



OPEN

Does entry to center-based childcare affect gut microbial colonization in young infants?

Gerben D. A. Hermes^{2,4}, Henrik A. Eckermann^{1,4}, Willem M. de Vos^{2,3} & Carolina de Weerth¹✉

Entry to center-based childcare (CC) at three months of life can be an important challenge for infants as it includes major stressors such as long maternal separations and frequently changing caregivers. Stress and the new environment may in turn alter the composition of the gut microbiota with possible implications for future health outcomes. As part of an ongoing longitudinal study, we investigated whether CC, as compared to being cared for by the parents at home, alters the composition of the gut microbiota, while accounting for known covariates of the infant gut microbiota. Stool samples of infants who entered CC ($n = 49$) and control infants ($n = 49$) were obtained before and four weeks after CC entrance. Using Redundancy analysis, Random Forests and Bayesian linear models we found that infant gut microbiota was not affected in a uniform way by entry to CC. In line with the literature, breastfeeding, birth mode, age, and the presence of siblings were shown to significantly impact the microbial composition.

The human gut microbiota refers to a complex and dynamic population of microorganisms that resides in the human gastrointestinal tract and has recently become object of much scientific endeavors. Intestinal bacteria as part of this ecosystem play central roles in human health and disease^{1,2}. They are essential for nutrition, intestinal function, the education of the developing immune system and the protection of the host from invading microbial pathogens^{3–5}. Via a bi-directional communication system, intestinal bacteria may even influence brain development and behavior^{2,6–8}.

Multiple factors contribute to the development of the human gut microbiota in infancy and the microbial composition becomes relatively stable within the first 3–5 years of life, although some reports describe a longer development phase^{9,10}. Because intestinal bacteria influence the development of important host physiological systems within this early and critical developmental time period¹¹ it is crucial to understand the factors that influence the establishment of the early gut microbiota. In the present study we will concentrate on an early life factor that can potentially disrupt healthy gut microbial colonization and negatively affect microbial composition: entry into childcare at the age of 3 months. In the following paragraphs, we will outline how gut microbial colonization occurs and why entering childcare at that young age may disrupt this process.

Gut microbiota development is a highly dynamic and individual process. The current consensus is that the first major exposure to microbes happens during the birthing process and is highly dependent on mode of delivery^{12–14}. The first inoculation during natural childbirth clearly resembles the maternal fecal microbiota, with potential input from the vagina and other parts of the urogenital tract¹⁴. In contrast, infants delivered through a Caesarean section (C-section) are colonized with common skin and environmental microbes¹². Nevertheless, this difference in microbiota composition between children born vaginally or by C-section seems to gradually decrease, although some later life impact has been reported^{15,16}.

The initial inoculum initiates a succession of events leading to the development of a child's own microbiome. In this dynamic process the microbial abundance increases over time, with large fluctuations in the microorganisms present and their relative abundance¹⁷. Diversity generally increases, aerobes are succeeded by facultative and then strict anaerobes and, roughly up until the introduction of the first solid foods, a well-constrained range of stereotypical bacteria emerge in the faeces. Exclusive breast-feeding generally selects for genera specialized

¹Donders Institute for Brain, Cognition and Behavior, Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands. ²Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands. ³Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

⁴These authors contributed equally: Gerben D. A. Hermes and Henrik A. Eckermann. ✉e-mail: Carolina.deWeerth@radboudumc.nl

in the utilization of complex human milk oligosaccharides, such as *Bifidobacterium*¹⁸ and to a lesser extent *Bacteroides* spp, as they can compete for the same ecological niche¹⁹. By studying both Western and non-Western populations it has been shown that differences exist with regards to community membership, but that the overall temporal dynamics are similar over populations, with aberrant development after C-section delivery, use of antibiotics or early termination of breast-feeding^{14,20–23}.

Microbial colonization is characterized by large inter- and intra-individual variability, with large, abrupt, community shifts with interludes of relative stability of varying lengths of time^{17,24–26}. Sometimes these shifts occur together with life events that likely instigate considerable environmental pressure, such as antibiotics use, fever, and introduction of formula feeding²⁶. Animal models that use maternal separation (MS) to induce early life stress, suggest that early life stress is another such environmental pressure that can induce shifts in the microbiota. MS provokes an adult depressive and anxiety-like phenotype, along with altered immune function, activation of the hypothalamic-pituitary-adrenal (HPA) axis and disruption of the offspring's microbiota^{27–31}. De Palma *et al.* demonstrated the importance of the microbiota to induce these behavioral changes in animals. In germ free mice HPA axis regulation was altered by early life stress, but for the actual induction of the behavioral changes the microbiota was required. This indicated that MS-induced changes in host physiology led to intestinal dysbiosis, which in turn was needed for behavioral changes to occur³².

The entrance into center-based childcare (CC) typically starts at three months of life in the Netherlands and includes early life stressors such as long maternal separations, new and frequently changing caregivers and peers, and exposure as well as adaption to a new physical environment. Entering CC produces significant increases in cortisol levels as compared to being cared for in the home environment. Cortisol levels continue to increase until at least a month after entering^{33–36}. CC has also been related to other symptoms and illnesses including diarrhea, respiratory illnesses, otitis media, and skin complaints³⁷. These findings indicate that entering CC at this young age can be an early life stressor for infants.

