spp.; ellipses – relative to the total bacterial load.

3.4.7 Receiving a multivitamin supplement during the first year of life

Relative to the total bacterial load, no correlation between vitamin intake and bifidobacterial composition was observed, whereas relative to *Bifidobacterium* spp., both 1- and 2-year-old children who had received multivitamins exhibited elevated levels of *B. bifidum* and decreased levels of *B. longum* (2-year-old children) (Table 3.14). However, abundances were distributed evenly along the abundance axis regardless of multivitamin intake (Appendix J). Surprisingly, relative both to *Bifidobacterium* spp. and to the total bacterial load, levels of *B. breve* were higher in those 10-day-old babies who had received multivitamins compared to those who had not (36.25 % and 4.43 % for yes- and no-groups relative to the total bacterial load, and 64.24 % and 9.17 % for yes-and no-groups relative to *Bifidobacterium* spp.; $p < 0.001$ in both cases; Appendix J). For the period from age 1 to 2 years, having received a multivitamin supplement correlated to changes in both the amount of *B. adolescentis* relative to *Bifidobacterium* spp. and *B. longum* relative to the total bacterial load (Table 3.15).

Table 3.14 Average abundance of MCR-predicted *Bifidobacterium* species

MCR predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)] Received a multivitamin supplement during the first year	
		Yes	No
		1 year	
*B. bifidum**	B*[3)]	30.99 (7.04)	17.55 (4.82)
		2 years	
*B. bifidum**	B*	32.89 (4.13)	20.52 (4.79)
*B. longum**	B*	28.45 (7.21)	50.49 (7.54)

1) B – relative to *Bifidobacterium* spp.;
1) SEM – standard error of the mean;
2) *$0.01 \leq p \leq 0.05$

Table 3.15 Average change in the abundance of MCR-predicted *Bifidobacterium* species from age 1 to 2 years in infants who were/were not given a multivitamin supplement during the first year of their life

MCR predicted bacterial species, relative to[1)]		Average change, % (SEM)[2)] From 1 to 2 years	
		Received a multivitamin supplement during the first year	
		Yes	No
B. adolescentis	B**[3)]	37.68 (10.37)	6.58 (7.32)
B. longum	TL**	3.80 (4.54)	-3.44 (2.09)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) **$0.005 \leq p \leq 0.01$

A summary of the correlation of multivitamin intake during the first year of life to the abundance and the development of MCR-predicted *Bifidobacterium* species is presented in Figure 3.16.

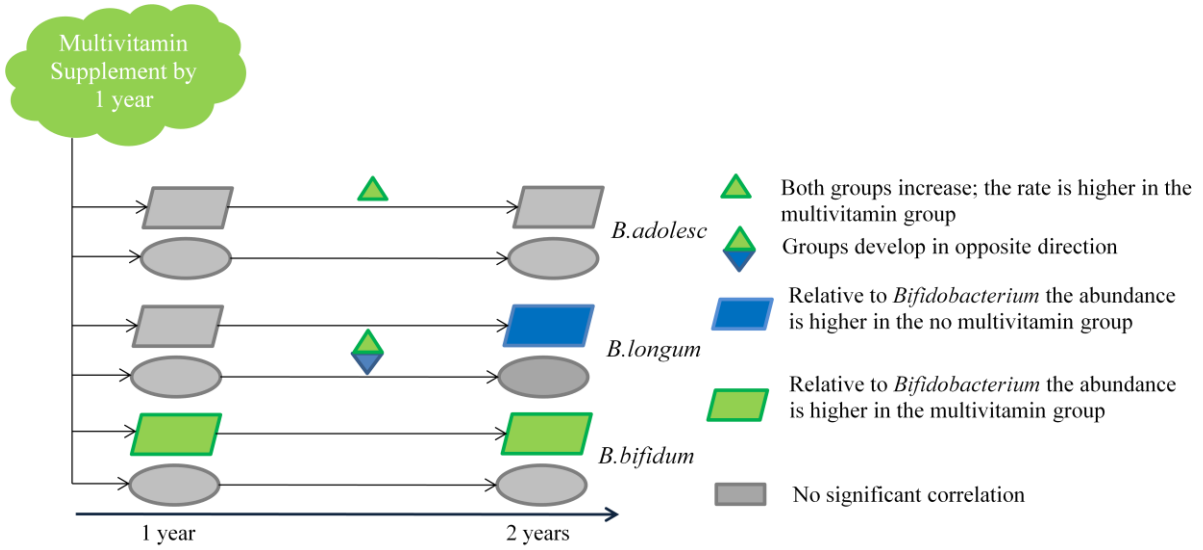

Figure 3.16 Correlation of having received a multivitamin supplement to the abundance and the development of MCR-predicted Bifidobacterium species in stool samples of infants. Two groups of infants are color-coded as: green – had received multivitamins during the first year of life; blue – had not received multivitamins. Time points, where no significant difference was observed, are colored grey. Colored parallelograms represent groups with significantly higher abundance relative to Bifidobacterium spp.; ellipses – relative to the total bacterial load. Triangles depict the situation, when the observed difference in the change between the two groups was significant. Green triangles pointing upwards indicate that the abundance increased in both groups, but the rate was higher in a group of infants who had received multivitamins. The combination of two triangles – two groups developed in opposite directions. The orientation of the triangle shows the direction of the change in both groups.

3.5 Correlation of *Bifidobacterium spp.* composition to allergy and eczema status by two years of age

3.5.1 Description of clinical criteria

The allergy and eczema status of children by the age of two years were compared to the relative amounts of *Bifidobacterium* species in their stool samples. With regard to sensitization, three concentrations of specific IgE serum were considered: IgE ≥ 0.1 kU/l; IgE ≥ 0.35 kU/l and IgE ≥ 0.7 kU/l. Eczema was assessed both using UK Working Party criteria and mothers' answers to a questionnaire asking whether or not the baby had experienced symptoms of eczema dermatitis during his/her first two years of life.

The cross-correlation between two health status criteria of infants was checked using Pearson's Chi-square test. Positive correlation between three IgE levels was detected. Medical diagnosis of eczema correlated to the parents' answers on whether or not the child experienced eczema symptoms (Appendix I). No correlation between the any of the three examined IgE levels and eczema was observed. We also checked for correlation between environmental factors and health status of the infant by the age of two years. There was a positive correlation between breast-feeding of the child throughout the first year of life and his/her serum IgE level by the age of two years (Appendix I).

3.5.2 IgE sensitization

Relative to the total bacterial load, the amount of *B. breve* was slightly higher in mothers whose children exhibited IgE levels ≥ 0.1 kU/l, as well as in 2-year-old children with specific IgE levels of more than 0.35 kU/l (Table 3.16). The amount of *B. bifidum* relative to *Bifidobacterium* spp. was elevated in 10-day-old children with IgE levels under the detection limit. From 1 to 2 years, the amount of this species increased in infants with an IgE level of more than 0.35 and 0.70 kU/l and decreased in a non-sensitized group (Table 3.17). Relative to the total bacterial load, the amount of *B. longum* in stool samples of sensitized newborns resembled that of their mothers' more than in nonsensitized, where it was higher (Table 3.17). *B. adolescentis* levels were elevated in sensitized 4-month-old children with both high and moderate IgE levels relative not only to a *Bifidobacterium* spp., but also to the total bacterial load (Table 3.16). From 10 days to 4 months, the abundance of *B. adolescentis* increased in sensitized (IgE ≥ 0.35 kU/l and 0.7 kU/l), and diminished in non-sensitized children (Table 3.17). Relative to total bacterial load during this period, however, a significant difference was only detected in children with IgE ≥ 0.35 kU/l. From 4 months to 1 year of age, however, the abundance of *B. adolescentis* decreased in sensitized children (IgE ≥ 0.7 kU/l) and increased in nonsensitized individuals.

