Early fecal microbiota composition in children who later develop celiac disease and associated autoimmunity

Anniina Rintala MSc[1,2]; Iiris Riikonen BM[1]; Anne Toivonen MD, PhD[3,4]; Sami Pietilä MSc[5]; Eveliina Munukka PhD[1]; Juha-Pekka Pursiheimo PhD[6]; Laura L Elo PhD[5]; Pekka Arikoski MD, PhD[7]; Kristiina Luopajärvi MD, PhD[8]; Ursula Schwab PhD[9]; Matti Uusitupa MD, PhD[9]; Seppo Heinonen MD, PhD[10]; Erkki Savilahti MD, PhD[8]; Erkki Eerola MD, PhD[2,1]; Jorma Ilonen MD, PhD[11,2]

[1]Department of Medical Microbiology and Immunology, University of Turku, Turku, Finland

[2]Department of Clinical Microbiology and Immunology, Tyks Microbiology and Genetics, Turku University Hospital, Turku, Finland

[3]Department of Bacteriology and Immunology, University of Helsinki and Laboratory Services (HUSLAB), Division of Clinical Microbiology, Helsinki University Hospital, Helsinki, Finland

[4]Department of Clinical Microbiology, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland

[5]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland

[6]Turku Clinical Sequencing Laboratory, University of Turku, Turku, Finland

[7]Department of Pediatrics, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland

[8]Children's Hospital, Department of Pediatrics, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[9]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

[10]Department of Obstetrics and Gynecology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[11]Immunogenetics Laboratory, Institute of Biomedicine, University of Turku

***Corresponding author:**

MSc Anniina Rintala, University of Turku, Department of Medical Microbiology and Immunology, Kiinamyllynkatu 13, 20520 Turku, Finland. Tel. +358 50 463 7602; Fax +358 2 233 0008; email: anniina.rintala@utu.fi

Conflicts of Interest and Source of Funding: The authors declare no conflict of interest. This study was financially supported by Päivikki and Sakari Sohlberg Foundation and Special Research Funds for Kuopio University Hospital.

Word count: 5075, including abstract, references, tables and figure captions

Figures: 3

Tables: 2

Abstract

Objectives: Several studies have reported that the intestinal microbiota composition of celiac disease (CD) patients differs from healthy individuals. The possible role of gut microbiota in the pathogenesis of the disease is, however, not known. Here we aimed to assess the possible differences in early fecal microbiota composition between children that later developed CD and healthy controls matched for age, sex, and HLA risk genotype.

Materials and methods: We used 16S rRNA gene sequencing to examine the fecal microbiota of 27 children with high genetic risk of developing CD. Nine of these children developed the disease by the age of 4 years. Stool samples were collected at the age of 9 and 12 months, before any of the children had developed CD. The fecal microbiota composition of children who later developed the disease was compared with the microbiota of the children who did not have CD or associated autoantibodies at the age of 4 years. Delivery mode, early nutrition, and use of antibiotics were taken into account in the analyses.

Results: No statistically significant differences in the fecal microbiota composition were found between children who later developed CD (n = 9) and the control children without disease or associated autoantibodies (n = 18).

Conclusions: Based on our results, the fecal microbiota composition at the age of 9 and 12 months is not associated with the development of CD. Our results, however, do not exclude the possibility of duodenal microbiota changes or a later microbiota-related trigger for the disease.

Keywords: celiac disease, HLA risk haplotype, fecal microbiota, next-generation sequencing, 16S rRNA gene sequencing

Introduction

Celiac disease (CD) is an immune-mediated disorder of the small intestine where gluten induces a mucosal inflammatory reaction. In Finnish population, the prevalence of CD has doubled during the last 30 years [1]. 90% of patients carry HLA-DQ2.5 molecule encoded by *HLA-DQA1*05* and *–DQB1*02* genes, and most of the remaining patients carry HLA-DQ8 molecule encoded by *HLA-DQA1*03* and *–DQB1*03:02* genes. Around 30-40% of people of Northern European descent are positive for HLA-DQ2.5 and/or DQ8, demonstrating the role of other etiological factors in the disease development in addition to these major genetic components [2].

Several studies have shown a relation between altered intestinal microbiota composition and CD [3-5]. However, the observed differences may arise from disease process itself, or be dependent on genetic differences. For example, the HLA-DQ2.5 haplotype has been reported to influence the gut microbiota composition of infants [6]. The possible changes and differences in the intestinal microbiota before the onset of CD are unclear.

The aim of our study was to screen the fecal microbiota composition of Finnish children genetically at the risk for CD before the onset of the disease. Our objective was to detect the possible differences in the microbiota between children who later developed CD and those who did not. Through this study, we hope to broaden the understanding on the impact of early microbial colonization in the pathogenesis of CD, and to reveal whether early fecal microbiota composition could reflect to the disease risk. To our knowledge, this is the first study so far where the samples have been collected before the onset of CD.