This study explores the effects of the entrance to CC on the developing microbiota of 3-month-old infants by comparing infants attending CC to infants being cared for at home. In line with the above-mentioned animal studies, we expected CC to be associated with changes in microbial composition. Specifically, we tested for differences in the relative abundances at a genus-like level (see Methods) as well as differences in alpha diversity by combining univariate Bayesian approaches with multivariate methods including the Random Forests machine learning algorithm that has previously been shown to be useful in microbiome data analyses²¹. We included breastfeeding as a potential protective factor for microbial development, and accounted for known confounders, namely birth mode, antibiotics, age, and the presence of siblings³⁸.

Methods

Participants. As part of a larger and ongoing longitudinal study (BIBO), 220 mothers were followed since the third trimester of pregnancy, to investigate the influences of early environmental and caregiving factors on child development. Uncomplicated singleton pregnancy, proficiency in the Dutch language, no drug use, and the absence of physical and mental health problems were criteria for initial inclusion. Eight of the 220 women were excluded due to preterm birth or for other medical reasons. In addition, 19 mothers discontinued participation in the study during the first three postpartum months because of personal circumstances. All remaining infants (N = 193) were healthy and born at full term (37 weeks). Infants who had used antibiotics (n = 4) were excluded from the current study. In the first four months of life, mothers collected 9 fecal samples from their infants. Two samples were available for use in this study: Ten weeks post-partum, before entrance to CC (PRE), and 4 weeks after the PRE sample (POST). After eliminating infants who did not provide stool samples for both time points, the final sample size consisted of 49 infants who entered CC (group CC) and 49 infants who were cared for at home (group HOME). Table 1 shows demographic variables for both groups. The age of the HOME group infants was slightly lower than that of the CC group, both for sample PRE ($p < 0.001$) as for sample POST ($p < 0.001$), using Welch's t-test. There were no significant differences between groups for any other of the shown variables. Within the CC group, infants varied in the number of half-days of childcare per week (Mdn = 4, IQR = 3–4). We tested in a separate analysis whether the number of half-days was associated with gut microbiota composition beyond just the grouping variable, but this effect did not modify the conclusions (Supplementary Table 1). This study and all its experimental protocols were approved by and carried out in accordance with the Ethical Committee of the Faculty of Social Sciences, Radboud University Nijmegen (ECG/AvdK/07.563). Informed consent was obtained from each mother.

Microbiota covariates. The mothers received diaries towards the end of their pregnancy with instructions to take weekly notes about breastfeeding and formula-feeding from week 1 until week 27 after birth. For each week, the average number of breast- and/or formula-feedings per day were noted. To determine the effect of breastfeeding we defined two breastfeeding variables: the average number of feedings per day before and during the investigative period (i.e. birth to when the PRE-sample was obtained and PRE to when the POST-sample was obtained). The average number of breastfeedings included the feeding of expressed breastmilk through a bottle. We included the breastfeeding variables, age, siblings and C-section as covariates in all linear models. For 5 infants (4 in CC (8%), 1 in HOME (2%)) breastfeeding data was missing completely. For the univariate analyses the missing values were imputed using multiple imputation (see methods). For the multivariate analysis and visualization the original data was used.

Feces collection, DNA isolation and microbiota profiling. The parents were instructed to collect the fecal samples at home and to store them at -20°C . For transportation, samples were kept in coolers and then stored at -20°C and later at -80°C before being processed at the Laboratory of Microbiology at Wageningen University. DNA isolation from fecal samples has been described elsewhere in detail³⁹. In brief, DNA was isolated

	CC (n = 49)	HOME (n = 49)	p-value
Gender			
Male	29	25	0.417
Female	20	24	
Age (days) PRE			
mean (sd)	87.8 ± 16.0	76.7 ± 6.3	<0.001
Min	58	68	
Max	123	90	
Age (days) POST			
mean (sd)	118.4 ± 16.1	106.5 ± 5.9	<0.001
Min	90	97	
Max	154	116	
Maternal Age (years)			
mean (sd)	32.9 ± 3.0	32.2 ± 3.6	0.306
Min	25.1	24.9	
Max	42.0	40.1	
Birthweight (grams)			
mean (sd)	3630.4 ± 508.9	3636.0 ± 438.4	0.955
Min	2708	2810	
Max	4600	4700	
Breastfeeding (Birth - PRE)			
mean (sd)	5.4 ± 2.9	5.7 ± 2.3	0.558
Min	0	0	
Max	11.4	8.9	
Breastfeeding (PRE - POST)			
mean (sd)	4.0 ± 2.9	3.8 ± 2.8	0.832
Min	0	0	
Max	8.5	8.2	
Formula-feeding (Birth - PRE)			
mean (sd)	1.5 ± 2.4	1.4 ± 2.1	0.558
Min	0	0	
Max	7.7	7.0	
Formula-feeding (PRE - POST)			
mean (sd)	1.8 ± 2.4	2.0 ± 2.3	0.832
Min	0	0	
Max	5.9	6.0	
Proportion breastfeeding (Birth - PRE)			
mean (sd)	0.8 ± 0.4	0.8 ± 0.3	0.558
Min	0	0	
Max	1	1	
Proportion breastfeeding (PRE - POST)			
mean (sd)	0.7 ± 0.5	0.6 ± 0.5	0.832
Min	0	0	
Max	1	1	
Siblings			
Yes	25	32	0.186
No	23	17	
C-section			
Yes	6	3	0.294
No	42	45	