Table 3.16 Average abundance of MCR-predicted *Bifidobacterium* species in sensitized and nonsensitized children and their mothers

MCR predicted bacterial species, relative to[1)]		Age category	Average abundance, % (SEM)[2)]	
			Specific serum IgE ≥ Threshold	
			Yes	No
Threshold – 0.10 kU/l				
B. breve	TL*[3)]	Late pregnancy	0.04 (0.02)	0.01 (0.01)
B. bifidum	B*	10 days	4.61 (1.41)	12.27 (2.63)
Threshold – 0.35 kU/l				
B. breve	TL*	2 years	1.51 (0.76)	0.30 (0.10)
B. adolescentis	B*	4 months	10.04 (4.89)	2.24 (0.78)
Threshold – 0.70 kU/l				
B. adolescentis	TL*	4 months	4.07 (2.13)	1.32 (0.44)
	B**		12.22 (5.89)	2.12 (0.73)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

Table 3.17 Average change in the abundance of MCR-predicted *Bifidobacterium* spp. in sensitized and nonsensitized children

MCR predicted bacterial species, relative to[1)]		Time period	Average change, % (SEM)[2)]	
			Specific serum IgE ≥ Threshold	
			Yes	No
Threshold – 0.10 kU/l				
B. longum	TL*[3)]	Late pregn – 2-4 days	-0.4 (0.28)	5.99 (1.47)
B. adolescentis	B**	10 days – 4 months	4.35 (2.72)	-5.02 (3.87)
B. breve	B*	10 days – 4 months	1.95 (12.68)	41.88 (11.68)
Threshold – 0.35 kU/l				
B. adolescentis	TL*	10 days – 4 months	2.77 (2.00)	-3.23 (2.64)
	B**		4.98 (3.16)	-4.65 (3.66)
B. longum	TL*	Late pregn – 2-4 days	-0.4 (0.28)	5.98 (1.47)
B. bifidum	B*	1 year – 2 years	6.41 (5.35)	-8.59 (4.46)
Threshold – 0.70 kU/l				
B. adolescentis	B*	10 days – 4 months	6.57 (4.12)	-3.93 (3.12)
	B*	4 months – 1 year	-4.52 (11.96)	15.87 (4.25)
	TL*		-4.98 (3.19)	0.20 (0.71)
B. bifidum	B*	1 year – 2 years	8.79 (10.00)	-8.34 (3.09)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

Relative to the total bacterial load, the abundance of *B. longum* in newborns with IgE ≥ 0.1 kU/l and 0.35 kU/l resembled its abundance in mothers more than in nonsensitized infants (Table 3.17).

A summary of the correlation between the abundance and the development of MCR-predicted *Bifidobacterium* species to specific serum IgE levels is presented in Figure 3.17.

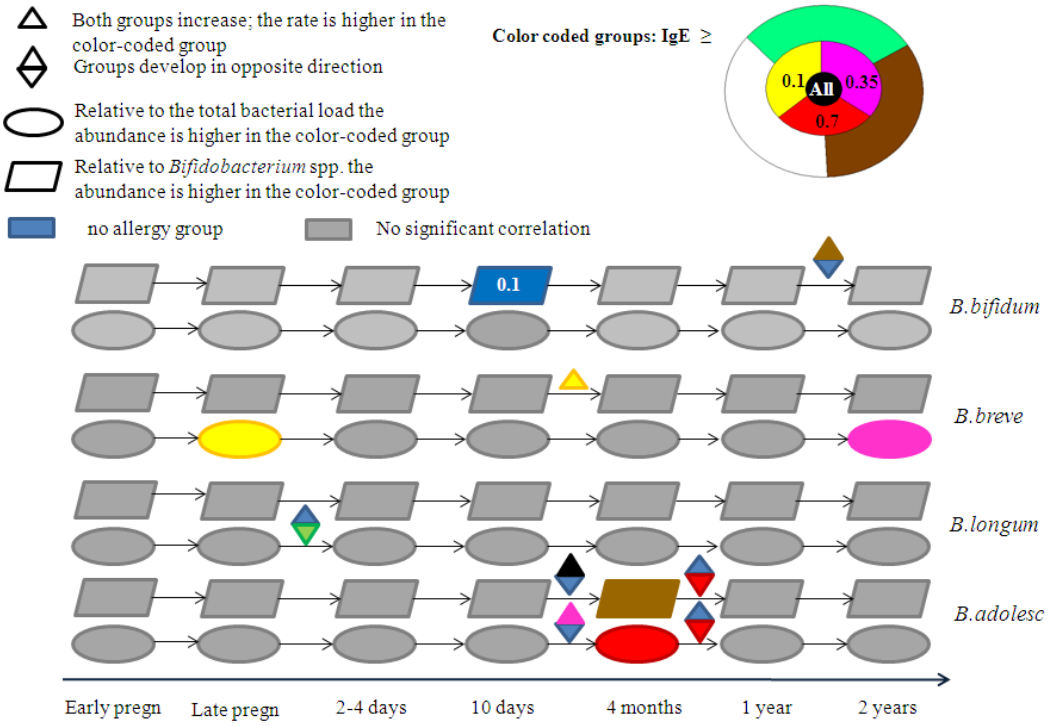

Figure 3.17 Age categories, in which differences in the relative amounts of Bifidobacterium species between sensitized/nonsensitized children and their mothers were significant. Two groups of infants and their mothers are color-coded as: IgE levels higher than 0.1 kU/l, 0.35 kU/l, 0.7 kU/l or the combination of them – see the color-coding circle; blue – no allergy. Time points, where no significant difference was observed, are colored grey. Colored parallelograms represent groups with significantly higher abundance relative to Bifidobacterium spp.; ellipses – relative to the total bacterial load. Triangles depict the situation, when the observed difference in the change between the two groups was significant. Triangle pointing to the top (or bottom) means that the abundance increased (or decreased) in both groups, but the rate was higher in a colored group. The combination of two triangles – two groups developed in opposite directions. The orientation of the triangle shows the direction of the change in both groups.

3.5.3 Eczema status

Relative to the *Bifidobacterium* spp. load, children whose eczema status was confirmed medically had higher amounts of *B. dentium* during the early days of their lives (10 days). Also, mothers of these babies showed elevated levels of *B. bifidum* during late pregnancy (Table 3.18). Relative to the total bacterial load, abundances of *B. dentium*, *B. longum* and all *Bifidobacterium* species were higher in 1-year-old children with medically diagnosed eczema. The relative amount of *B. longum* in stool samples of 10-day-old children also correlated to

the incidence of eczema, though in this case it was higher in the healthy group without eczema (Table 3.18).

Table 3.18 Average abundance of MCR-predicted *Bifidobacterium* species in mothers and their children who by two years of age developed/did not develop eczema as assessed by UK Working Party criteria

MCR predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)]	
		Eczema diagnosed (UKWP)	
		Yes	No
		Early pregnancy	
B. bifidum	B*[3)]	37.32 (2.94)	22.76 (2.96)
		10 days	
B. longum	TL*	6.15 (3.09)	29.09 (5.54)
B. dentium	B*	17.05 (9.70)	2.35 (0.82)
		1 year	
B. dentium	TL*	1.08 (0.15)	0.35 (0.22)
B. longum	TL*	7.01 (3.09)	2.54 (5.54)
All *Bifidobacterium* spp.	TL**	16.88 (4.07)	6.59 (1.33)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

Ten-day-old children with eczema symptoms had higher amounts of *B. dentium* relative to *Bifidobacterium* spp. (Table 3.19). At 4 months, both *B. longum* and *B. adolescentis* were elevated in children with eczema symptoms, whereas *B. breve* was less abundant relative not only to *Bifidobacterium* spp., but also to the total bacterial load (Table 3.19).

Table 3.19 Average abundance of MCR-predicted *Bifidobacterium* species in children showing/not showing eczema symptoms

MCR predicted bacterial species, relative to[1)]		Average abundance, % (SEM)[2)]	
		Eczema symptoms (questionnaire)	
		Yes	No
		10 days	
B. bifidum	B*[3)]	13.75 (2.94)	6.62 (1.92)
		4 months	
B. longum	TL*	32.70 (7.66)	16.90 (3.52)
	B*	55.33 (9.65)	32.01 (5.63)
B. breve	TL*	12.93 (6.48)	31.19 (4.58)
	B*	19.39 (9.34)	50.00 (6.38)
B. adolescentis	TL*	3.89 (1.83)	1.25 (0.45)
	B**	11.68 (5.49)	2.06 (0.73)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

There was no difference detected in the change of *Bifidobacterium* species' abundances in children who had eczema symptoms compared to those who did not. However, at some time points, the changes in the abundance of *B. longum* and *B. dentium* showed opposite trends in children with diagnosed eczema compared to healthy individuals (Table 3.20).