Materials and Methods

Study subjects

Stool samples were obtained from 27 Finnish children (recruited at the Kuopio University Hospital and the Kätilöopisto Maternity Hospital in Helsinki) with a high genetic risk for CD participating in a follow-up study [7]. Newborn children were screened for the presence of HLA-DQB1*02 and HLA-DQA1*05 alleles [8] and followed until 3 or 4 years of age by screening for tissue transglutaminase autoantibodies (tTGA) [9]. The diagnosis of CD was based on typical histological findings, villous atrophy and crypt hyperplasia, in the duodenal biopsy of small intestine [10]. Nine children (all girls) were diagnosed with duodenal biopsy at the median age of 3.5 years (range 2.6 - 4.2 years) after the development of tTGA autoantibodies at the median age of 3 years (18 months to 3 years). They were selected for the study together with 18 control infants, matched by gender and date of birth, remaining negative for tTGA during the follow-up.

Stool samples were collected at the age of 9 and 12 months (9mo and 12mo samples, respectively). Samples were stored in home freezers (-20°C) for a maximum period of two months, after which they were delivered to the study centers in cool boxes filled with ice bags and stored at -80°C until processing. The 9mo sample from one case infant and 12mo stool sample from one control infant had not been delivered. Data on mode of delivery were obtained from the participating families or obstetric records, and data on nutrition and antibiotics were collected using dietary questionnaires and phone calls at the age of 1-2 weeks, 1.5 months, 3.5 months, 6 months and 9 months. After the age of 9 months, the health-related information was collected by general follow-up questionnaires. Written informed consent was obtained from all

families. The study plan was accepted by the ethics committee of the Kuopio University Hospital.

DNA extraction

Bacterial DNA from ~100 mg of the frozen stool specimens was extracted with a semi-automatic GXT Stool Extraction Kit VER 2.0 (Hain Lifescience GmbH, Nehren, Germany) combined with an additional homogenization by bead-beating in 0.1 mm Glass Bead Tubes (MO BIO Laboratories, Inc., Carlsbad, CA, USA) at 1000 rpm for 3 minutes with MO BIO PowerLyzer™ 24 Bench Top Bead-Based Homogenizer (MO BIO Laboratories, Inc., Carlsbad, CA, USA) to enhance the cell lysis. The DNA concentrations of the DNA extracts were measured with Qubit dsDNA HS assay kit and Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA). The DNA extracts were stored at -80 °C.

Fecal microbiota composition analysis

Bacterial profiles of the stool specimens were analyzed with 16S rRNA gene sequencing. The 16S rRNA gene libraries were generated in a single PCR with custom-designed dual-indexed primers. The approach is described in Supplemental Materials and Methods.

Data analysis

The raw sequence quality was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and data analysis was performed with Qiime (v. 1.9) pipeline as described previously [11-13]. Sequence reads were filtered with a quality score acceptance rate of 20 or better. Chimeric sequences were filtered using usearch (v.

6.1), and operational taxonomic units (OTUs) were picked using uclust algorithm with 97% sequence similarity. OTUs representing less than 0.05% of the total sequence count were excluded. To minimize the effect of inter-sample variation in the sequencing efficiency, samples were subsampled (rarefied) by random sampling without replacement to the lowest common sequencing depth (134 414 reads). Annotations for the resulting OTUs were derived from GreenGenes database [14].

All analyses of the 16S rRNA data were made from the randomly subsampled OTU tables. To study the bacterial diversity of the samples, α-diversity metrics were computed and α-rarefaction plots were generated with Qiime. Differences in the Shannon diversity indices were then assessed with JMP Pro 12 (SAS Institute, Inc., Cary, NC, USA), applying non-parametric methods and considering $P < 0.05$ as statistically significant. Outliers were excluded before the analyses. Taxonomic summary produced by Qiime was visualized as bar charts and statistical differences in the taxonomic richness, i.e. in the OTU abundances, were assessed with non-parametric Kruskal-Wallis test. Taxonomic levels phylum and genus were studied, and False Discovery Rate (FDR) adjusted *P*-value < 0.05 was considered as statistically significant. OTUs existing in less than 25% of the samples were excluded before statistical testing. To analyze the differences in the overall bacterial diversity across the samples, weighted UniFrac distance matrices were generated from the randomly subsampled OTU tables and principal coordinate analysis (PCoA) plots were produced. The PCoA plots were visualized with EMPeror. To confirm the visual observations, adonis analyses were performed. Adonis returns an R^2 value showing the amount of variation explained by the grouping variable, and a *P*-value for statistical significance [15].

Differences in the nutritional variables (duration of breastfeeding, introduction of solid food, and introduction of gluten), and in the number of received antibiotic courses were analyzed with JMP Pro 12 (SAS Institute, Inc.), using non-parametric methods and considering $P < 0.05$ as statistically significant. To eliminate the mode of birth as a confounding factor, all statistical analyses were first performed for the whole data set (n = 27), and then repeated for the vaginally born infants (n = 24).