Table 1. Descriptive statistics for demographic variables of infants and mothers included in the present study. *Notes.* CC = center-based childcare. Breastfeeding (Birth - PRE) = average number of breastfeedings per day for the time period between birth and when the first stool sample was obtained. Breastfeeding (PRE - POST) = average number of breastfeedings per day for the time period between the collection of the first and the second stool sample. Proportion breastfeeding (Birth - PRE) = proportion of total feedings per day that were breastfeeding for the time period between birth and when the first stool sample was obtained. Breastfeeding (PRE - POST) = proportion of total feedings per day that were breastfeeding for the time period between the collection of the first and the second stool sample was obtained. Siblings = presence of siblings at time of birth.

using a combination of column purification and Repeated-Bead-Beating. Purity and concentration of DNA were assessed with a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA). The analysis was then performed utilizing a previously benchmarked custom made, phylogenetic microarray, the Human Intestinal Tract Chip (HITChip)^{40,41}. The HITChip contains a duplicated set of 3,631 probes, which target the V1 and V6 hypervariable regions of the 16S rRNA gene of 1140 intestinal bacterial phylotypes. After extraction of DNA, the full-length 16S rRNA gene was amplified by PCR using primers T7prom-Bact-27-for and Uni-1492-rev⁴¹. This was followed by *in vitro* transcription and labelling of the resulting RNA with Cy3/Cy5 before hybridization to the array. The signal intensity data from the microarray hybridizations were collected from the Agilent G2505C scanner (Agilent Technologies) using the Agilent Feature Extraction software, version 10.7.3.1 and pre-processed using an in-house MySQL database and custom R scripts. Each scanner channel from the array was spatially normalized separately using polynomial regression, followed by outlier detection and filtering in each set of probes with a χ^2 test. Each sample was hybridized at least twice to ensure reproducibility. Duplicate hybridizations with a Pearson correlation <0.95 were not considered for further analysis. Microbiota profiles were summarized to genus-like 16S rRNA gene sequence groups with a sequence similarity $>90\%$ referred to as species and relatives ('et rel.'). Measurements of probes that belong to the same phylotype were normalized with Robust Probabilistic Averaging^{42,43}. Log10-transformed hybridization signals were used as a proxy for bacterial abundance.

Statistical analyses. Microbiota analysis. All analyses were performed in R version 3.5⁴⁴. Bacterial richness was calculated at the probe level by using an 80% quantile threshold for detection of each individual probe. Diversity within a sample is termed alpha diversity and was calculated using the Shannon metric. HITChip signals were transformed to relative abundance. To determine the dynamics of microbial groups we calculated their coefficient of variation (CoV). CoV is a standardized measure of dispersion defined as the ratio of the standard deviation (σ) to the mean (μ). For the redundancy analysis (RDA) and the Bayesian robust linear models we applied centered-log-ratio (clr) transformation⁴⁵. The clr-transformation of relative abundances allows for the application of statistical methods that have been developed for real random variables, such as RDA and the Bayesian models⁴⁶. To determine the multivariate effects of CC and the number of half-days in CC on overall microbiota composition, we performed redundancy analysis (RDA) while accounting for the following variables that are known to influence microbiota composition: breastfeeding (average number of feedings during and before the investigative period), birth mode (C-section vs natural birth), age in days, and the presence of siblings using the function *rda* from the *vegan* package⁴⁷.

To determine the univariate effects of CC entry on alpha diversity and each bacterial group individually, we performed Bayesian hierarchical robust linear models as described by Krushke *et al.*⁴⁸. Bayesian data analysis provides several advantages over classical null hypothesis group comparison methods. These include richer information about parameter estimates, as the method provides complete distributional information about model parameters such as means and standard deviations, including credibility intervals of all possible combinations of these parameters. Furthermore, Bayesian data analysis delivers more precise information about the uncertainty when estimating group differences. The robust linear model presented by Krushke *et al.* in particular has advantages over standard linear models: The model is able to accurately estimate the mean (μ) and standard deviation (σ) when outliers are present as it utilizes the student t-distribution instead of the gaussian distribution. Furthermore, standard linear models assume homogeneity of variance between groups while this assumption is often violated in the context of differential abundance testing⁴⁸. In addition, an environmental factor could lead to a change in the variance of a distribution rather than (only) a change in the mean. This possibility is not considered in standard models. Our model allows the standard deviation σ to vary between the groups by modeling it as a linear function of the grouping variables CC, time, siblings and birth mode. The model can be written as:

