Trackability of proteins from probiotic *Bifidobacterium* spp. in the gut using metaproteomics

**D.M. Hendrickx**1*, R. An1, S. Boeren2, S.K. Mutte2, the PRESTO study team, H. Wopereis3 and C. Belzer1*

1Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, the Netherlands; 2Laboratory of Biochemistry, Wageningen University, P.O. Box 8128, 6700 ET Wageningen, the Netherlands; 3Danone Nutricia Research, Uppsalalaan 12, 3584 CT Utrecht, the Netherlands; *diana.hendrickx@wur.nl; clara.belzer@wur.nl

Received 9 December 2022 | Accepted 7 June 2023 | Published online 1 September 2023

**Abstract**

Beneficial effects of *Bifidobacterium* spp. on gut microbiota development and infant health have been reported earlier. Therefore, supplementation of infant formulas with probiotic bifidobacteria, as well as prebiotics stimulating bifidobacterial growth, has been proposed. Here, we studied the faecal microbiome of infants supplemented with specialized nutrition, of which some received a standard amino acid-based formula (AAF) and others an AAF with a specific mixture of prebiotics and a probiotic (synbiotics) using metaproteomics and 16S rRNA gene sequencing. Faecal samples were taken at baseline, as well as after 6 and 12 months fed with the specialized formula. The aim was to compare microbial differences between infants treated with the standard AAF and those who received the additional synbiotics. Our findings show that infants who received AAF with synbiotics have significantly higher levels of *Bifidobacteriaceae* DNA as well as significantly increased levels of *Coriobacteriaceae* proteins, over time. Moreover, at visit 12 months higher levels of some bifidobacterial carbohydrate-active enzymes, known to metabolize oligosaccharides, were observed in the synbiotic group compared to the non-synbiotic group. The results indicate that the synbiotics provided in our study are trackable in faecal samples on the DNA and protein level during the intervention period.

**Keywords**

infant formula – synbiotics – microbiome

1 **Introduction**

*Bifidobacterium* spp. have been reported to play a central role in gut microbiota development in early life (Kumar *et al.*, 2020). They are important for the development of the immune system (Hart *et al.*, 2004) and for the maintenance of health (Turroni *et al.*, 2012). They are known to have beneficial effects on human health, like reducing allergy symptoms (Singh *et al.*, 2013), preventing gastrointestinal infections (Weizman *et al.*, 2005) and reducing cancer risk (Rafter *et al.*, 2007).

Because of their beneficial effects, probiotic *Bifidobacterium* strains have been added to infant formula. Bazanella *et al.* (2017) showed that supplementation of infant formula with a mixture of bifidobacterial (*Bifidobacterium bifidum* BF3, *Bifidobacterium breve* BR3, *Bifidobacterium longum* BG7, *B. longum* subspecies...
infantis BT1) influences gut microbiota and metabolism in early life. Escribano et al. (2018) studied the effect of probiotic B. longum subsp. infantis CECT7210 and showed a reduction in diarrhoea as well as a lower constipation prevalence. Chi et al. (2020) reported a beneficial effect of Bifidobacterium lactis on the gut microbiota in infants with a low birth weight. Holscher et al. (2012) showed a beneficial effect of B. lactis Bb12 on the intestinal antibody response. Taipale et al. (2011) reported a reduction of respiratory infections in infants that received a formula with Bifidobacterium animalis subsp. lactis BB-12.

Often, probiotic strains are combined with prebiotic oligosaccharides that stimulate bifidobacterial growth (Marini et al., 2003). Candy et al. (2018) showed that amino acid-based formula (AAF) supplemented with a symbiotic mixture comprising of prebiotic fructo-oligosaccharides (scFOS/lcFOS; 9:1) and probiotic B. breve M-16V (AAF-syn) has a beneficial effect on the composition of the gut microbiota of infants with non-IgE-mediated cow’s milk protein allergy (CMPA) (Candy et al., 2018).

Previous studies showed that the AAF-syn was hypoallergenic and supported normal growth and weight gain (compared to WHO standards) in healthy, full-term infants, and in infants with IgE- or non-IgE-mediated CMA (Burks et al., 2015; Candy et al., 2018; Harvey et al., 2014). The number of children developing tolerance to cow’s milk was similar comparing the AAF-syn with AAF and in line with the natural rates of CMA outgrowth (Chatchatee et al., 2022). The same studies (Burks et al., 2015; Chatchatee et al., 2022; Harvey et al., 2014) also confirm that both formulas are safe based on (serious) adverse events, medication use and blood parameters and that they are well-tolerated (based on gastrointestinal symptoms). Interestingly, the AAF-syn was recurrently associated with fewer infectious events and use of medications including antibiotics, suggesting a clinical benefit of the specific symbiotic beyond the modification of gut microbiota (Sorensen et al., 2021).