Table 3.20 Average change in the abundance of MCR-predicted *Bifidobacterium* species in children with diagnosed eczema and healthy children

MCR predicted bacterial species, relative to[1)]		Average change, % (SEM)[2)]	
		Eczema diagnosed (UKWP)	
		Yes	No
From 4 days to 10 days			
B. dentium	B**[3)]	5.00 (5.85)	-5.11 (2.03)
B. longum	TL*	-1.74 (8.62)	26.65 (7.41)
From 1 year to 2 years			
B. longum	B*	-20.1 (6.77)	12.04 (7.85)

1) B – relative to *Bifidobacterium* spp.; TL – relative to the total bacterial load;
2) SEM – standard error of the mean;
3) *$0.01 \leq p \leq 0.05$; **$0.005 \leq p \leq 0.01$

A summary of the correlation between the abundance and the development of MCR-predicted *Bifidobacterium* species to diagnosed eczema, and eczema reported by parents, is presented in Figure 3.18.

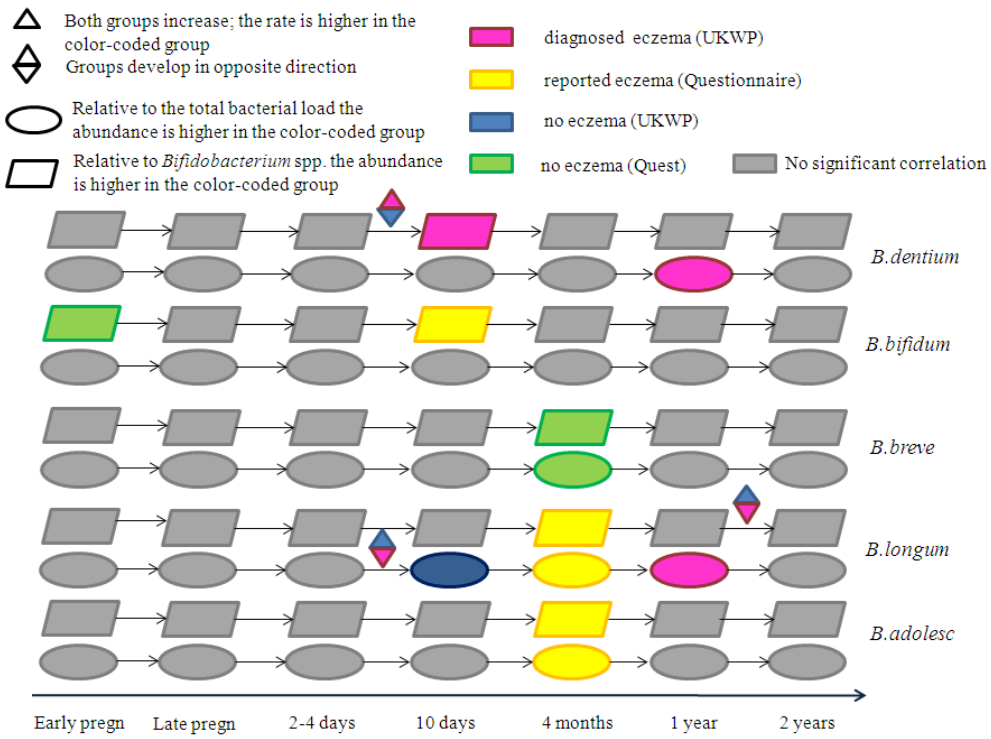

Figure 3.18 Age categories, in which differences in the relative amounts of Bifidobacterium species between mothers and their children who had/had not suffered from eczema was significant. Two groups of infants and their mothers are color-coded as: pink – diagnosed eczema; blue – no eczema diagnosed; yellow – eczema reported by parents; green – no eczema reported by parents. Time points, where no significant difference was observed, are colored grey. Colored parallelograms represent groups with significantly higher abundance relative to Bifidobacterium spp.; ellipses – relative to the total bacterial load. The combination of two triangles depicts the situation when two groups developed in opposite directions. The orientation of the triangle shows the direction of the change in both groups.

4 Discussion

4.1 Identification of dominant *Bifidobacterium* species

4.1.1 Assessment of MCR-ALS for *Bifidobacterium* species resolution

The MCR-ALS method for resolving mixed *clpC* sequence data was verified using 35 predefined mixtures of four *Bifidobacterium* species commonly associated with the human gut. The approach was proven to be a suitable method for identification and quantification of bifidobacterial species in mixtures. However, there was a tendency of overestimating the abundance of *B. adolescentis* and underestimating that of *B. breve*, whereas the abundance of the *B. longum* group was predicted more accurately. The *B. longum* group comprised of two bacteria, *B. longum longum* and *B. longum infantis*, and therefore there was twice as much information about this group in mixed sequences than for *B. breve* and *B. adolescentis*. This may partially explain why the *B. longum* group was better resolved than the other species. During the design and assessment of the method, we encountered many problems with the quality of sequences used for the resolution. If not removed, these sequences caused improper spectra alignment correction and thus hindered accurate MCR-ALS prediction. Instead of removing the samples, we checked the possibility of removing the part of information about the sequences that had caused the abnormality. Not only did this not weaken the predictability of the method, but even slightly improved MCR-ALS performance by reducing the prediction error.

4.1.2 Prevalence of *Bifidobacterium* species in stool samples of infants and their mothers

MCR-ALS identified five predominant components in stool samples of infants and their mothers from Trondheim – *B. adolescentis*, *B. dentium*, *B. breve*, *B. bifidum* and *B. longum*. Confirmation of the identified components in the dataset was acheived by cloning and sequencing of the cloned amplicons, which showed 98 % – 99 % identity with the corresponding GenBank reference sequences. Two MCR-resolved component sequences, identified as *B. bifidum* and *B. dentium*, showed low similarity to their respective reference sequences deposited in GenBank. However taking into account that these species were detected by sequencing cloned inserts from the corresponding samples, we suppose that this

low similarity was caused by noise. The fact that apart from *B. animalis* only those *Bifidobacterium* species, which were predicted by MCR-ALS, were identified by sequencing of cloned inserts, strongly speaks in favor of correct prediction of *Bifidobacterium* species.

We did not identify *B. pseudolongum*, which is normally associated with a human gut. When we searched mixed sequences against NCBI nucleotide database, these species were not detected in any of the samples either. The *clpC* sequences of *B. pseudolongum* (AY722386.1; DQ238019.1; DQ238034.1) deposited in GenBank are short and miss the region of the primer binding, so we were unable to check whether the absence was caused by primer bias. However, the GenBank-deposited sequences were obtained with the same *clpC* primers which we used for this work (Ventura et al., 2006). Therefore primer bias is unlikely to be the reason for the lack of *B. pseudolongum* in the IMPACT dataset. *B. pseudolongum* is believed to be present exclusively in adults (Ventura et al., 2010a). In our dataset, only 24 % of all samples (78 out of 330) represented stool samples from adults, therefore the information about this species could have been lost during MCR-ALS analysis, as it probably was not present in 76 % of the analyzed samples. Interestingly, we also did not detect *B. pseudocatenulanum* – the species which was shown to be present in the human gut irrespective of the age (Turroni et al., 2009). However, in the Turroni study of 59 healthy individuals of different age, the prevalence of *B. pseudocatenulanum* varied from 3 % to 14 % depending on the age and the sample type (stool or mucosal samples). The MCR method searches for components that are common for all the samples in the dataset, whereas other information is regarded as noise. Therefore, the species of *B. adolescentis*, *B. breve*, *B. longum*, *B. bifidum* and *B. dentium* most likely comprise the most commonly identified *Bifidobacterium* species in the gut, while others (like *B. animalis*, which was detected by sequencing all cloned amplicons generated from one sample), though still being present, and even possibly being present at high concentrations, are not shared in the cohort population. This can also be supported by the fact that only 57 % of the variation among the samples is explained by MCR, leaving the rest of it to other species, present in some, but not all individuals.