$$y_i \sim T(\nu, \mu, \sigma)$$

where

$$\begin{aligned} \mu_i = & \beta_{0[i|j]} + \beta_1 \times CC + \beta_2 \times time + \beta_3 \times CC \times time + \beta_4 \times age + \beta_5 \times bf + \beta_6 \\ & \times sib + \beta_7 \times csec + \beta_8 \times sib \times csec \text{ and } \sigma_i = \beta_{\sigma 0} + \beta_{\sigma 1} \times CC + \beta_{\sigma 2} \times time \\ & + \beta_{\sigma 3} \times CC \times time + \beta_{\sigma 4} \times sib + \beta_{\sigma 5} \times csec + \beta_{\sigma 6} \times sib \times csec \end{aligned}$$

Besides μ (mean) and σ (standard deviation (SD)), the model consists of the parameter ν , which represents the normality parameter (low values lead to long and heavy tails, whereas as ν increases, the distribution approaches the gaussian distribution). The j in $\beta_{0[i|j]}$ indicates that each subject can deviate from the overall mean (often referred to as partial pooling or mixed effects modeling).

The goal is to infer differences in μ and σ of the assumed distribution of relative bacterial abundances between the subgroups (CC PRE, CC POST, HOME PRE and HOME POST), while accounting for the effects of the other environmental variables. This implies multiple comparisons of model parameters. In the Bayesian framework we can use a normal prior centered at 0 with a standard deviation of 1 for each comparison of interest across all the models so that the Bayesian 95% credible interval (CI) of the effect size is always more likely to include zero compared to the classical confidence interval, which makes Bayesian inference a more conservative approach⁴⁹. The robust linear models were fitted using the package *brms*, which uses the probabilistic programming language Stan^{50,51}. Stan utilizes Hamilton Monte Carlo (HMC), a Markov chain Monte Carlo (MCMC) method, to estimate parameters. To ensure proper convergence of the chains, we investigated individual chains using Shinystan and screened diagnostic parameters (divergent transitions and rhat values)⁵². To visualize the posterior predictive

intervals we utilized the tidybayes package⁵³. Brms allows multiple data sets as input, which enables the use of multiple imputation. We used predictive mean matching (PMM) as implemented in the *mice* package⁵⁴ to impute the missing data points within the breastfeeding variable. PMM is robust against misspecification of the imputation model while performing as well or better than common parametric imputation models under various missingness conditions with up to 20–30% missing values^{55,56}. In contrast, the variable breastfeeding had 10% missing. The multiple imputation resulted in 10 new datasets that were used to fit the Bayesian models. We also explored the same models using listwise deletion.

Finally, to evaluate whether we could predict childcare entrance based on microbiota composition after one month, we used the Random Forests machine learning algorithm (RF) with relative abundance data. RF is a tree-based ensemble learning method that is well suited for classification based on microbial abundances⁵⁷. We performed repeated 10-fold cross-validation (10 × 10) to estimate the classification accuracy using the caret package⁵⁸.

Results

Infant microbiota composition and dynamics. Overall, the infant microbiota was dominated by only a few genera from four Phyla; Actinobacteria, Firmicutes (specifically members from the Class Bacilli), Bacteroidetes and Proteobacteria. At the genus level, these were *Bifidobacterium* spp. with a mean relative abundance of 51%, followed by facultative anaerobes from the Bacilli (bacteria related to *Streptococcus* (16.8%), *Enterococcus* (3.4%), *Lactobacillus plantarum* et rel (2.9%) and *Granulicatella* (1.0%)). The cumulative mean relative abundance of these groups was more than 75%. Additionally, the variation in relative abundance of these taxa at the population level was very high. For instance, the relative abundance of *Bifidobacterium* spp. ranged from completely dominating (89%) to almost absent (0.2%) (Fig. 1A,B). Apart from being the most dominant taxa at the population level these were also among the most variable within subjects as determined by their CoV (Fig. 1C). The fact that the most abundant taxa display the highest variability implies a highly variable microbiota.

Effects of childcare entry on infant microbiota. *Redundancy analysis of the effects of childcare entry on infant gut microbiota composition.* We determined whether childcare (CC) entry affected overall microbiota community composition using Redundancy Analysis (RDA). RDA is a direct gradient analysis technique which summarizes linear relationships between components of response variables (microbiota) explained by a set of explanatory variables (CC and covariates) by multiple linear regression of the multiple response variables on the multiple explanatory variables. To determine how the different environmental variables interact with and impact the microbiota, we calculated their simple effects (i.e. the effect of the environmental variable on the microbiota without any other covariates) as well as the conditional effects (the impact on the microbiota when the effect of the other variables are partialled out). This allowed us to determine the effect of each variable on its own, but also their combined effects.