Results

The characteristics of the participating children are presented in Table 1. Of the 27 infants in this study, 24 were born by vaginal delivery while three were born by cesarean section (CS). All infants were breastfed at the hospital, but the continuation of breastfeeding varied from seven days to over 18 months (Table 1). The median duration of the breastfeeding was 11 months, yet breastmilk was often supplemented with formula. The median age for solid food introduction was 4.1 months (range 2.0 - 6.2 months), and for gluten introduction 5.7 months (range 4.4 - 7.7 months). Infants that were earlier introduced to solid foods were also introduced to gluten earlier (Spearman's correlation 0.679, $P < 0.001$). No differences in the average breastfeeding duration or in the age of solid food or gluten introduction were observed between the case and control infants (Mann-Whitney U $P = 0.5395$, $P = 0.4712$ and $P = 0.5539$, respectively; Table 2). Control children seemed to be breastfed at the age of gluten introduction more often than the case infants, but the difference was not statistically significant (Pearson's chi-squared test $P = 0.2113$; Table 2). Furthermore, no differences were observed in the prescription of antibiotics between the case and control infants (Table 2). In total, 80.8 % of the study subjects had received at least one antibiotic course by the age of 4 years. Exclusion of the infants born by CS did not significantly affect the results regarding the nutritional information or the antibiotics (Table 2).

The 16S rRNA gene sequencing of the 52 infant stool specimens resulted in 134k – 330k OTUs per sample (mean 191k, SD 40k), the overall sequence count being 9.9×10^6 OTUs. The average bacterial diversity of the stool samples, represented as median Shannon index values, was 3.32 (range 2.66 - 4.32) for 9mo samples and 4.10 (range 3.27 - 4.54) for 12mo samples, showing significant increase in the bacterial diversity between 9 and 12 months ($FDR = 0.0002$). The

difference remained significant after the exclusion of the CS infants (*P = 0.0014*). When all study subjects were included in the analysis, the Shannon indices did not differ between the case and control infants (*P = 0.1113* for 9mo samples and *P = 0.686* for 12mo samples). The "observed species" metric of Qiime confirmed these findings, representing little difference between the case and control infants (Fig. 1a and 1b) but a clear difference between the 9mo and 12mo samples (Fig. 1c). However, when only the vaginally born infants were studied, the Shannon indices in the 9mo samples tended to be higher in the cases than in the controls (*P = 0.049*). In the 12mo samples, the difference between the case and control infants remained insignificant after exclusion of the infants born by CS (*P = 0.302*).

The average phylum and genus level bacterial composition of the case and control samples, i.e. the average relative bacterial OTU abundances, are represented as bar charts in Fig. 2. Minor differences can be visually seen in the bacterial taxonomy between the case and control infants in both 9mo and 12mo samples (Fig. 2a and 2b). However, Qiime reported no statistically significant differences in the bacterial composition between the case and control infants when the microbial abundances at phylum and genus levels were assessed. Further, no differences were observed in *Bacteroides-Prevotella* or *Bacteroides-Bifidobacterium* ratios between the case and control infants in neither 9mo nor 12mo samples (*P = 0.201* and *P = 0.388* for *Bacteroides-Prevotella* and *P = 0.955* and *P = 0.572* for *Bacteroides-Bifidobacterium*, respectively). In principal coordinate analysis (PCoA) plot, where individual samples with similar microbiota composition cluster together, no differences could be observed between the case and control infants (Fig 3a and b). Further, adonis analysis confirmed that no significant differences between the cases and controls occurred in neither 9mo nor 12mo samples (*P = 0.808* and *P = 0.696*,

respectively). One 9mo control sample had a significantly deviating bacterial profile (Fig. 3a) with more than 50% of the bacteria being *Enterobacteriaceae*, but as this had no effect on the results of the statistical analyses and did not significantly affect the bar chart figures, the sample was not excluded from the final analyses. The results regarding the fecal microbiota composition between the case and control infants remained unaltered when excluding the infants born by cesarean section, with only minor differences in the *P* values (results not shown).

The results concerning the effect of age, delivery mode and nutritional variables on the fecal microbiota composition are presented in Supplemental Results. Briefly, the bacterial composition differed significantly between the 9mo and 12mo samples (Fig. 2b and 3c) and between the infants born vaginally or by CS. Further, the duration of breastfeeding correlated with several bacterial genera. However, the fecal microbiota composition did not significantly differ between the children that had been breastfed during gluten introduction and the ones that had not (*FDR > 0.1* for all bacterial phyla and genera in both 9mo and 12mo samples).

Discussion

Previous studies have reported that the gut microbiota composition of CD patients differs from healthy individuals [3,16]. It is, however, unclear whether the altered gut microbiota has a role in the pathogenesis of CD or merely is a consequence of the disease [17]. The HLA-DQ2.5 haplotype has been reported to influence the intestinal bacterial community of infants [6], but most of these children do not develop CD. In the Finnish population, less than 10% of the DQ2.5 positive subjects develop the disease. We compared the early fecal microbiota composition between children with the risk genotype who later developed CD and genetically susceptible children who did not develop the disease or associated autoantibodies during the follow-up. Based on the 10% lifetime risk and identification of already 6.5% of children with tTGA (cases included) in the whole follow-up group, we don´t expect many new cases among the selected tTGA negative controls. Follow-up studies also suggest that the appearance of new CD associated autoantibodies is leveling off after 3 to 4 years of age [18].