B. dentium, mostly isolated from oral cavities and believed to be associated with dental caries (Tannock, 2010), was identified as one of five the most common components of *Bifidobacterium* group among all 83 individuals examined in this study. However, stool samples represent the luminal microbial consortium of the human gut. So, in the study of

Turroni et al., they detected *B. dentium* in stool, but not in mucosal samples, suggesting that this species most likely constitutes part of the transient microbiota (Turroni et al., 2009).

4.2 Average composition of *Bifidobacterium* species in infants and their mothers and the interactions revealed

We calculated average abundances of the five most common *Bifidobacterium* species both within the *Bifidobacterium* spp. group and relative to the total microbial load. It has been shown by Palmer et al., that the total gut microbiota undergoes crucial changes in its development up to the age of 1 year, and then it remains stable over time and resembles that of the adult (Palmer et al., 2007). Strikingly, in our data, there were still high changes in bifidobacterial composition between 1 and 2 years of age, though development was directed towards an adult profile. Palmer et al. detected low prevalence of *Bifidobacterium* species in stool samples of infants. However, for our study population, the presence of *Bifidobacterium* species at all ages was confirmed both by the analysis of 16S rRNA and *clpC* genes.

B. dentium was one of the most stable species and its relative load remained nearly the same regardless the age. If *B. dentium* is commonly present in the mouth, then it would most likely be transferred constantly to the intestines with food and then excreted. The levels of *B. dentium*, however, suggest some growth of this species in the gut lumen.

A majority of the adult bifidobacterial load consisted of *B. adolescentis*. This goes in accordance with the fact that *B. adolescentis* is predicted to have a plant-derived oligosaccharides fermentation machinery (Ventura et al., 2007a). In newborns, the average levels of *B. adolescentis* were much lower than in adults, and this species constituted just a 20^{th} of the whole *Bifidobacterium* spp. load up to 1 year of age, when it then started increasing in abundance towards an adult profile. However, relative to the total bacterial load, by the age of 10 days the percentage of *B. adolescentis* reached the adult level and then remained rather stable over time regardless of the change in the total *Bifidobacterium* spp. group. Therefore the commonly accepted view of *B. adolescentis* as characteristic of an adult gut microbiota (Haarman M., Knol J., 2005) might possibly be the result of comparing the abundances of the species relative to other *Bifidobacterium* spp. rather than to the total gut microbiota.

B. breve was almost exclusively found in infants. *B. breve* is regarded to be an important species during weaning, as amylopullulanase, the enzyme responsible for starch degradation, is encoded in its genome (Ventura et al., 2007b). The peak of *B. breve* abundance was found in 4-month-old infants, where it comprised one fifth of the total bacterial load. By the age of 1 year it diminished and by 2 years, *B. breve* was nearly absent.

Despite the fact that *B. bifidum* is predicted to possess lacto-N-biosidase and galacto-N-biosidase activity, enabling it to ferment HMOs (Wada et al., 2008), its relative abundance compared to other *Bifidobacterium* spp. groups was higher in adults and 2-year-old children than in newborns and 4-month-old infants. However, relative to the total bacterial load, this group of bacteria exhibited peak abundance at 4 months, when all infants were breast-fed, and became nearly absent in 2-year-olds and adults. Therefore, *B. bifidum* seems to be less affected by the total diminishing of *Bifidobacterium* spp. group than other species of this genus.

The *B. longum* group was one of the two most abundant groups of *Bifidobacterium* species in 1- and 2-year-old children and the third most abundant group in adults. In newborns *B. longum* comprised most of the *Bifidobacterium* spp. load. Compared to other *Bifidobacterium* species, *B. longum* showed the greatest correlation to outer bacterial groups as revealed by LSA, whereas *B. breve*, *B. dentium* and *B. bifidum* seemed to form a co-correlating triangle. Both *B. longum* and *Lactobacillales* were predicted to have a positive correlation to the *Clostridium* group, which may point to their interaction with these bacteria. Falony et al. (2006) studied the co-cultures of *B. longum* BB536 with *Anaerostipes caccae* or *Roseburia intestinalis*, both belonging to the *Clostridium* cluster. They have shown that when cultured on oligofructose as a sole carbon source, one of the *Clostridium* cluster species could not ferment oligofructose, whereas another was growing slowly. However, when *B. longum* was present in the system, both bacteria grew well by utilizing the acetate released by *B. longum*. The correlation detected in the work presented here, however, showed a comparative delay of three time points indicating that, apart from the feeding cross-pathway, the presence of *B. longum* during a mother's pregnancy may by some means prepare the infant's gut for the colonization with *Clostridium* species. Also, *B. longum* negatively correlated to *Proteobacteria*. We propose that this could be explained not by interactions between those two bacteria, but instead by the change in the environmental conditions. *Proteobacteria* are predominantly found among newborns, and then, when the oxygen in the gut becomes

eliminated, *Proteobacteria* become outcompeted by strict anaerobes such as *Bifidobacterium* spp. It is worth noting, that the correlations revealed by LSA seem to be self-supportive, as bacterial groups which were found to be correlated with each other, had a similar correlation to a third bacterial group. Apart from the interaction between the *Lactobacillalles*, *B. longum* and *Clostridium* groups, this can also be demonstrated by the correlation between the abundances of *B. breve*, *B. dentium* and *Faecalibacterium*; between the *B. longum*, *B. breve* and *Bacilli* groups and between the *B. longum*, *Bacilli* and *Proteobacteria*.

4.3 *B. longum* group resolution

Despite the fact that *B. longum longum* is normally characteristic of an adult rather than of an infant gut, this species was detected in the majority of infants at all ages. Right after the birth, *B. longum* group was identified in 25 out of 32 children, and all sequences, apart from two that were unclassified, exclusively belonged to *B. longum longum*. The prevalence of the *B. longum infantis* group was quite low as it was detected only in 13 children, out of which 9 were 4 months old. *B. longum infantis* is normally associated with a healthy infant microbiota and is believed to reduce the risk of allergy development as it suppresses pro-inflammatory IL-17 and induces anti-inflammatory IL-10 production (Miyauchi et al., 2010). In our study, we did not find any correlation between the presence of *B. longum infantis* in the stool samples and allergy development. However, we only used IgE levels as an indicator of allergy and not the IL-17/IL-10 ratio.

Surprisingly, all sequences obtained from stool samples with the specific *B. longum* group primer were pure. In those children, who had *B. longum infantis* at 4 months, *B. longum longum* was detected earlier and later in life, suggesting a competition between the two species. *B. longum longum* is more suited for fermentation of plant-derived oligosaccharides, whereas *B. longum infantis* harbors breakdown machinery for milk-derived oligosaccharides (LoCascio et al., 2010). To check whether the reason we could not detect the co-existence of these two subspecies had been caused by infant diet, we incubated co-cultures of the two bacteria using lactose or lacto-N-biose as a single carbon source. The results, however, were inconclusive and the experiment has to be repeated with more stringent anaerobic conditions and better control of cell mixture concentrations.

4.4 Correlation to environmental factors

To minimize the risk of false discovery, we assumed that if the development of some bacterial species is truly dependent on one or another factor, then it would be logical to expect a significant correlation of that factor to the abundance, to the change in the abundance or to both during some period of time rather than only at one time point. Also, the distribution of bacterial loadings can indicate whether or not the difference is real – if, graphically, the percentage represents a continuous line and both groups seem to be distributed evenly along this line, it is likely that even though the difference between two groups was found statistically significant, it might have no biological meaning.