We did not find a significant effect of CC entry or the number of half-days in CC (Supplementary Table 1), compared to staying at home, on the microbiota. Neither in separation nor combined with other environmental variables. Nevertheless, birth mode, feeding mode, age, and siblings, were significantly correlated to the microbiota in concordance with literature³⁸. The strongest effect was from breastfeeding, with decreasing effect sizes for birth mode, siblings and age. All simple and conditional effects and corresponding p-values and their respective effect sizes (R^2) are shown in Table 2. R^2 reflects the percentage of variation explained out of the total microbiota variation; i.e. a higher R^2 implies a stronger effect. All these findings are combined in a tri-plot visualizing the relation of the environmental variables with each other and their resulting effect on the microbiota (Fig. 2A). The relation between the variation explained by the environmental variables and their overlapping conditional effects is visualized in a Venn diagram (Fig. 2B).

Bayesian group comparisons of effects of entry on individual microbial groups and microbiota diversity. To gain more insight on the association of CC entry and other environmental variables with individual bacterial groups and microbiota diversity, we performed Bayesian hierarchical robust linear models. Bayesian approaches provide more detailed information about the uncertainty when estimating parameters such as group differences or slope parameters in linear models. The robust linear model as described by Kruschke *et al.* is particularly well suited to model distributions when outliers are present and to address common model violations in standard linear models such as heterogeneity of variance⁴⁸. In the following, for covariates, *effect* refers to the magnitude of the slopes whereas for the group comparisons *effect* refers to the difference in the means. To compute the difference in means between two groups (e.g. CC-PRE – CC-POST) the calculated posterior distributions of their means are subtracted. We make a statement with confidence about the effect size being larger than zero when 95% of the posterior distribution excludes zero. The use of listwise deletion instead of multiple imputation led to similar results.

Figure 3 shows the posterior distributions of interest whereby red coloring indicates that we can make a claim with confidence. Within the CC or HOME group (Fig. 3A), temporal effects were as follows. Within CC, bacteria related to *Streptococcus bovis* and *Staphylococcus* were lower after one month, whereas within HOME bacteria related to *Granulicatella* and *Aerococcus* were higher, while those related to *Enterobacter aerogenes* and *Oxalobacter formigenes* were lower. We did not detect differences in relative bacterial abundances between the two groups (Fig. 3B, right graph) before CC entry, while after one month the relative abundance of the Proteobacteria related to *Enterobacter aerogenes* and *Klebsiella pneumoniae* was lower in the HOME group, while that of *Streptococcus intermedius* was higher (Fig. 3B, left graph).

Figure 3C shows that infants born via C-section showed higher relative abundances of bacteria related to *Granulicatella*, *Aerococcus* and Micrococaceae, but lower relative abundances of *Bifidobacterium*. Infants with siblings were found to have lower levels of *Staphylococcus* and also a lower SD of this taxon. A lower SD was also