Stool samples were collected at the age of 9 and 12 months and analyzed with 16S rRNA gene sequencing enabling the detection and relative quantification of the bacterial taxa present in the samples. Previous studies have reported differences in the levels of *Bacteroides*, *Clostridium* and *Staphylococcus* genera, alongside with the *Bacteroides-Prevotella* ratio between the CD patients and healthy controls [16]. In addition, one certain *Bacteroides* species, *Bacteroides dorei,* has been associated with active CD [19]. In this present study, the abundance of *Bacteroides dorei* was not assessed, as the species level identification by 16S rRNA gene sequencing is uncertain [11,20]. At the genus level, however, neither *Bacteroides* abundance nor the *Bacteroides-Prevotella* ratio differed between the case and control infants. Further, the abundance of

Clostridium genera did not differ between the case and control infants, and *Staphylococcus* genus was completely undetectable in this sample cohort. Altogether, in this study, no statistically significant differences in the fecal microbial diversity or composition were found between the children who later developed CD and the ones who did not. Thereby, our results suggest that early fecal microbiota composition would not be associated with the pathogenesis of CD, while the previous findings concerning the differences in the gut microbiota between CD patients and healthy individuals might have risen for example from dietary changes after the disease onset or from the disrupted gut homeostasis due to altered mucosal immune responses [21,22]. To some extent, the findings regarding the gut microbiota composition in CD may also be methodology dependent. Further, as the 16S rRNA sequencing only reveals the microbial composition of the stool samples, functional profiles of the microbiota remain uncovered. In order to study the functionality of the gut microbes, metaproteomic or metabolomic analyses should be performed [23,24]. It has been previously shown that the metabolic activity of the gut microbiota in CD children differs from healthy individuals both before and after the implementation of gluten-free diet [25,26].

Even though the results of this study suggest that the early fecal microbiota composition may not predict the development of CD, this does not exclude the possibility of a microbiota-related trigger for the disease. As e.g. dietary changes, antibiotics and infections can disrupt the intestinal homeostasis [27], it remains possible that some external trigger could later disastrously disturb the gut microbiota balance, leading to the onset of the disease [28]. In fact, one hypothesis for the CD onset is that intestinal infections could interfere the intestinal homeostasis and lead to increased intestinal permeability, i.e. leaky gut (for review, see [29]), which could

then allow the absorption of undigested gliadin molecules that initiate the immune processes leading to the disease [30]. However, few studies have investigated the role of specific infectious agents in the development of the disease, yet in some studies an increased prevalence of CD has been associated with repeated early infections [31,32]. In this study, use of antibiotics was taken into account in the analyses, but the possible impact of early viral infections was not assessed.

In addition to sudden external triggers, gradual unfavorable progression of the gut microbiota may also lead to leaky gut and the activation of the inflammatory pathways [29]. For example, epidemiological studies have reported that children born by CS have an increased risk for CD compared to the vaginally delivered infants [33-35]. In this study cohort, the mode of delivery had a significant effect on the fecal microbiota composition, but due to the limited number of CS children, no conclusions regarding the role of delivery mode on CD risk could be drawn. However, based on the results of this study, the individuals who develop CD do not already in the early infancy have a distinct fecal microbiota composition compared to other infants with risk-HLA-haplotype, suggesting that the onset of CD is more likely a consequence of a strong external trigger rather than gradual development due to a peculiarly vulnerable gut microbiota. However, feces may not be the most optimal sample material for CD studies, as CD primarily affects the small intestine and fecal microbiota inadequately reflects the duodenal microbiota [4,36]. Thus, this study does not rule out the possibility of small intestinal microbiota changes prior to the onset of the disease. On the other hand, the duodenal mucosal homeostasis was not assessed, as studies on the mucosal host-microbial crosstalk and the possibility of unfavorable immune functions would have required the collection of duodenal biopsies. Previous studies

regarding the role of duodenal microbiota in CD-linked mucosal immune responses remain rather inconclusive [22,37,38], making this an extremely interesting area for future research.

The possible protective role of breastfeeding against CD onset has been analyzed in several studies, and the gradual gluten introduction during ongoing breastfeeding has been suggested to protect against the disease [39,40]. Some recent studies, however, have not been able to confirm these results [39-42]. In this study, the average duration of breastfeeding did not differ between the case and control children, but the controls tended to be more likely breastfed during the gluten introduction. The duration of breastfeeding correlated with several bacterial genera, of which especially *Lactobacillus* and *Bifidobacterium* have been previously linked to positive health outcomes (for review, see [43]). However, no difference in the abundance of *Lactobacillus* or *Bifidobacterium* could be observed between the infants who later developed CD and the infants who did not. Further, no differences in the fecal microbiota composition were seen between the infants that were breastfed during gluten introduction and the ones who did not, suggesting that even though breastfeeding might induce changes in the gut microbiota, the possible protective role of breastfeeding against CD might not be related to gut microbiota. However, due to the limited number of infants not breastfed during the gluten introduction in this study, these results are merely indicative.

Conclusions

Our results indicate that the fecal microbiota composition at the age of 9 and 12 months is not associated with the development of CD. Our results do not exclude the possibility of duodenal microbiota differences or a later microbiota-related trigger for the disease, but suggest that the

infants developing CD do not originally have distinct fecal microbiota composition compared to individuals that possess the risk-HLA-haplotype but do not develop the disease.