Faecal metaproteomics can provide insight in the composition of gut microbiota, as well as their functional roles (Kingkaw et al., 2020; Kleiner, 2019). However, metaproteomics studies on the effect of synbiotics are scarce. Ke et al. (2019) combined metaproteomics with 16S rRNA gene sequencing (16S rRNASeq) and metabolomics to study the effect of synbiotics on the gut microbiota in mice. Korpela et al. (2018) analysed faecal samples of C-section born infants to investigate the effect of probiotic supplementation on microbiota composition.

In this study, we combined 16S rRNASeq with faecal metaproteomics to investigate presence of proteins of the probiotic and the changes in the microbiome related to feeding with AAF-syn in comparison with AAF.

2 Materials and methods

Study description

For this study, stool samples from a subset of 40 subjects who participated in the PRESTO study (NTR3725) (Chatchatee et al., 2022) were retrieved from Danone Nutricia research as described elsewhere (Hendrickx et al., 2022). In short, the PRESTO study included infants aged ≤ 13 months with confirmed IgE-mediated CMA who were randomised to receive specialised nutrition (Chatchatee et al., 2022). Of the 40 subjects selected, 23 received AAF-syn and 17 received AAF. The symbiotic blend was composed of probiotic Bifidobacterium breve M-16V (1.47 × 10⁹ cfu/100 ml) and prebiotic oligosaccharides (oligofructose and inulin in a 9:1 ratio, total concentration of 0.63 g/100 ml) (Chatchatee et al., 2022).

Faecal samples of the 40 infants were analysed by metaproteomics and 16S rRNASeq at three visits: baseline (before start of the study), and after 6 and 12 months (6M and 12M) of intervention. Clinical characteristics of the 40 infants are provided in Supplementary Table S1. One infant was excluded because of unknown allergy status at visit 12M as described previously (Hendrickx et al., 2022).

This multicenter study was performed according to the World Medical Association (WMA) Declaration of Helsinki and the International Conference on Harmonization guidelines for Good Clinical Practice (Chatchatee et al., 2022). Ethical approval of the study protocol and amendments, and written informed consent from the parents of the infants were obtained before the start of the study (Chatchatee et al., 2022). A detailed list of all national ethics committees, institutional review boards and regulatory authorities that approved the study is provided in the Supplementary Table S0.

16S rRNA gene amplificon sequencing and pre-processing

Sequencing of the V3-V4 region of the 16S rRNA gene and pre-processing of the raw data was performed as described elsewhere (Hendrickx et al., 2022). In sum-
mary, read pairs obtained using the 2 × 300 bp paired-end MiSeq protocol (Illumina) were demultiplexed and trimmed to remove low-quality sequences. The trimmed reads were dereplicated and counted, and after which filtering was applied to remove reads with less than two reads over all samples, chimeras, reads including PhiX and reads including Adapter sequences. Taxonomy at the genus level was assigned to amplicon sequence variants (ASVs) using the SILVA 138 database (Quast et al., 2013), resulting in 173 genera.

**Processing of samples for metaproteomics analysis**

Sample preparation, nLC-MS/MS and protein identification were performed as described elsewhere (Hendrickx et al., 2022). In summary, after protein extraction and protein digestion into peptides, samples were analysed with nLC/MS-MS. The obtained MS/MS spectra were analysed with MaxQuant (Cox and Mann, 2008) using the Andromeda peptide search engine (Cox et al., 2011) and an in-house proteomics database (Hendrickx et al., 2022), which after strict filtering, left 2705 protein groups with at least 2 identified peptides of which at least one should be unique and at least one should be unmodified. Of these 2705 protein groups, 2481 were from bacteria.

**Statistical analysis**

Statistical analyses were performed with R version 3.6.1 (R Core Team, 2019). Aggregation of ASVs at the family level and conversion to relative abundances was performed with the R microbiome package (Lahti and Shetty, 2012-2019) version 1.8.0. The relative abundances for the metaproteomics data were obtained from the Intensity Based Absolute Quantification (iBAQ) intensities and aggregated at the family level. Core microbiota at the family level were identified with the R microbiome package and defined as families with a relative abundance higher than 1% in at least half of the 16S rRNAseq or metaproteomics samples.