Factors related to the reported health status of the mother during pregnancy had the greatest correlation to the percentage and development of *Bifidobacterium* species. Whether or not the mother reported having had a cold during pregnancy seemed to correlate with significantly different levels of *B. adolescentis*, *B. breve* and *B. longum* at a given time point, whereas reporting of a vaginal fungal infection showed the greatest correlation to the change in the abundance of *B. longum* and all *Bifidobacterium* spp. group over time.

People with weakened immune systems are more likely to exhibit symptoms of common cold infections (Cohen, S., 1996). IgA is the main immunoglobulin found in the mucous secretion of intestinal and respiratory tract (Wood, 2006) which prevents the penetration of pathogenic microorganisms by preventing their adsorption to mucosal epithelium. In line with this, *B. breve*, which is known for inducing serum IgA production (Yasui et al., 1992), was detected in significantly higher counts in the group of newborns whose mothers had not reported cold. Interestingly, both in 3- and 10-day-old infants, there was a clear separation between *B. breve* abundances in the two groups. It was especially evident at 10 days of age when almost all the infants of mothers who had reported a cold, had low *B. breve* counts, while its abundance in healthy group infants was high. Its increase in abundance from 10 days to 4 months was also higher in infants whose mothers had reported a cold. And then, by the age of 4 months, the amount of *B. breve* became significantly higher in this group of infants. However, at 2 years of age, relative amount of *B. breve* was higher in children of healthy mothers. The significant difference in the amount of *B. breve* at 2 years of age might indicate a long-term impact of a mother's health on the gut microbiota of her child later in life. Taking into account the number of various time points where the significant correlation to the incidence of a reported

common cold was detected, it seems plausible that there is some mechanism behind, which may be connected both to the severity of cold symptoms and to the development of certain bifidobacterial species.

For the most part, reported vaginal fungal infection correlated to the change in the abundance of *Bifidobacterium* species. Interestingly, out of 2 time points and 4 periods over time, during which the significant difference in *B. longum* group levels was detected between individuals, the all *Bifidobacterium* spp. group showed the same pattern at 1 time point and 3 periods over time. During pregnancy, both *B. longum* and all *Bifidobacterium* spp. groups increased in those mothers who reported having had a fungal infection and decreased in those, who had not. The same was observed in infants between 1 and 2 years of age. Interestingly, both *B. longum* and all *Bifidobacterium* spp. were lower in 1-year-old infants of mothers who had reported an infection. Also, both *B. longum* and all *Bifidobacterium* spp. groups increased to a higher extent in 3- to 10-day-old newborns whose mothers reported a fungal infection during pregnancy. The parallel correlation of *Bifidobacterium* spp. and *B. longum* to a reported fungal infection in children of various ages and the absence of this correlation to other bifidobacterial species might indicate that the change detected in all *Bifidobacterium* spp. was actually caused by the particular change in *B. longum*. However, during these developmental stages, *B. longum* does not constitute the majority of *Bifidobacterium* species in the gut. Moreover, even though at 10 days of age, the *B. longum* group constituted the majority of *Bifidobacterium* species, the abundance of *B. longum* in 10-day-old infants correlated to the reported incidence of vaginal fungal infection during pregnancy, but of all *Bifidobacterium* species did not. Similarly, both the abundances of all *Bifidobacterium* spp. and of *B. adolescentis* groups in newborns resembled that of the mother more in the no-infection group, whereas for *B. longum* a significant difference was not detected. Therefore the significant difference in the change of all *Bifidobacterium* spp. may mostly be explained by the change in the *B. longum* group, but it is also compensated by other bifidobacterial species to some extent. Vaginal fungal infection normally refers to *Candida* spp., a fungus which is a commensal intestinal resident, but causes an infection in vagina if the conditions, such as acidity and sugar content, are right (Janssen EMEA, 2011). In general, bifidobacterial flora is known to antagonize fungal infection (Romani, 2004). The observation that vaginal fungal infection correlated with changes in the abundances of bifidobacterial species over time might indicate that the immune system stimulates the growth of healthy microbiota as a response to the fungal infection. Unfortunately, we do not possess information on the treatment of women

who reportedly suffered from fungal infection. However, it is possible that pregnant women were advised to use vaginal suppositories containing lactobacilli and bifidobacteria.

Both breast-feeding and multivitamin intake during the first year of life seemed, generally, to lead to elevated bifidobacterial counts in 1- and 2-year-old children. Intuitively, these results seem logical as all of the dominant bifidobacterial species are present in breast milk (Grönlund et al., 2007) and riboflavin, vitamin B2, is required for the growth of *Bifidobacterium* spp. (Rossi, Amaretti, 2010). However, in the case of breast-feeding, the significant difference seems to be caused only by 2 individuals who had high counts, whereas other infants had comparable levels of this species. As for multivitamin intake, the abundances of *B. bifidum* and *B. longum* were evenly distributed, regardless to which group infants belonged. Therefore, further confirmation of these results is required.

We observed a very highly significant correlation between the percentage of *B. breve* at 10 days and whether or not the children were given multivitamins later in life, with a very clear separation of the two groups. Obviously, the idea that a low *B. breve* percentage at 10 days would, in any way, make an impact on parents' decision to give a multivitamin supplement to an infant few months later in life, is ridiculous. Most likely, this separation is caused not by the vitamin intake, but by some other factor, which coincidentally was observed in those children who were later given vitamins, and was absent in those who were not. We also detected a correlation between *B. breve* and *B. bifidum* abundance in infant guts and whether or not the mother had eaten vegetables regularly during pregnancy. However, this correlation was detected only at one age. Moreover, the abundances of *B. bifidum* in stool samples from infants formed an even distribution regardless mothers' vegetable consumption. As for *B. breve*, even though it was nearly absent in infants whose mothers had been consuming vegetables on a nearly daily basis, whereas present in the other group, this separation might have actually been caused not by vegetable consumption, but by whether or not the mother reported having a cold during pregnancy, as these two factors were found cross-correlated to each other.

4.5 Correlation to allergy and eczema development

The concentration of specific IgE in the serum was used for allergy assessment as, currently it is the most reliable immunological indicator (Størro et al., 2011). Three concentrations of IgE were considered – IgE ≥ 0.1 kU/l, 0.35 kU/l and 0.7 kU/l. Normally, the concentration of IgE ≥ 0.35 kU/l indicates possible clinical allergy (Nevro-NEL, 2011). However, it still is just an indication, not the final diagnosis. Therefore for the discussion, we will focus on the correlation to IgE levels of more than 0.35 kU/l. The only consistent correlation of IgE levels to *Bifidobacterium* species was detected for *B. adolescentis* levels at 4 months of age and bordering periods over time. The relative amount of *B. adolescentis* was higher in 4-month-old sensitized infants. From 10 days to 4 months of age, *B. adolescentis* increased in infants who had elevated IgE levels by the age of two years, while levels decreased in the non-sensitized group. Conversely, from 4 months to 1 year of age, its levels increased in healthy, but decreased in allergic, children. This clear difference in the development of *B. adolescentis* over time may indicate that not the bacterial load itself, but changes of it may play role in, or be an indication of, allergy development.

The symptoms of eczema include rash, itching and dryness which may also indicate allergy and can be misinterpreted by parents. Overall, there were five infants, each of whom had been diagnosed with allergy (IgE > 0.7 kU/l) and had eczema symptoms; in two of them eczema was confirmed by UKWP criteria. We find it very interesting that three of these 4-month-old infants, including two with medically confirmed eczema, had much higher abundance of *B. adolescentis* compared to other children. *B. adolescentis* has been isolated more frequently from stool samples of allergic children (Ouwehand et al, 2001; Stsepetova et al., 2007). This species has been reported to trigger pro-inflammatory TNF-α, IL-6 and IL-12 cytokines, while it lacks the ability to induce anti-inflammatory IL-10 (He et al., 2002). In addition, *B. adolescentis* was shown to induce IL-5 which is also associated with atopic responses (Tannock, 2010). Taken together, this suggests that highly elevated levels of *B. adolescentis* around 4 months of age may be correlated to the development of a range of allergic disorders.