Acknowledgements

Anne Björk and Sirpa Nolvi are acknowledged for organizing the sample collection and for collecting the nutritional and health data. Heidi Isokääntä and Minna Lamppu are warmly thanked for excellent technical assistance in the sample processing.

Declaration of interest

The authors declare no conflicts of interest.

References

[1] Lohi S, Mustalahti K, Kaukinen K, et al. Increasing prevalence of coeliac disease over time. Aliment Pharmacol Ther. 2007;26(9):1217-1225.

[2] Dieli-Crimi R, Cénit MC, Núñez C. The genetics of celiac disease: A comprehensive review of clinical implications. J Autoimmun. 2015;64:26-41.

[3] Nadal I, Donant E, Ribes-Koninckx C, et al. Imbalance in the composition of the duodenal microbiota of children with coeliac disease. J Med Microbiol. 2007;56(12):1669-1674.

[4] Collado MC, Donat E, Ribes-Koninckx C, et al. Specific duodenal and faecal bacterial groups associated with paediatric coeliac disease. J Clin Pathol. 2009;62(3):264-269.

[5] Collado MC, Donat E, Ribes-Koninckx C, et al. Imbalances in faecal and duodenal Bifidobacterium species composition in active and non-active coeliac disease. BMC microbiology. 2008;8(1):232.

[6] Olivares M, Neef A, Castillejo G, et al. The HLA-DQ2 genotype selects for early intestinal microbiota composition in infants at high risk of developing coeliac disease. Gut. 2015;64(3):406-417.

[7] Lammi A, Arikoski P, Hakulinen A, et al. Development of gliadin-specific immune responses in children with HLA-associated genetic risk for celiac disease. Scand J Gastroenterol. 2016;51(2):168-177.

[8] Lundin KE, Scott H, Hansen T, et al. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J Exp Med. 1993;178(1):187-196.

[9] Dieterich W, Ehnis T, Bauer M, et al. Identification of tissue transglutaminase as the autoantigen of celiac disease. Nat Med. 1997;3(7):797-801.

[10] Marsh MN. Gluten, major histocompatibility complex, and the small intestine: a molecular and immunobiologic approach to the spectrum of gluten sensitivity (‘celiac sprue’). Gastroenterology. 1992;102(1):330-354.

[11] Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nature methods. 2010;7(5):335-336.

[12] Kuczynski J, Stombaugh J, Walters WA, et al. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Current protocols in microbiology. 2012:1E. 5.1-1E. 5.20.

[13] Bokulich NA, Subramanian S, Faith JJ, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature methods. 2013;10(1):57-59.

[14] DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-5072.

[15] Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, et al. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol. 2013;531:371-444.

[16] Collado MC, Calabuig M, Sanz Y. Differences between the fecal microbiota of coeliac infants and healthy controls. Curr Issues Intestinal Microbiol. 2007;8(1):9.

[17] Cenit MC, Codoner-Franch P, Sanz Y. Gut Microbiota and Risk of Developing Celiac Disease. J Clin Gastroenterol. 2016;50 Suppl 2, Proceedings from the 8th Probiotics, Prebiotics & New Foods for Microbiota and Human Health meeting held in Rome, Italy on September 13-15, 2015:S148-S152.

[18] Liu E, Lee H, Aronsson CA, et al. Risk of pediatric celiac disease according to HLA haplotype and country. N Engl J Med. 2014;371(1):42-49.

[19] Sanchez E, Donat E, Ribes-Koninckx C, et al. Intestinal Bacteroides species associated with coeliac disease. J Clin Pathol. 2010;63(12):1105-1111.

[20] Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4516-4522.

[21] De Palma G, Nadal I, Medina M, et al. Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children. BMC microbiology. 2010;10(1):63.

[22] Cheng J, Kalliomäki M, Heilig HG, et al. Duodenal microbiota composition and mucosal homeostasis in pediatric celiac disease. BMC gastroenterology. 2013;13(1):113.

[23] Matysik S, Le Roy CI, Liebisch G, et al. Metabolomics of fecal samples: a practical consideration. Trends Food Sci Technol. 2016;57:244-255.

[24] Wilmes P, Heintz-Buschart A, Bond PL. A decade of metaproteomics: where we stand and what the future holds. Proteomics. 2015;15(20):3409-3417.

[25] Tjellström B, Stenhammar L, Högberg L, et al. Gut microflora associated characteristics in children with celiac disease. Am J Gastroenterol. 2005;100(12):2784-2788.

[26] Nistal E, Caminero A, Vivas S, et al. Differences in faecal bacteria populations and faecal bacteria metabolism in healthy adults and celiac disease patients. Biochimie. 2012;94(8):1724-1729.

[27] Lozupone CA, Stombaugh JI, Gordon JI, et al. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489(7415):220-230.

[28] Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nature Reviews Immunology. 2016;16(6):341-352.

[29] Marasco G, Di Biase AR, Schiumerini R, et al. Gut microbiota and celiac disease. Dig Dis Sci. 2016;61(6):1461-1472.

[30] Green PH, Cellier C. Celiac disease. N Engl J Med. 2007;357(17):1731-1743.