Protein identifiers of bacterial proteins were assigned to KEGG (Kanehisa and Goto, 2000) orthology (KO) identifiers. Subsequently, the KO identifiers were classified using KEGG Brite hierarchy level c, to obtain functional annotation for the proteins. Furthermore, we selected proteins that were classified as carbohydrate-active enzymes (CAZymes) using the CAZY database (http://www.cazy.org/, Drula et al., 2022).

Relative abundances of core taxa, top 10 microbial functional protein classes, top 10 human protein classes, amino acid (AA) metabolism proteins and CAZymes were compared between interventions (AAF-syn versus AAF) and between visits using linear mixed models (LMM), with intervention, visit and intervention by visit as fixed effects and subject as random effect. CAZymes and AA metabolism proteins were studied separately for bifidobacterial and other bacteria. Prior to LMM analysis, the data were normalised using centred log-ratio (CLR) transformation (Aitchison, 1986). P-values for pairwise comparison were corrected for multiple testing using the method of Benjamini-Hochberg (Benjamini and Hochberg, 1995).

To determine discriminative microbial features between AAF-syn versus AAF, linear discriminant analysis Effect Size (LEfSe) analysis (Segata et al., 2011) with an alpha value of 0.05 for the factorial Kruskal-Wallis test and a logarithmic LDA score threshold of 2.0 was applied. To compare visits within each intervention group, paired LEfSe analysis was performed by including information on which samples belonged to the same subject. All LEfSe analyses were performed through the online Galaxy module (http://huttenhower.sph.harvard.edu/galaxy). To determine how much variance was explained by the intervention, we performed partial RDA with intervention as explanatory variable and the remainder of the environmental variables as covariates.

### 3 Results

**Differences between AAF-syn and AAF in 16S rRNA gene-based relative abundance levels**

LMM and LEfSe analysis were conducted on the 16S rRNA gene-based relative abundances at the taxonomic family level to determine microbial variables most likely explaining differences between AAF-syn and AAF. At baseline no significant differences between both groups were observed (Supplementary Table S2). At 6 and 12M, both analyses identified significantly higher Bifidobacteriaceae in the AAF-syn compared to the AAF group (Supplementary Table S2, Figure 1 and 2, Supplementary Figure S1). Furthermore, significantly lower Bacteroidaceae were observed by LMM in the AAF group compared to the AAF-syn group at 6M (Supplementary Table S2). Moreover, at 6M, LEfSe identified higher Morganellaceae, Enterococcaceae and Lactobacillaceae, and lower Sutterellaceae, Monoglobaceae and Butyricicoccaceae in the AAF-syn compared to the AAF group (Figure 1). Furthermore, LEfSe identified significantly lower Akkermansiaceae, Verrucomicrobiaceae and Fusobacteriaceae at 12M in the AAF-syn compared to the AAF group (Figure 2).
higher Bifidobacteriaceae at 6M compared to baseline were observed by both analyses in the AAF-syn group (Supplementary Table S3, Supplementary Figure S2A). Moreover, LEfSe identified higher Bifidobacteriaceae at 12M compared to baseline in the AAF-syn group (Supplementary Figure S2B). No significant differences in Bifidobacteriaceae between visits were observed in the AAF group (Supplementary Table S3, Supplementary Figures S2D-F). In both intervention groups, both analyses showed lower Enterobacteriaceae at 12M compared to baseline (Supplementary Table S3, Supplementary Figures S2B and E). LEfSe also identified lower Enterococcaceae in the AAF-syn group at 12M compared to 6M, but not in the AAF group (Supplementary Figure S2A-D). Bacteroidaceae increased in both intervention groups (Supplementary Table S3). However, only the overall increase of Bacteroidaceae between baseline and 12M in the AAF-syn, but not in the AAF group was found significant by both analyses (Supplementary Table S3, Supplementary Figures S2B and E). For the increase in Bacteroidaceae between baseline and 6M in the AAF group, significance was only found by LEfSe (Supplementary Figures S2A and D), while LMM found marginal significance (Supplementary Table S3).