UKWP-diagnosed eczema correlated to the abundance of *B. longum* load at 10 days and 1 year of age and to the change in its abundance in newborns. In a recent study of Hong et al. (2010), higher abundances of *B. longum* were detected in 1-year-old non-eczema infants, but our data suggest the opposite. However, in Hong's study, the correlation was detected among

caesarian-delivered children, whereas the *B. longum* abundance in stool samples from vaginally-delivered infants showed no correlation to eczema. Interestingly, for the IMPACT dataset, the abundance of all *Bifidobacterium* spp. was also found significantly higher in 1-year-old infants with medically diagnosed eczema.

4.6 Transmittance from mother to child

We did not find correlation between the presence of a given bifidobacteria species in mothers with that in their newborns. For this test, we have binarized the data based on the percentage of the species in a given sample. However, the binarization approach is not the best way to test for the independence between two variables because of uncertainity regarding which threshold should be used. The only statistically justified threshold we could use was the estimated error of the MCR prediction. Unlike the detection threshold, the error of prediction does not imply that if the amount of the species is lower than this value, then the species will not be detected, it only implies is that if the real relative amount of the species is 10 % then the predicted amount may (or may not) deviate, in average, by 5 %. The best approach would be to look for the independence in actual, not binarized, data. A graphical representation of the percentage in mothers vs. the percentage in children clearly suggests that the presence of all five dominant bifidobacterial species is independent of that in the mothers. However, that still does not exclude the transmittance of bifidobacteria. We have to keep in mind that we have been analyzing only the stool samples. At birth, the child is exposed to bacteria which are present in the birth canal of the mother. Therefore vaginal swabs, but not stool samples, are more suitable for the analysis. Moreover, the child receives bifidobacteria during breast-feeding as well – *B. longum*, *B. animalis*, *B. bifidum*, *B. breve*, *B. adolescentis* and *B. catenulanum* have been isolated from the breast milk (Grönlund et al., 2007). Additionally, bifidobacterial DNA have been detected in 8 out of 9 caesarian-delivered placenta samples raising a question about the sterility of the infant gut at birth (Satokari et al., 2009). On the other hand, dendritic cells supposedly transport bacteria from the gut to mother's mammary glands (Perez et al., 2007) and the same mechanism might be used for the transmittance of bacteria (or alternatively bacterial DNA) to placenta. Therefore it may be still possible to show the transmittance from mother to child by assessing maternal gut microbiota.

When we compared *B. longum longum* sequences in mother/child pairs, only four out of nine pairs had identical sequences. However, taking into account that apart from unclassified

sequences, only *B. longum longum* was detected in both mothers' and newborns' stool samples, we believe that infants receive this species from their mothers. *B. longum infantis* was mostly detected at 4 months of age – there was only one sample where it was detected earlier. This indicates that *B. longum infantis* is most likely transmitted child-to-child and competes out *B. longum longum* as it is more suited for a milk diet (LoCascio et al., 2010). By the age of 1 year, however, when the change to a more adult-like diet occurs, *B. longum longum* regains its position and persists through life. There were only two infants where *B. longum infantis* persisted up to 1 year of age. Unfortunately for both infants, the questionnaires were not submitted and the information on their diet or lifestyles is absent.

4.7 Suggestions of further work

MCR is a promising approach for the analysis of mixed DNA samples; however, it would be beneficial to improve its predictability. We tested the possibility to use PLS regression for the correction of MCR predicted concentration profiles. This allowed diminishing the error of the concentration prediction in mixed samples and reducing the level of over- and underestimating some bacterial species. However, the PLS regression model has to be constructed for each specific dataset, so the researcher must be sure of the bacterial diversity in the dataset in question. This method might be a useful approach for concentration profile corrections when performing various co-culturing time series or, alternatively, when the bacterial diversity of the dataset is known in advance.

The basecaller script designed in this project can also be improved. In case of mixed spectra with two peaks at one position, the basecaller calls for the IUPAC nucleotide ambiguity code, however, in case three peaks are mixed, the script assigns it not one, but instead two mixed nucleotides, which then requires manual correction. Also, when looking for the mixed peaks the basecaller script does not take into account the height ratio between these peaks. That means that if one of the peaks is slightly higher than the noise (e.g. 0.012 rfu), whereas the other is much higher (e.g. 0.08 rfu), the script would still assign a mixed nucleotide at this position. So implementation of both height ratio and three-mixed-peak nucleotide recognition would be a great improvement to the script.

To further confirm the correct identification and quantification of dominant bifidobacterial species in stool samples, it is beneficial to perform the analysis with another gene. More

preferably, the full MLST should be performed on the data, not only to ensure the right identification of *Bifidobacterium* species, but also to examine the transmittance of these species by comparing the mother/child pairs.

Moreover, it would be very interesting to isolate *B. longum longum* and *B. longum infantis* from IMPACT stool samples and perform a competition test with lacto-N-biose or, preferably purified HMOs, as a single carbon source to find out if the competition is driven solely by the diet. Our preliminary results on *B. longum* strains obtained from DSMZ collection suggest that *B. longum longum* DSM20219 outcompetes *B. longum infantis* DSM20088 at high cell densities, whereas it loses the competition at low cell concentration, which may point to their r/K selection. However, the experiment with each cell density was performed only once, so it should be repeated.

IMPACT dataset has information on 720 individuals; we have analyzed only a small subset of it. We failed to determine correlations with regard to many environmental factors of interest, like pets at home or smoking parents, because there were not enough samples to separate groups into two. Therefore, it would be interesting to include more samples for the analysis to increase the strength of separation. Also, the questionnaires contained a lot of information on living conditions or health and diet both during pregnancy and the first year of life which we did not investigate due to time constraints. Therefore, it would also be interesting to increase the number of factors.

In many studies, *B. adolescentis* has been detected more often in allergic infants. The consistent correlation of its abundance and the change in it around 4 months of age makes it very interesting to follow what happens during this period of development. Moreover, the extremely high abundance of *B. adolescentis* in three infants who had both high levels of IgE and eczema, raises even more interest in this bacterium. It would be very interesting to select more children with both high IgE and diagnosed and reported eczema to find out whether this extremely high abundance is a trend or a coincidence. If it is a trend, then it would be interesting to isolate *B. adolescentis* from the stool samples of these infants and perform the experiments with dendritic cells to see what cytokine profile is induced.

Surprisingly, some *Bifidobacterium* species were consistently correlated to reports of having a common cold and vaginal fungal infection during pregnancy. To confirm the correlation findings, first we need to increase the number of samples. Also, we have not specified the

stage of the pregnancy when the infection occurred. Instead, we were solely concentrating on the fact that these diseases did or did not reportedly occur during pregnancy. However, the information about which trimester the women suffered from the disease is also available. Therefore it would be interesting to determine if the correlation of bifidobacterial species in infants with the occurrence or absence of these infections is independent of the pregnancy stage or not. If the correlation between the *Bifidobacterium* species composition and development is observed only if the cold or fungal infection happened at a single period, rather than during the entire pregnancy, then what happens during this particular period of time should be investigated. It would be of great interest to examine breast-milk samples from the mothers. It might be that both contracting a common cold and vaginal fungal infection during pregnancy may have an impact on the development of milk secretory cells, and the difference we see in infant's stool samples is actually caused by the difference in the bacterial composition of the mother's milk. Moreover, there was a positive cross-correlation between breast-feeding throughout the first year of life and allergy development. Therefore, it would also be interesting to see what changes in breast milk composition occur 1 year after the lactation starts.