[31] Myléus A, Hernell O, Gothefors L, et al. Early infections are associated with increased risk for celiac disease: an incident case-referent study. BMC pediatrics. 2012;12(1):194.

[32] Stene LC, Honeyman MC, Hoffenberg EJ, et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. Am J Gastroenterol. 2006;101(10):2333-2340.

[33] Decker E, Engelmann G, Findeisen A, et al. Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children. Pediatrics. 2010;125(6):e1433-40.

[34] Cho CE, Norman M. Cesarean section and development of the immune system in the offspring. Obstet Gynecol. 2013;208(4):249-254.

[35] Mårild K, Stephansson O, Montgomery S, et al. Pregnancy outcome and risk of celiac disease in offspring: a nationwide case-control study. Gastroenterology. 2012;142(1):39-45. e3.

[36] Stearns JC, Lynch MD, Senadheera DB, et al. Bacterial biogeography of the human digestive tract. Scientific reports. 2011;1:170.

[37] Nistal E, Caminero A, Herrán AR, et al. Differences of small intestinal bacteria populations in adults and children with/without celiac disease: effect of age, gluten diet, and disease. Inflamm Bowel Dis. 2011;18(4):649-656.

[38] McLean MH, Dieguez D,Jr, Miller LM, et al. Does the microbiota play a role in the pathogenesis of autoimmune diseases?. Gut. 2015;64(2):332-341.

[39] Ivarsson A, Myleus A, Norstrom F, et al. Prevalence of childhood celiac disease and changes in infant feeding. Pediatrics. 2013;131(3):e687-94.

[40] Ivarsson A, Hernell O, Stenlund H, et al. Breast-feeding protects against celiac disease. Am J Clin Nutr. 2002;75(5):914-921.

[41] Vriezinga SL, Auricchio R, Bravi E, et al. Randomized feeding intervention in infants at high risk for celiac disease. N Engl J Med. 2014;2014(371):1304-1315.

[42] Lionetti E, Castellaneta S, Francavilla R, et al. Introduction of gluten, HLA status, and the risk of celiac disease in children. N Engl J Med. 2014;371(14):1295-1303.

[43] de Sousa Moraes LF, Grzeskowiak LM, de Sales Teixeira TF, et al. Intestinal microbiota and probiotics in celiac disease. Clin Microbiol Rev. 2014;27(3):482-489.

Appendices

Supplemental Materials and Methods

Supplemental Results

Tables

Table 1: Background information of the study subjects

ID	Case/control	Mode of delivery	Solid food introduction (months)	Gluten introduction (months)	Duration of breastfeeding (months)
1	case	vaginal	6.24	6.24	> 18.0
2	control	vaginal	4.27	7.66	6.0
3	control	vaginal	3.02	4.50	0.23
4	case	vaginal	3.98	4.40	1.0
5	control	vaginal	4.63	5.55	N/A
6	control	vaginal	3.88	5.75	11.0
7	case	vaginal	4.01	5.03	2.5
8	control	vaginal	4.60	5.13	13.0
9	control	vaginal	4.04	5.52	6.0
10	case	vaginal	4.04	5.62	N/A
11	control	vaginal	4.17	5.85	13.0
12	control	vaginal	3.84	5.36	12.0
13	case	vaginal	2.00	4.50	0.33
14	control	vaginal	4.07	5.26	6.0
15	control	vaginal	4.50	5.26	10.0
16	case	vaginal	5.55	6.31	12.0
17	control	cesarean section	3.94	6.05	7.0
18	control	vaginal	5.16	5.85	12.0
19	case	vaginal	5.36	6.67	14.0
20	control	vaginal	5.59	5.95	18.0
21	control	vaginal	5.55	6.44	7.5
22	case	vaginal	5.42	6.08	16.0
23	control	cesarean section	3.81	5.22	7.5
24	control	vaginal	4.14	5.65	16.0
25	case	cesarean section	4.07	6.28	> 18.0
26	control	vaginal	4.01	5.65	11.0
27	control	vaginal	3.48	5.00	5.0

Table 2: Nutritional information and antibiotic courses of the study subjects

	All infants in this study (n = 27)			Vaginally born infants (n = 24)		
	Cases (n = 9)	**Controls (n = 18)**	***P***	**Cases (n = 8)**	**Controls (n = 16)**	***P***
Average duration of breastfeeding (months)	10.2 (SD 7.69)	9.48 (SD 4.43)	0.5395[M]	9.12 (SD 7.59)	9.78 (SD 4.64)	0.9155[M]
Average age for solid food introduction (months)	4.52 (SD 1.27)	4.26 (SD 0.66)	0.4712[M]	4.57 (SD 1.34)	4.31 (SD 0.69)	0.8301[M]
Average age for gluten introduction (months)	5.68 (SD 0.84)	5.65 (SD 0.67)	0.5539[M]	5.61 (SD 0.86)	5.65 (SD 0.69)	0.6025[M]
Breast feeding during gluten introduction (% of subjects)	66,7 %	87.5 %	0.2113[P]	62.5 %	85.7 %	0.2113[P]
Antibiotic courses before 9 months (% of subjects)	37.5 %	35.3 %	0.9146[P]	28.6 %	26.67 %	0.9256[P]
Antibiotic courses before 12 months (% of subjects)	42.9 %	47.1 %	0.8511[P]	33.3 %	40.0 %	0.7763[P]
Antibiotic courses before 4 years (% of subjects)	87.5 %	77.8 %	0.5615[P]	85.7 %	75.0 %	0.5665[P]