In the AAF-syn, but not in the AAF group, higher Micrococcaceae at 6M and 12M compared to baseline were obtained by LEfSe (Supplementary Figure S2A-B and D-E). In both groups, higher Butyricicoccaceae, Oscillospiraceae, Monoglobaceae, Rikenellaceae and Aerococcaceae were found at 12M compared to baseline (Supplementary Figure S2B and D). For the AAF-syn group, higher Butyricicoccaceae were obtained at 12M compared to 6M (Supplementary Figure S2C), while for the AAF group Butyricicoccaceae and Oscillospiraceae were higher at 6M compared to baseline (Supplementary Figure S2B and D). For the AAF-syn group, higher Butyricicoccaceae were obtained at 12M compared to 6M (Supplementary Figure S2D). For the AAF group, Prevotellaceae were higher at 12M than at 6M (Supplementary Figure S2B-C). In the AAF group, Sulfolobaceae were higher at 12M than at baseline and 6M in the AAF-syn group only (Supplementary Figure S2B-C). In the AAF-syn group, Prevotellaceae were higher at 12M compared to baseline, and lower at 12M compared to 6M (Supplementary Figure S2B-C). In the AAF group, Prevotellaceae were higher at baseline and 12M than at 6M (Supplementary Figure S2D and F). Aerococcaceae was higher at 12M than at 6M in the AAF group only (Supplementary Figure S2F). Christenellaceae were higher at 12M than at baseline and Coriobacteriaceae were higher at 12M than at 6M in the AAF-syn, but not in the AAF group. In the AAF, but not in the AAF-syn group, Enterococcaceae were higher at baseline than at 6M and 12M (Supplementary Figures S2D-E). Furthermore, Morganellaceae were higher, and Desulfovibrioaceae and Ruminococcaceae were lower at baseline than at 6M in the AAF group only (Supplementary Figure S2D). Moreover, higher Ruminococcaceae, Acidaminococcaceae, Akkermansiosiaceae and Verrucomicrobiaceae at 12M than at baseline were identified in the AAF, but not in the AAF-syn group (Supplementary Figure S2E).

Differences between AAF-syn and AAF in metaproteomics-based relative abundance levels
To determine significant differences in relative protein abundances at the taxonomic family level between
interventions, we also conducted a LMM analysis on the core taxa and a LEfSe analysis on the whole data set. At baseline, no significant differences between AAF-syn and AAF could be observed (Supplementary Table S4). At 6M and 12M, both LMM and LEfSe identified significantly higher *Bifidobacteriaceae* proteins in the AAF-syn group compared to the AAF group (Supplementary Table S4, Figures 3-4 and Supplementary Figure S3). At 12M but not at 6M, LMM identified significantly lower relative protein abundances of *Coriobacteriaceae* in the AAF-syn group than in the AAF group at both 6M and 12M (Figures 3-4 and Supplementary Figure S3). Lower relative protein abundances of *Coriobacteriaceae*, *Oscilllospiraceae* and *Tannerellaceae* in the AAF-syn group compared to the AAF group were only identified at 6M and not at 12M (Figure 3 and Supplementary Figure S3). For the following families, lower relative protein abundances in the AAF-syn group than in the AAF group were only obtained at 12M (Figure 4 and Supplementary Figure S3): *Ruminococcaceae*, *Bacteroidaceae*, *Enterobacteriaceae* and *Prevotellaceae*.

LMM identified significantly lower relative abundances of *Enterobacteriaceae* at 12M compared to baseline (Supplementary Table S5). In the AAF-syn group, but not in the AAF group, LEfSe identified significantly higher *Enterobacteriaceae* and *Enterococcaceae* proteins at baseline and 6M compared to 12M (Supplementary Figure S4A-B). Moreover, higher *Coriobacteriaceae* proteins were identified at 12M compared to 6M in the AAF-syn group only by both LMM and LEfSe (Supplementary Figure S4B and Supplementary Table S5). In both intervention groups, higher relative protein abundances of *Coriobacteriaceae* were obtained by LEfSe at 12M compared to baseline (Supplementary Figure S4A and D). In the AAF-syn group only, lower *Prevotellaceae* were identified at 12M compared to 6M (Supplementary Figure S4B). In the AAF-syn group only, the LEfSe results show higher *Enterococcaceae* proteins at baseline and 6M compared to 12M (Supplementary Figure S4A-B). Moreover, higher *Tannerellaceae* proteins at 6M than at baseline, higher *Rikenellaceae* proteins at 12M than at baseline, and lower relative abundances of *Eggerthellaceae* proteins at 12M than at 6M were identified in the AAF group only (Supplementary Figure S4C-E).

**Intervention (AAF-syn or AAF) is predictive for microbial signatures at 6 and 12 months**

In a previous study (Hendrickx *et al.*, 2022), we showed through redundancy analysis (RDA) that intervention (AAF-syn versus AAF) significantly contributes to the variance in microbial proteome profiles and 16S rRNA gene-based taxonomy at 6M. For all types of profiles, except protein-based taxonomy, a significant contribution of intervention was identified at 12M. In this study, these findings were further analysed by partial RDA with intervention as explanatory variable and the remainder of the environmental variables as covariates. Significance was obtained for proteome profiles at 12M (Supplementary Table S6). As a consequence, the partial RDA plot for proteome profiles at 12M shows the clearest separation between interventions, while
for the other partial RDA plots there is some overlap between the two groups (Supplementary Figure S5).