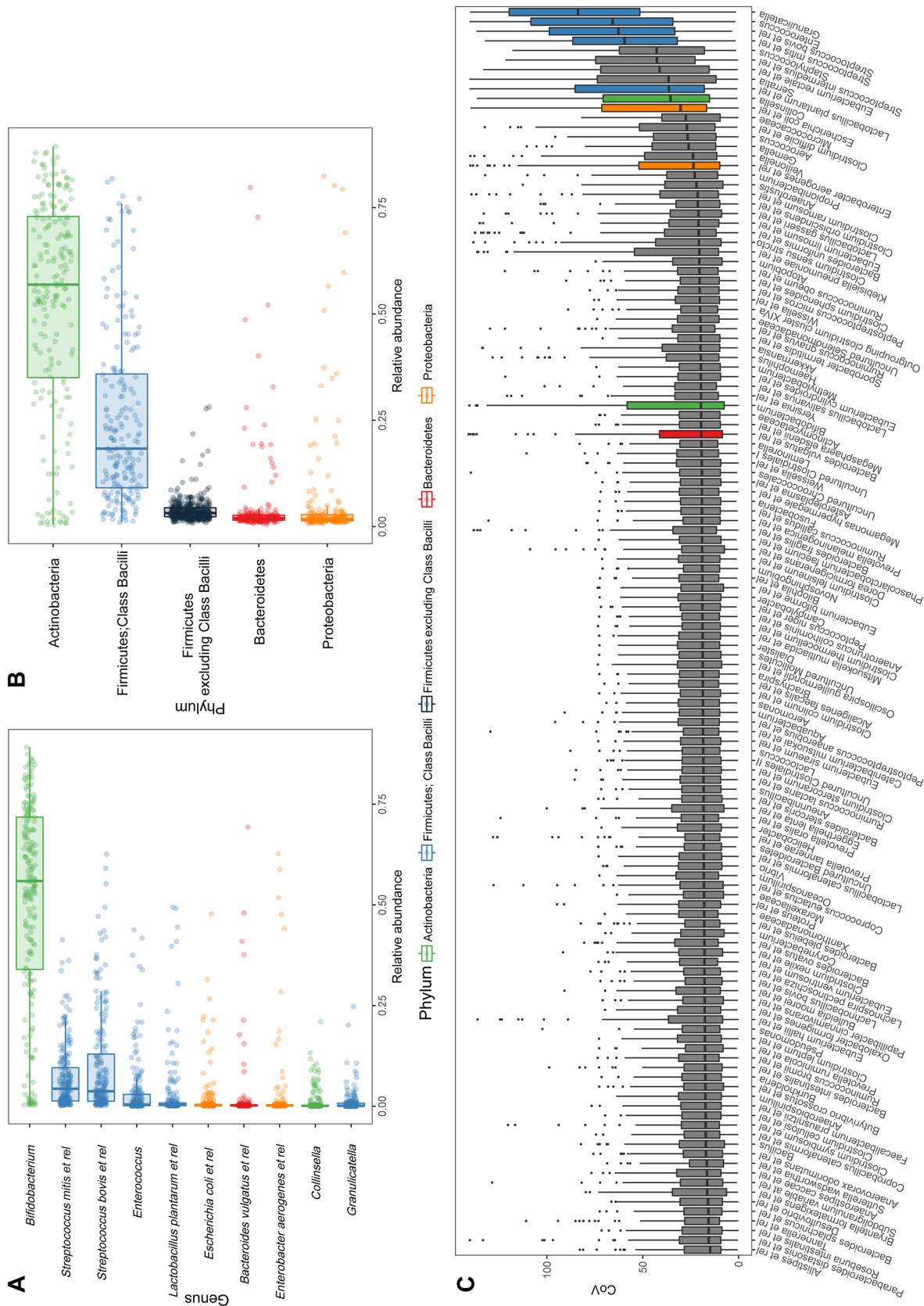


Figure 1. Infant microbiota composition and dynamics. **(A,B)** Relative abundance of major genus-like groups and their Phyla with mean abundance of >1% at the population level. **(C)** Coefficient of Variation (CoV) of all genus-like groups. The infant microbiota is highly variable as it is dominated by only a few taxa, which are also among the most variable. GREEN – Actinobacteria; BLUE – Firmicutes, Class Bacilli; BLACK – Firmicutes excluding Class Bacilli; RED – Bacteroidetes; ORANGE – Proteobacteria.

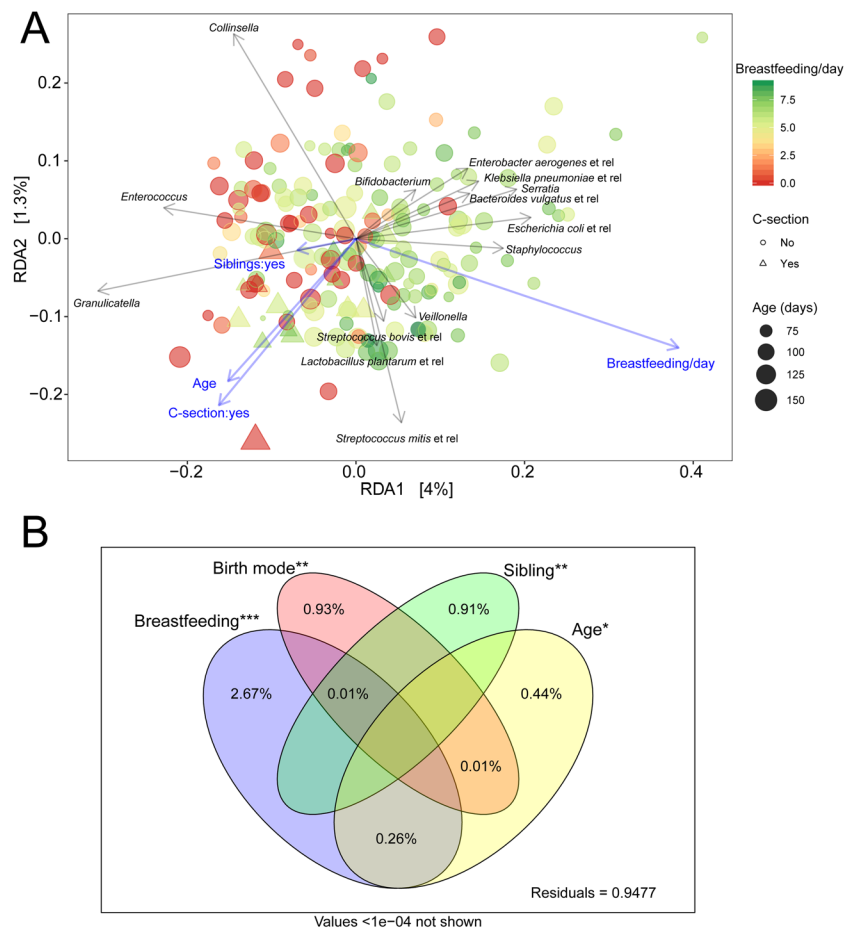


Figure 2. The impact of birth mode, breastfeeding, siblings and age on the infant gut microbiota. **(A)** Redundancy analysis (RDA) visualizing microbiota composition of all fecal samples ($n = 196$) colored by the number of breast-feedings and the size of the points scaled by age in days. Individuals born by C-section are represented as triangles. RDA displays and explains the variation explained in the microbiota, constrained by the predictor variables. Blue arrows depict the significant environmental variables and grey arrows the abundance of bacterial groups. Length of the arrows is a measure of fit. The longer the arrow the higher the association. **(B)** Venn diagram visualizing the partitioning of the variation explained by the significant predictors. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