[M]Mann-Whitney U test, [P]Pearson's chi-squared test

Figure Legends

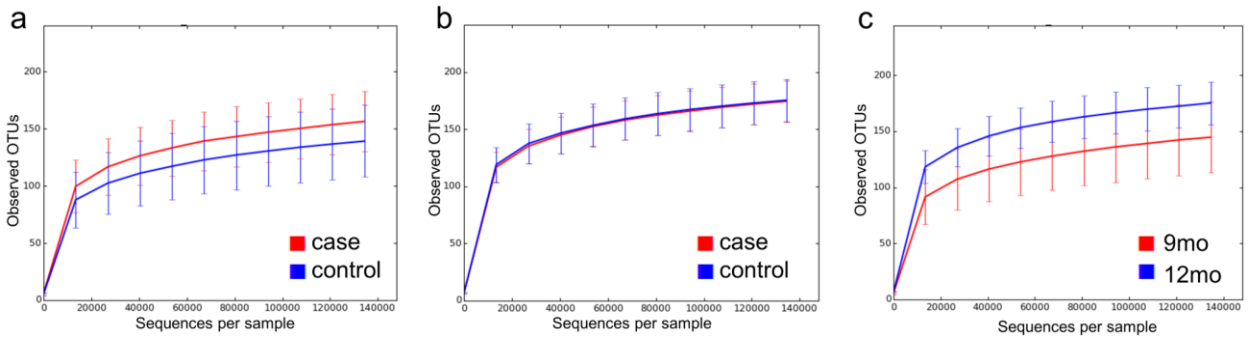

Figure 1: Bacterial diversity of the study samples, represented as observed different OTUs per sequences. At the age of nine months, case infants tend to have a slightly more diverse fecal microbiota than the controls (a), but this difference is statistically significant only when the infants born by cesarean section are excluded from the analysis. At the age of 12 months, no differences can be seen (b). By contrast, a clear difference can be observed in the bacterial diversity between the 9mo and 12mo samples (c). The rarefaction level is 134 414 reads per sample.