5 Conclusion

Five dominant *Bifidobacterium* species and their relative abundance were identified in stool samples of 83 infants of age 3 days, 10 days, 4 months, 1 year and 2 years and of their mothers during pregnancy. By the age of 10 days, *B. adolescentis* comprised the same relative amount of the total bacterial load in infant gut as in adults. Relative to *Bifidobacterium* spp., *B. breve* was the most abundant species in 4-month-old infants, whereas by the age of two years, stool samples were richest in *B. longum*, *B. adolescentis* and *B. bifidum* and the development of bifidobacterial microbiota was directed towards the adult profile. The *B. longum* group mostly consisted of *B. longum longum* irrespective of age, and *B. longum infantis* was mostly identified in 4-month-old individuals indicating the infant-to-infant nature of its transmittance. Possible interactions between *B. longum* and outer bacterial groups were revealed by LSA. Mother's health status during pregnancy was found to be correlated to the composition and development of bifidobacterial microbiota of infants. The amount of *B. breve* in newborns, 4-month-old and 2-year-old infants correlated with the reported incidence of a common cold during pregnancy. The change in the relative amount of all *Bifidobacterium* spp. and the *B. longum* group in stool samples of infants during almost all age periods studied correlated to the reported occurrence of vaginal fungal infection during pregnancy. Elevated amounts of *B. adolescentis* in 4-month-old infants correlated to high IgE levels and eczema symptoms; levels of the *B. longum* group in infants correlated with the occurrence of eczema.

6 References

Agresti, A. (1992). A Survey of Exact Inference for Contingency Tables. *Statistical Science, 7* (1), 131-153.

Ammehjelpen. (2011, 02 01). *Ammeslutt/avvenning*. Retrieved April 2011, from http://ammehjelpen.no/artikkel/55/Ammesluttavvenning

Barrangou R. (2003). Functional and comparative analysis of an operon involved in FOS utilization by *L.acidophilis*. *PNAS, 100,* 8957-8962.

Blaser M., Falkow S. (2009). What are the consequences of the disappearing human microbiota? *Nature Reviews. Microbiology, 7,* 887-893.

Brisse S. et al. (2009, 06 09). *Bifidobacterium MLST Database*. Retrieved from Institut Pasteur: http://www.pasteur.fr/recherche/genopole/PF8/mlst/Bifidobacterium.html

Brugere J. et al. (2009). Tools for stools: the challenge of assessing human intestinal microbiota using molecular diagnostics. *Expert Rev. Mol. Diagn., 9* (4), 353-365.

Candela M. et al. (2004). Evaluation of the rrn operon copy number in *Bifidobacterium* using real-time PCR. *Letters in Applied Microbiology, 38,* 229-232.

Carr P. et al. (1998). Evaluation and management of vaginitis. *J Gen Intern Med, 13,* 335-346.

Cline M. et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols, 2,* 2366-2382.

Cohen, S. (1996). Psychological stress, immunity, and upper respiratory infections. *Current Directions in Psychological Science, 5* (3), 86-90.

de Muinck E. et al. (2011). Diversity, transmission and persistence of *Escherichia coli* in a cohort of mothers and their infants. *Environmental Microbiology Reports*.

de Vries W., Stouthamer A. (1967). Pathway of glucose fermentation in relation to the taxonomy of bifidobacteria. *J. Bacteriol., 93,* 574-576.

Deguchi Y. et al. (1985). Comparative studies on synthesis of water-soluble vitamins among human species of bifidobacteria. *Agric. Biol. Chem., 49* (1), 13-19.

Deinzer R., Schuller N. (1998). Dynamics of stress-related decrease of salivaryImmunoglobulin A (sIgA): relationship to symptoms of the common cold and studying behavior. *Behavioral Medicine, 23* (4), 161-169.

Deletoile A. et al. (2010). Species delineation and clonal diversity in four *Bifidobacterium* species as revealed by multilocus sequencing. *Research in Microbiology*, doi:10.1016/j.reismic.2009.12.006.

Department of Health. (2007). *Weaning. Starting solid food*. Retrieved April 2011, from http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_084164.pdf

Dereeper A., Guignon V. et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res., 36*.

Dethlefsen L. et al. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature, 449*, 811-817.

Eckburg P. et al. (2005). Diversity of the human intestinal microbial flora. *Science, 308*, 1635-1638.

Edgington, E. S. (1995). *Randomisation tests* (3rd ed.). NY: Marcel Dekker.

Falony G. et al. (2006). Cross-feeding between *Bifidobacterium longum* BB536 and acetate-converting, butyrate-producing colon bacteria during growth on oligofructose. *Appl. Environ. Microbiol., 72*, 7835-7841.

Ferraris L. et al. (2010). New selective medium for selection of bifidobacteria from human feces. *Anaerobe* (16), 469-471.

Fliss I. et al. (2010). Antimicrobial activity of the genus *Bifidobacterium*. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 125 - 144). Caister Academic Press.

Forno E. et al. (2008). Diversity of the gut microbiota and eczema in early life. *Clinical and Molecular Allergy, 6* (11).

Garrido M. et al. (2008). Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reaction processes. *Anal Bioanal Chem, 390,* 2059-2066.

Glazowska S., Czebreszuk U. (2010). *Multilocus sequence type characterization of Bifidobacteria in infants and their mothers.* Bachelor's degree project, Hedmark University College.

Grönlund M. et al. (2007). Maternal breast-milk and intestinal bifidobacteria guide the compositional development of the *Bifidobacterium* microbiota in infants at risk of allergic disease. *Clinical and Experimental Allergy, 37,* 1764-1772.

Guiemonde M. et al. (2007). Competitive exclusion of enteropathogens from human intestinal mucus by *Bifidobacterium* strains with acquired resistance to bile - A preliminary study. *Int. J. Food Microbiol., 113,* 228-232.

Hattori M., Taylor T. (2009). The human intestinal microbiome: a new frontier of human biology. *DNA Research* (16), 1-12.

He F. et al. (2002). Intestinal *Bifidobacterium* species induce varying cytokine production. *J Clin Immunol, 109,* 1035-1036.

Hong P. et al. (2010). Comparative analysis of fecal microbiota in infants with and without eczema. *PLoS ONE, 5* (4).

Huson D. et al. (2007). Dendroscope- An interactive viewer for large phylogenetic trees. *BMC Bioinformatics, 8* (460).

Haarman M., Knol J. (2005). Quantitative real-time PCR assays to identify and quantify fecal *Bifidobacterium* species in infants receiving a prebiotic infant formula. *Appl. Env. Microbiol., 71* (5), 2318-2324.

Janssen EMEA. (2011, February 8). *What are the symptoms of fungal infections*. Retrieved April 2011, from Fungal infections: http://www.janssen-cilag.com/bgdisplay.jhtml?itemname=fungal_symptoms#3

Jaumot J. et al. (2006). Analysis of DNA microarrays by multivariate curve resolution. *Anal. Biochem., 358,* 76-89.

Jolley K. (2011, May 12). *PubMLST. MLST databases and software.* Retrieved May 12, 2011, from http://pubmlst.org/

Kibbe, W. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res., 35(webserver issue): May 25*.

Kimura K. et al. (1997). Analysis of fecal populations of bifidobacteria and lactobacilli and investigation of the immunological responses of their human hosts to the predominant strains. *Appl. Envir. Microbiol, 63* (9), 3394-3398.

Kirjavainen P. et al. (2002). Aberrant composition of gut microbiota of allergic infants: a target of bifidobacterial therapy at weaning? *Gut, 51,* 51-55.

Kovatcheva-Datchary P. et al. (2009). Tools for the tract: understanding the functionality of the gastrointestinal tract. *Therapeutic advances in gastroenterology, 2,* 9-22.

Lamendella R. et al. (2008). Bifidobacteria in feces and environmental waters. *Appl. and Environ. Microbiol., 74,* 575-584.

Langsrud, Ø. (2004, 01 09). *Fisher's Exact test*. Retrieved from http://www.langsrud.com/fisher.htm

Ley R. et al. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell, 124,* 837 - 848.

LoCascio R. et al. (2010). Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. *infantis* as revealed by comparative genomic hybridization. *Appl. and Env. Microbiol., 76* (22), 7373-7381.

Maiden M. et al. (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *PNAS Microbiology, 95,* 3140-3145.

Makras L., de Vuyst L. (2006). The in vitro inhibition of Gram-negative pathogenic bacteria by bifidobacteria is cause by the production of organic acids. *Int. Dairy J., 16,* 1049-1057.

Marques T. et al. (2010). Programming infant gut microbiota: influence of dietary and environmental factors. *Current opinion in biotechnology, 21,* 149-156.