In all taxonomy-based partial RDA plots, Bifidobacteriaceae is among the variables that together explain 80% of the total variance in the ordination space.

**Significant differences between AAF-syn and AAF at the microbial functional protein level**

The results of LMM conducted on the relative abundances of the top 10 microbial functional classes show significantly lower relative abundances of proteins from the fructose and mannose metabolism in the AAF-syn group than in the AAF group at visit 6M (Supplementary Table S7). In the AAF-syn, but not in the AAF group, a significantly higher relative abundance of glycolysis/glucoseoneogenesis was found at visit 12M compared to baseline (Supplementary Table S8). In the AAF group, but not in the AAF-syn group, a significantly lower relative abundance of ABC transporters was obtained at visit 12M compared to baseline (Supplementary Table S8). For both intervention groups, a significantly lower relative abundance of RNA degradation proteins was obtained at visit 12M compared to baseline (Supplementary Table S8). When repeating the LMM analysis with all obtained functional classes instead of only the top 10, significantly higher relative abundances of propanoate metabolism, and significantly lower relative abundances of tryptophan metabolism and cyanoamino acid metabolism were found at 12M compared to baseline in the AAF-syn group, but not in the AAF group (Supplementary Table S9).

**Differences between AAF-syn and AAF in relative abundance levels of top 10 human protein classes**

Significant differences in relative abundance of the top 10 human protein classes between interventions within each visit, and between visits within each intervention were determined by LMM. At baseline and 6M, serine peptidases (total of 8 protein groups) were significantly higher in the AAF-syn group than in the AAF group (Supplementary Table S10). At 6M, glycoside hydrolases (total of 6 protein groups) were significantly higher in the AAF-syn than in the AAF group (Supplementary Table S10). Within the AAF-syn group, significantly higher proline-rich proteins (total of 4 protein groups), and significantly lower S100 proteins (total of 4 protein groups) were observed at 6M and 12M compared to baseline (Supplementary Table S11). Immunoglobulins (total of 37 protein groups) were significantly lower at 12M compared to baseline in the AAF-syn group (Supplementary Table S11). Furthermore, in the AAF-syn group, transthyretin/hydroxyisourate hydrolases (1 protein, transthyretin) were significantly higher at 12M compared to 6M (Supplementary Table S11). No significant differences between visits were observed within the AAF group (Supplementary Table S11).

**Activity of Bifidobacterium spp.**

As the AAF-syn group received a probiotic strain of *Bifidobacterium breve* an LMM analysis of relative abundance levels of *Bifidobacterium* spp. proteins, in particular CAZymes and AA metabolism proteins, was performed. At 12M, significantly higher relative abundances of the CAZymes alpha-1,4 glucan phosphorylase and beta-galactosidase (from *B. breve*) were identified in the AAF-syn compared to the AAF group (Supplementary Table S12). In the AAF-syn group, but not in the AAF group, significantly higher relative abundances of beta-galactosidase (from *B. breve*) were observed at visit 12M compared to baseline (Supplementary Table S13).

At 6M, significantly higher 6-phosphogluconate dehydrogenase, originating from *B. breve*, was observed in the AAF-syn compared to the AAF group (Supplementary Table S14). There were also several AA metabolism proteins that showed significantly higher relative abundances in the AAF-syn compared to the AAF group at 12M: alanine aminopeptidase (from *B. breve*), aldehyde-alcohol dehydrogenase (from *B. breve*), arginine biosynthesis bifunctional protein, argininosuccinate lyase (from *B. breve*), branched-chain-amino-acid transaminase (from *B. breve*) and homoserine dehydrogenase (Supplementary Table S14). No significant differences between visits within were observed within the two intervention groups (Supplementary Table S15).

**Activity of proteins from other bacteria**

An LMM analysis was also performed on relative abundance levels of CAZymes and AA metabolism proteins of other bacteria. At 6M, a significantly lower relative abundance of N-acetylglucosamine-6-phosphate deacetylase (from *Blautia* spp.) was found for the AAF-syn group compared to the AAF group (Supplementary Table S16). In the AAF-syn group, but not in the AAF group, a significantly lower relative abundance of alpha-1,4 glucan phosphorylase (from *Aneaeostipes, Clostridium* and *Ruminococcus* spp.) was found at visit 12M compared to baseline (Supplementary Table S17).

At 12M, lower relative abundance levels of the following non-bifidobacterial AA metabolism proteins were identified in the AAF-syn group compared to the AAF group: methionine-tRNA ligase (from *Blautia* spp.) and
NADP-specific glutamate dehydrogenase (from *Megasphaera* spp.) (Supplementary Table S18).