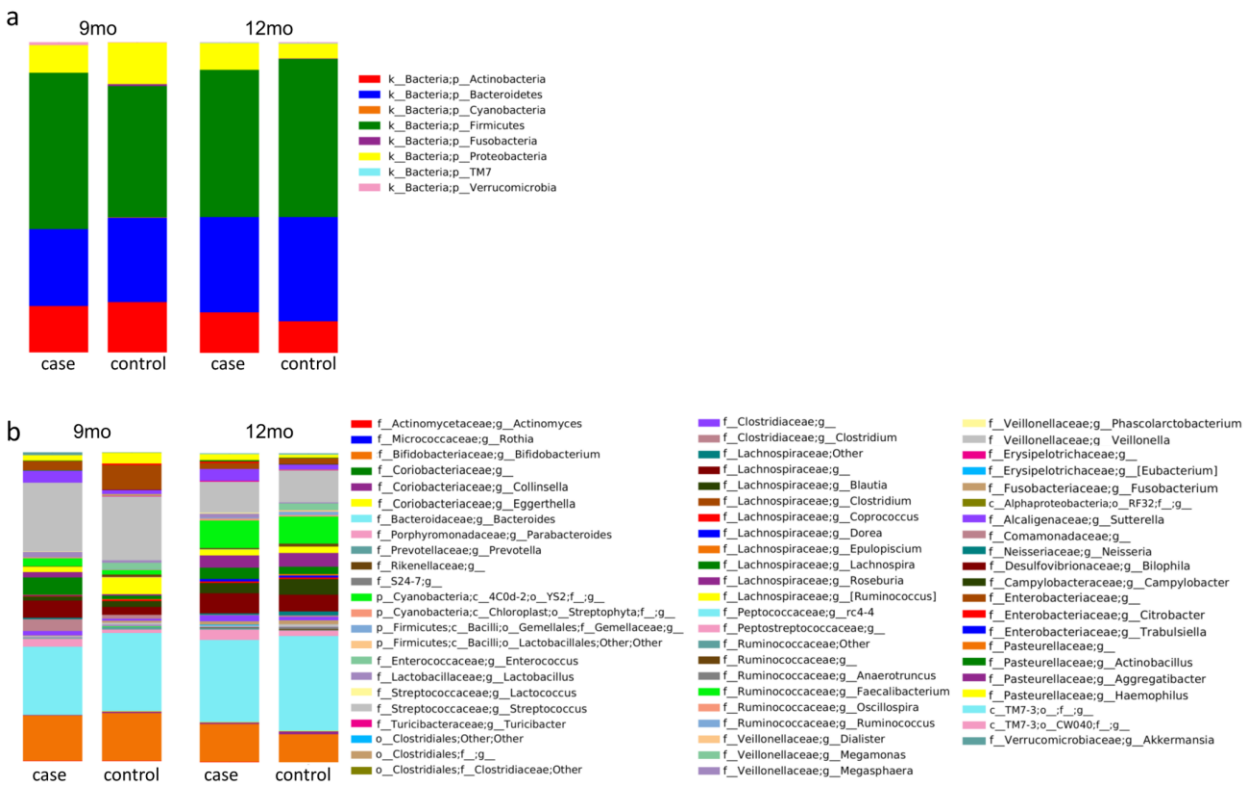

Figure 2: The average bacterial composition of the study samples. No significant differences can be visually observed in the phylum level bacterial composition neither between the case and control infants nor between the 9mo and 12mo samples (a). Differences in the genus level bacterial composition are more prominent; 9mo and 12mo samples differ clearly from each other, and in the 9mo samples, minor differences can be observed between the case and control infants (b). The rarefaction level is 134 414 reads per sample.

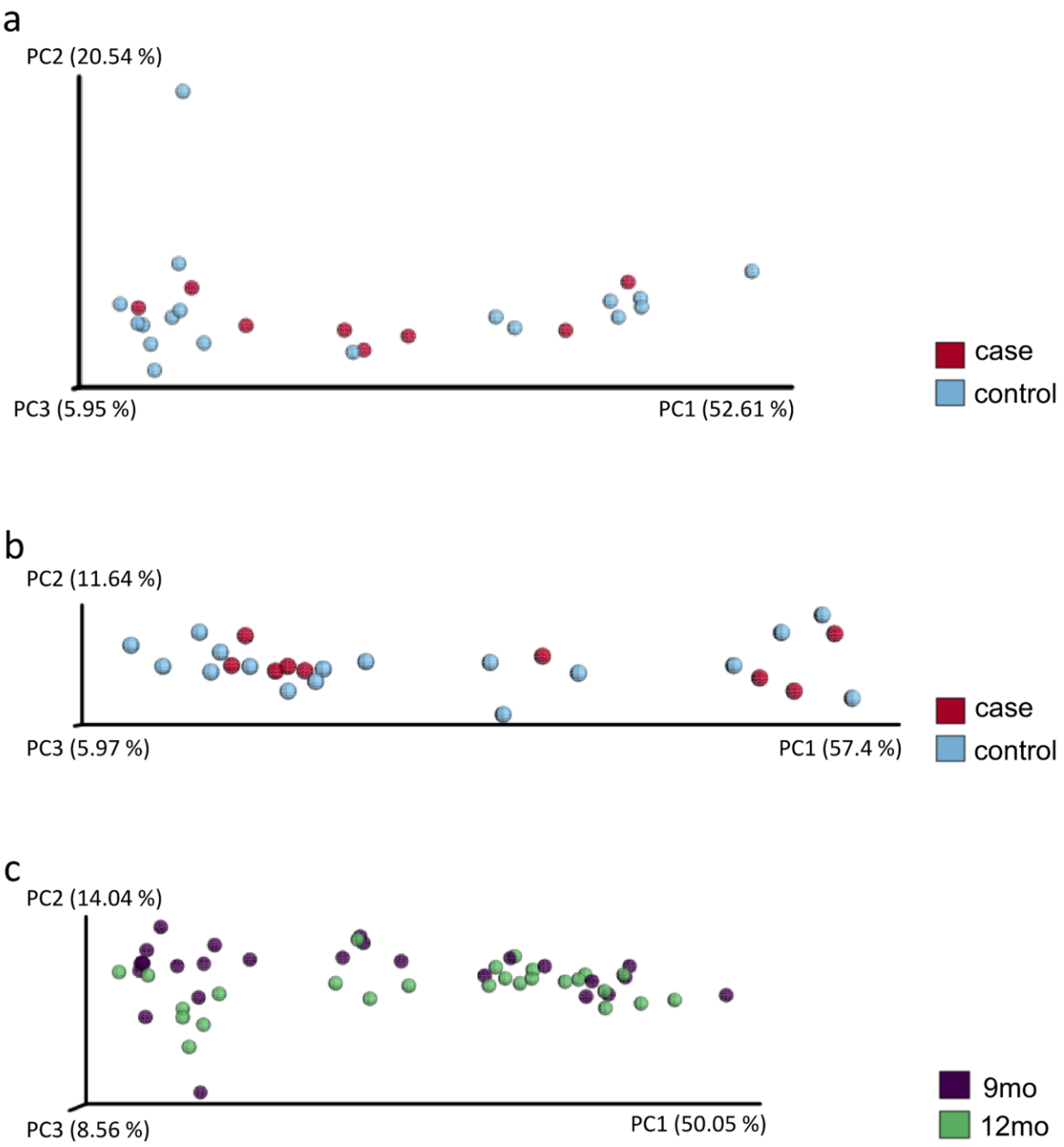

Figure 3: Principal Coordinate Analysis (PCoA) of the study samples. Case and control infants do not cluster separately in the PCoA in either 9mo (a) or 12mo (b) samples. Furthermore, 9mo and 12mo samples do not form clearly separate clusters in the PCoA (c). The rarefaction level is 134 414 reads per sample.