	Df	Variance	F	Pr (>F)	R ²
Simple effects					
Time	1	0.317	1.4265	0.129	0.008
CC	1	0.274	1.2314	0.256	0.007
Age	1	0.436	1.9677	0.028*	0.011
Sibling	1	0.456	2.0602	0.019*	0.011
Birth-mode	1	0.559	2.5288	0.015**	0.014
Breastfeeding	2	1.605	3.7085	0.001***	0.032
CC × Time	3	0.781	1.172	0.226	0.019
Conditional effects					
Age	1	0.436	2.0652	0.018*	0.004
Sibling	1	0.598	2.8338	0.003**	0.010
Birth-mode	1	0.609	2.8843	0.002**	0.010
Breastfeeding	2	1.663	3.9364	0.001***	0.027

Table 2. RDA models output.

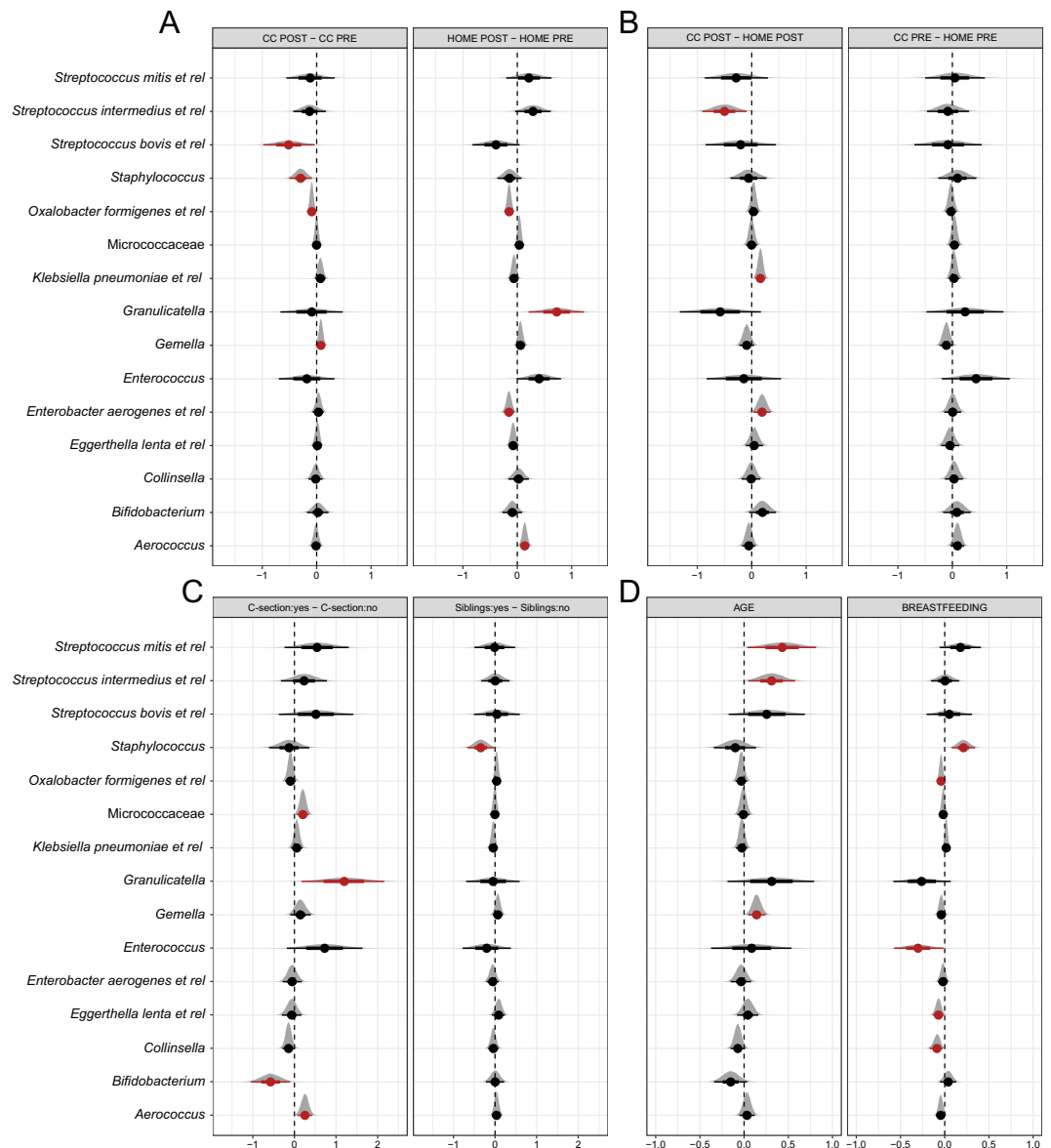


Figure 3. Bayesian hierarchical robust linear model posterior distributions for individual bacterial group differences. (A) within group effects, (B) between group effects and (C) C-section and siblings (D) age and breastfeeding. For the covariates (D) the x-axis refers to the magnitude of the slopes whereas for the group comparisons (A–C) it refers to the difference in the means. Taxa are shown in red when the probability that the absolute effect size is >0 exceeds 0.95, given our model and the data.

confirmed for *Enterococcus*, without differences in the mean. Finally, as can be seen in Fig. 3D, a higher number of daily breast-feedings was associated with more bacteria related to *Staphylococcus* and less to *Enterococcus*, *Collinsella*, *Eggerthella lenta* and *Oxalobacter formigenes*. A higher age was positively associated with bacteria related to *Streptococcus mitis*, *Streptococcus intermedius* and *Gemella*.