No significant differences in non-bifidobacterial AA metabolism proteins were obtained between visits within the AAF-syn group (Supplementary Table S19). For the AAF group, relative abundances of non-bifidobacterial NADP-specific glutamate dehydrogenase (from the taxa mentioned above) were significantly higher at 12M than at baseline (Supplementary Table S19).

4 Discussion and conclusions

In this study, we applied 16S rRNAseq and metaproteomics on faecal samples from infants with IgE-mediated cow’s milk protein allergy, who received a standard AAF or AAF with specific synbiotics. LMM and LEfSe analysis were applied to compare infants that received a standard AAF with those that received an AAF with synbiotics.

At 6M and 12M, both analyses of the 16S rRNAseq data showed higher *Bifidobacteriaceae* in the AAF-syn than in the AAF group. Similar results were obtained for metaproteomics-based relative abundances. Moreover, higher 16S rRNA gene-based relative abundances of *Bifidobacteriaceae* were obtained at 6M compared to baseline by both analyses in the AAF-syn group only. Furthermore, LEfSe identified also higher 16S rRNA gene-based relative abundances of *Bifidobacteriaceae* at 12M compared to baseline in the AAF-syn, but not in the AAF group.

Furthermore, *Bifidobacteriaceae* was among the variables that explained 80% of the total variance in the partial RDA plots. The findings indicate both the delivery of the probiotic *B. breve* strain and the induced growth and activity of bifidobacteria (endogenous species already present in the gut and *B. breve* M-16V delivered via the study product) by the prebiotic component.

At 6M, significantly lower 16S rRNA gene-based relative abundances of *Bacteroidaceae* were obtained by LMM in the AAF-syn group compared to the AAF group. In contrast, higher levels of members of the *Bacteroidaceae* family (*Bacteroides*) were observed in toddlers that received fortified milk with a symbiotic blend (*B. breve* M-16V, long-chain galactooligosaccharides (lcGOS) and short-chain fructooligosaccharides (scFOS)) at 12M of intervention (Matsuyama *et al.*, 2019). However, these infants were older than those in our study and did not have CMA.

For both AAF-syn and AAF, both analyses showed lower 16S rRNA-gene based relative abundances of *Enterobacteriaceae* at 12M compared to baseline. At the metaproteome level, this result was only obtained for AAF-syn. Decreasing relative abundances of *Enterobacteriaceae* after feeding with a symbiotic blend (scGOS/lcFOS and *B. breve* M-16V) have also been observed in infants delivered by C-section (Chua *et al.*, 2017). Decrease of *Enterobacteriaceae* has also been reported as an effect of infants growing older (Laursen, 2021).

Higher 16S rRNA gene-based relative abundances of *Coriobacteriaceae* were observed at 12M compared to 6M for the synbiotics treatment only. For the metaproteomics analysis, higher relative abundances of *Coriobacteriaceae* were observed at 12M compared to baseline for both intervention groups. Increased relative abundance of *Coriobacteriaceae* by symbiotic treatment (scGOS/lcFOS and *B. breve* M-16V) has been reported earlier in healthy toddlers (Matsuyama *et al.*, 2019).

LEfSe showed lower *Lachnospiraceae* proteins at both 6M and 12M, and lower *Bacteroidaceae* proteins at 12M in the AAF-syn group than in the AAF group. Lower relative abundance of members of the *Lachnospiraceae* family (*Coproccoccus, Blautia* and *Dorea*) were also reported in healthy toddlers after 3 months of feeding with scGOS/lcFOS and *B. breve* M-16V compared to control (Matsuyama *et al.*, 2019). However, the same study reported also higher relative abundance of *Lachnospiraceae* – [Ruminococcaceae] genus after 3 months of intervention compared to control, and higher levels of *Dorea* and *Bacteroidaceae – Bacteroides* after 12M intervention compared to control. Decreased relative abundances of *Lachnospiraceae* (*Tyzzerella, Blautia* and *Lachnoclostridium*) were also observed in infants with non-IgE mediated (delayed) cow’s milk allergy after 8 weeks of treatment with an AAF with a symbiotic blend (prebiotic fructo-oligosaccharides and probiotic *B. breve* M-16V) compared to treatment with a standard AAF (Wopereis *et al.*, 2019).

At 6M, but not at 12M, lower *Coriobacteriaceae* proteins were found in the AAF-syn compared to the AAF group. In healthy toddlers, members of the *Coriobacteriaceae* family (*Colinsella*) were reported to be higher in the symbiotic than in the control group after 3 months of intervention (Matsuyama *et al.*, 2019).