Marteau P. et al. (2002). *Bifidobacterium animalis* strain DN-173 010 shortens the colonic transit time in healthy women: a double-blind, randomized, controlled study. *Aliment Pharmacol Ther, 16*, 587-593.

Menard O. et al. (2008). Gnotobiotic mouse response induced by *Bifidobacterium* sp. strains isolated from infants. *Appl. Environ. Microbiol., 74* (3), 660-666.

Miyauchi E. et al. (2010). Interaction of Bifidobacteria with the intestinal mucosa. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 145-162). Caister Academic Press.

Mohan R. et al. (2008). Effects of *Bifidobacterium lactis* Bd12 supplementation on body weight, fecal pH, acetate, lactate, calprotectin, and IgA in preterm infants. *Pediatr. Res., 64*, 418-422.

Neish, A. (2009). Microbes in gastrointestinal health and disease. *Gastroenterology, 136* (1), 65-80.

Nevro-NEL. (2011, April 16). *Spesifikk IgE (Tidligere RAST)*. Retrieved April 26, 2011, from NEL - Nevrologiske prosedyrer: http://nevro.legehandboka.no/prover-og-svar/immunologi/rast-3050.html

Nishimoto M. et al. (2007). Practical praparation of lacto-n-biose I, a candidate for the bifidus factor in human milk. *Biosci. Biotechnol. Biochem., 71* (8), 2101-2104.

Ouwehand A. et al. (2001). Differences in *Bifidobacterium* flora composition in allergic and healthy infants. *J. Allergy Clin. Immunol., 108*, 144-145.

Palmer C. et al. (2007). Development of the human infant intestinal microbiota. *PLoS Biol, 5* (7), 1556-1573.

Penders J. et al. (2007). Gut microbiota composition and development of atopic manifestation in infancy: the KOALA birth cohort study. *Gut, 56*, 661-667.

Perez P. et al. (2007). Bacterial imprinting of the neonatal immune system: lessons from maternal cells. *Pediatrics, 119*, e724-e732.

Pompei A. et al. (2007). Folate production by bifidobacteria as a potential probiotic property. *Appl. Environ. Microbiol., 73*, 179-185.

Romani, L. (2004). Immunity to fungal infections. *Nature reviews. Immunology, 4*, 1-13.

Rossi M., Amaretti A. (2010). Probiotic properties of Bifidobacteria. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 97-123). Caister Academic Press.

Round J., Mazmanian S. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Immunology, 9*, 313-323.

Roy, D. (2001). Media for the isolation and enumeration of bifidobacteria in dairy products. *International Journal of Food Microbiology, 69*, 167-182.

Ruan Q. et al. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics, 22* (20), 2532-2538.

Sanchez B. et al. (2010). Acid and bile resistance and stress response in Bifidobacteria. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 71-94). Caister Academic Press.

Satokari R. et al. (2009). *Bifidobacterium* and *Lactobacillus* DNA in the human placenta. *Letters in Applied Microbiology, 48*, 8-12.

Sekelja M. et al. (2010). Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *The ISME Journal*, 1-13.

Sela D. et al. (2010). Metabolism of Bifidobacteria. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 45-70). Caister Academic Press.

Shanahan, F. (2009). Therapeutic implications of manipulating and mining the microbiota. *The Journal of physiology, 587.17*, 4175-4179.

Smith, L. (2002). *A tutorial on Principal Component Analysis.* Retrieved January 2010, from http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Statistics solutions. (2006). *Conduct and interpret the Chi-square test of independence*. Retrieved April 2011, from http://www.statisticssolutions.com/products-services/login/standard-membership/conduct-interpret-analyses/chi-square

Stecher B. et al. (2010). Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathogens, 6* (1).

Stsepetova J. et al. (2007). Molecularly assessed shifts of *Bifidobacterium* spp. and less diverse microbial communities are charactersitic of 5-year-old allergic children. *FEMS Immunol Med Microbiol, 51*, 260-269.

Størro O. et al. (2010). A primary health-care intervention on pre- and postnatal risk factor behaviour to prevent childhood allergy. The Prevention of Allergy among Children in Trondheim (PACT) study. *BMC Public Health, 10:443*.

Størro O. et al. (2011). Temporal differences in early gut microbial colonization are associated with atopic sensitization at 2 years of age. *Clinical and Experimental Allergy*.

Tannock, G. (2010). Analysis of bifidobacterial populations in bowel ecology studies. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and Molecular Aspects* (pp. 1-15). Caister Academic Press.

Tauler R. et al. (2010). *Multivariate Curve Resolution - Theory*. Retrieved April 2011, from Multivariate Curve Resolution Homepage: http://www.ub.edu/mcr/web_mcr/theory.html

Tauler R. et al. (1995). Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometrics, 9*, 31-58.

Tomasi G. et al. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemometrics, 18*, 231-241.

Trosvik P. et al. (2007). Multivariate analysis of complex DNA sequence electropherograms for high-throughput quantitative analysis of mixed microbial populations. *Appl. Env. Microbiol., 73* (15), 4975-4983.

Turroni F. et al. (2009). Bifidobacterial diversity in the human intestine. *Appl. Environ. Microbiol., 75* (6), 1534-1545.

van der Meulen R. et al. (2006). Kinetic analysis of bifidobacterial metabolism reveals a minor role for succinic acid in the regeneration of NAD+ through its growth-associated production. *Appl. Environ. Microbiol., 72*, 5204-5210.

Ventura M. et al. (2006). Analysis of bifidobacterial evolution using a multilocus approach. *Int. J. Sys. Evol. Microbiol., 56*, 2783-2792.

Ventura M. et al. (2010a). Bifidobacterial ecology and comparative genomics: perspectives and prospects. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 31-44). Caister Academic Press.

Ventura M. et al. (2005). Genetic and transcriptional organization of the *clpC* locus in *Bifidobacterium breve* UCC 2003. *Appl. Env. Microbiol., 71* (10), 6282-6291.

Ventura M. et al. (2009). Genome-scale analysis of health-promoting bacteria: probiogenomics. *Nat. Rev. Microbiol., 7*, 61-71.

Ventura M. et al. (2007b). Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie van Leeuwenhoek, 91*, 351-372.

Ventura M. et al. (2007a). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews, 3*, 495-548.

Ventura M. et al. (2010b). Genomics of Bifidobacteria: where we stand now and where we are going. In B. Mayo, & D. van Sinderen (Eds.), *Bifidobacteria. Genomics and molecular aspects* (pp. 17-30). Caister Academic Press.

Wada J. et al. (2008). *Bifidobacterium bifidum* lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. *Appl Environ Microbiol, 74*, 3996-4004.

Wall R. et al. (2009). Role of gut microbiota in early infant development. *Clinical Medicine: Pediatrics, 3*, 45-54.

Wang Q. et al. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol, 73* (16), 5261-7.

Wood, P. (2006). Specific immune recognition: the antibody molecule. In P. Wood, *Understanding Immunology* (pp. 41-58). Science.

Xia L. et al. (2008). *Local Similarity Analysis*. Retrieved 01 2011, from Microbial Community and Metagenomics Analysis Server: http://meta.usc.edu/lsa

Yasui H. et al. (1992). Detection of *Bifidobacterium* strains that induce large qiantities of IgA. *Microbial Ecology in Health and Disease, 5*, 155-162.

Yildirim Z., Johnson M. (1998). Characterization and antimicrobial spectrum of bifidocin B, a bacteriocin produced by *Bifidobacterium bifidum* NCFB 1454. *Journal of Food Protection, 61* (1), 47-51.

Zimonja M. et al. (2008). Multivariate curve resolution of mixed bacterial DNA sequence spectra: identification and quantification of bacteria in undefined mixture samples. *J.Chemometrics, 22*, 309-322.

Zoetendal E. et al. (2001). The host genotype affects the bacterial community. *Microbial Ecology in Health and Disease, 13*, 129-134.

Øyen T. et al. (2006). Intestinal microbiota and its effect on the immune system — A nested case-cohort study on prevention of atopy among smallchildren in Trondheim: The IMPACT study. *Contemporary Clinical Trials* (27), 389-395.