At 12M, but not at 6M, our results showed lower *Ruminococcaceae* proteins in the AAF-syn compared to the AAF group. Decreased relative abundances of *Ruminococcaceae* (*Ruminococcus*) were also observed in infants with non-IgE mediated cow’s milk allergy after 8 weeks of treatment with an AAF with synbiotics.
compared to a standard AAF (Wopereis et al., 2019). In healthy toddlers, the lower relative abundance of Ruminococcus in the synbiotic group compared to the control was also found after 3 months of intervention, while after 12M of intervention the effect was opposite (Matsuyama et al., 2019).

Normal development of the gut microbiota in early life consists of a first stage where the gut microbiome is dominated by Enterobacteriaceae, and a second stage, where Enterobacteriaceae decrease an anaerobic bacteria such as members of the Bifidobacteriaceae, Bacteroidaceae, Ruminococcaceae and Clostridiaceae families increase (Matamoros et al., 2013). At the 16S rRNA gene-based taxonomy level, we observed a similar profile after the introduction of AAF-syn: a decrease in Enterobacteriaceae and an increase in anaerobic bacteria of the families mentioned above (Supplementary Table S3, Supplementary Figure S2). After the introduction of the AAF, we observed a decrease instead of an increase of Bifidobacteriaceae (Supplementary Table S3, Supplementary Figure S2).

At the metaproteomics-based taxonomy level, a slightly different profile was observed after introducing the AAF-syn: Bifidobacteriaceae and Ruminococcaceae proteins increased, while both Bacteroidaceae and Enterobacteriaceae proteins decreased (Supplementary Table S5, Supplementary Figure S4). After introducing the AAF, Bifidobacteriaceae and Enterobacteriaceae proteins decreased, while Bacteroidaceae and Ruminococcaceae increased (Supplementary Table S5, Supplementary Figure S4). These results suggest that after treatment with AAF-syn, gut microbiota development is more similar to normal development in healthy infants than after treatment with AAF.

The analysis of the top 10 microbial functional protein classes identified some significant results. In particular, the relative abundance of proteins from glycolysis/gluconeogenesis was significantly higher at 12M than at baseline in the AAF-syn, but not in the AAF group. When repeating the analysis with all functional protein classes, several lower abundant classes showed significant differences between visit 12M and baseline for the AAF-syn group, but not for the AAF group (Supplementary Table S9). This means that differences in microbiota composition are not only reflected in the top 10, but that several lower abundant classes are also informative.

Apart from microbial changes, also some changes in the human proteome could be associated with synbiotic intervention. At 12M, some bifidobacterial CAZymes were significantly higher in the AAF-syn than in the AAF group. There were also several bifidobacterial AA metabolism proteins that showed higher relative abundances in the AAF-syn than in the AAF group at one or both of the 6M and 12M visits. Bifidobacterial CAZymes are known to metabolize oligosaccharides (fructooligosaccharides (FOS)/galactooligosaccharides (GOS)) (Boger et al., 2019; Tanno et al., 2021). These results suggest that the proteins from the probiotic B. breve M-16V strain and the increased abundance of proteins from other endogenous bifidobacterial contribute to the degradation of prebiotics (as a proxy for metabolic activity).

When studying non-bifidobacterial protein activity, we observed that at visit 12M the relative abundance of several AA metabolism proteins was higher in the AAF than in the AAF-syn group. Moreover, higher non-bifidobacterial NADP-specific glutamate dehydrogenase was observed at 12M than at baseline. These results suggest that amino acids are mainly metabolized by Bifidobacterium spp. in the presence of a synbiotic, while they are mainly metabolized by other bacteria in absence of a synbiotic.

While studies based on 16S rRNA gene sequencing only are limited to the identification of taxonomic differences between two interventions, including metaproteomics also reveals differences in bacterial activity. In particular, we focused on differences in activity of CAZymes and AA metabolism proteins in our study. The validity and added value of a high-throughput metaproteomics approach has been shown for other types of dietary interventions, like dietary fibre versus fermented foods (Gonzalez et al., 2020), and a probiotic intervention in adults (Kolmeder et al., 2016).

In summary, infants that received an amino acid formula with synbiotics were characterized by significantly increased levels of Bifidobacteriaceae over visits, and a significant increase in Coriobacteriaceae proteins over visits.

Moreover, at 12M relative abundances of some bifidobacterial CAZymes, known to metabolize oligosaccharides, were significantly higher in the synbiotic than in the non-synbiotic group. This is desirable, as it suggests that proteins from the probiotic B. breve M-16V strain and the increased abundance of proteins from other endogenous bifidobacteria contribute to the degradation of prebiotics when applying a synbiotics intervention.
Supplementary Material

Supplementary material is available online at: https://doi.org/10.6084/m9.figshare.23560572

Table S0. List of ethics committees, institutional review boards and regulatory authorities that approved this study.

Table S1. Clinical characteristics.

Table S2. Differences in 16S rRNA gene-based relative abundance between interventions.

Table S3. Differences in 16S rRNA gene-based relative abundance between visits.

Table S4. Differences in metaproteomics-based relative abundance between interventions.

Table S5. Differences in metaproteomics-based relative abundance between visits.

Table S6. Results of partial RDA with intervention as explanatory variable.

Table S7. Differences of top 10 microbial protein functional classes between interventions.

Table S8. Differences of top 10 microbial protein functional classes between visits.

Table S9. Differences of top 10 human protein classes between interventions.

Table S10. Differences of top 10 human protein classes between visits.

Table S11. Differences of top 10 microbial protein classes between visits.

Table S12. Differences of bifidobacterial CAZymes between interventions.

Table S13. Differences of bifidobacterial CAZymes between visits.

Table S14. Differences of bifidobacterial AA metabolism proteins between interventions.

Table S15. Differences in of bifidobacterial AA metabolism proteins between visits.

Table S16. Differences of non-bifidobacterial CAZymes between interventions.

Table S17. Differences of non-bifidobacterial CAZymes between visits.

Table S18. Differences of non-bifidobacterial AA metabolism proteins between interventions.

Table S19. Differences of non-bifidobacterial AA metabolism proteins between visits.

Figure S1. Heat maps of discriminative features between treatments found by LEfSe analysis of 16S rRNA-based relative abundances.

Figure S2. LEfSe analysis of 16S rRNA-based relative abundances at the family level. Pairwise comparison between visits within the intervention groups.

Figure S3. Heat maps of discriminative features between treatments found by LEfSe analysis of metaproteomics-based relative abundances.

Figure S4. LEfSe analysis of metaproteomics-based relative abundances at the family level. Pairwise comparison between visits within the intervention groups.

Figure S5. Partial RDA with intervention as explanatory variable and adjusting for other environmental factors.

Acknowledgements

This study was part of the EARLYFIT project (Partnership programme NWO Domain AES-Danone Nutricia Research), funded by the Dutch Research Council (NWO) and Danone Nutricia Research (project number: 16490). Heleen de Weerd (Danone Nutricia Research) is gratefully acknowledged for pre-processing the 16S rRNAseq data. We thank Jolanda Lambert (Danone Nutricia Research) for project management and for providing the 16S rRNAseq, the clinical data and the samples for proteomics. We thank Guus Roeselers (Danone Nutricia Research) for his input in the study design, and Jan Knol (Danone Nutricia Research, Wageningen University) for his input during meetings.

Author contributions

DMH performed the statistical analysis and wrote the majority of the manuscript, with input of all co-authors. RA and SB performed the metaproteomics experiments. DMH and SKM developed the proteomics database. The PRESTO study team generated the clinical data. HW reviewed the manuscript in detail. CB designed the study, coordinated the project and participated in all phases of the research and article preparation. All authors contributed to writing the manuscript. All authors have approved the final article.

The PRESTO study team members are: P. Chatchatee, Pediatric Allergy and Clinical Immunology Research Unit, Division of Allergy and Immunology, Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand; A. Nowak-Wegrzyn, Allergy and Immunology, Department of Pediatrics, Gastroenterology and Nutrition, New York University Langone Health, US & Department of Pediatrics, Gastroenterology and Nutrition, Collegium Medicum, University of Warmia and Mazury, Olsztyn, Poland; L. Lange, St Marien Hospital,
under accession number PRJEB56782. Raw proteomics data are available from ProteomeXchange via the PRIDE (Vizcaíno et al., 2016) partner repository under accession number PXD037190. The R code for the statistical analysis in this study has been deposited in Gitlab (https://git.wur.nl/afsg-microbiology/publication-supplementary-materials/2022-hendrickx-et-al-earlyfit_presto_synbiotics_study).

References

growth and tolerance and hypoallergenicity of an amino acid-based formula with synbiotics. Pediatric Research 75:343-351. https://doi.org/10.1038/pr.2013.211


Kumar, H., Collado, M.C., Wopereis, H., Salminen, S., Knol, J. and Roeselers, G., 2020. The bifidogenic effect revisited—ecology and health perspectives of bifidobacterial colo-