We used the same approach to compare microbiota diversity between groups. Table 3 shows the estimated difference in the means and standard deviations between the groups as well as the magnitude of the slopes for the covariates. Within the CC or HOME groups our model estimates that there is no temporal effect on diversity as well as between the two groups before CC entrance. However, when comparing HOME and CC one month after entrance, the average diversity was estimated to be lower in the CC group. Figure 4 shows the means of the Shannon diversity index per group with 95% CI (black point range) as well as the observed values (black points) and the posterior predictive interval (blue bar). To calculate the predictive interval, we used the median for the average number of breast-feedings and the median age at PRE and POST, respectively. The Bayesian predictive intervals illustrate the uncertainty of the predictions: the estimated distributions for the CC/HOME groups are very much overlapping despite the mean difference, whereas C-Section seems to have a stronger impact on diversity. Age and breastfeeding did not reach our predefined threshold to make a statement with confidence about the effect being >0 and there was no difference in the estimated standard deviations between groups.

Comparison	Median	95% CI
Difference in means		
CC PRE - HOME PRE	0.01	[-0.18, 0.19]
HOME POST - HOME PRE	0.11	[-0.01, 0.23]
CC POST - CC PRE	-0.12	[-0.26, 0.03]
CC POST - HOME POST	-0.22	[-0.41, -0.03]
Covariates		
C-section:yes - C-section:no	0.43	[0.15, 0.7]
Sibling:yes - Sibling:no	-0.11	[-0.27, 0.06]
Age	0.06	[-0.07, 0.2]
Breastfeeding	-0.06	[-0.13, 0.01]
Difference in SD		
CC PRE - HOME PRE	0.07	[-0.02, 0.17]
HOME POST - HOME PRE	0.01	[-0.09, 0.1]
CC POST - CC PRE	0.02	[-0.13, 0.19]
CC POST - HOME POST	0.08	[-0.04, 0.22]
C-section:yes - C-section:no	0.06	[-0.06, 0.24]
Sibling:yes - Sibling:no	-0.02	[-0.12, 0.07]

Table 3. Estimated model parameters for microbiota diversity (Shannon).

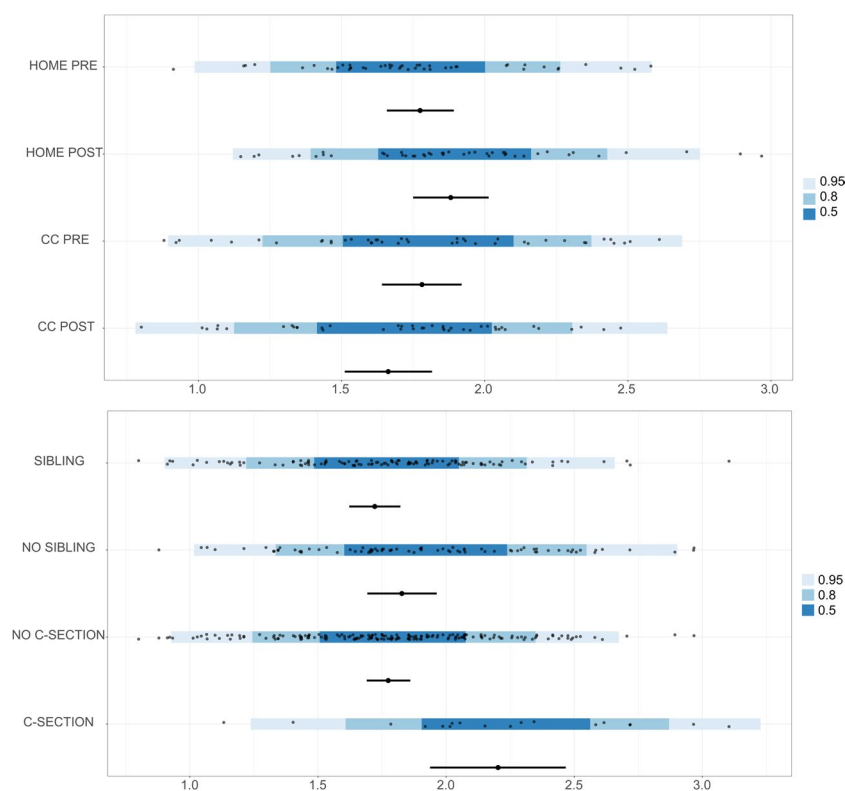


Figure 4. Bayesian hierarchical robust linear model group posteriors for microbiota alpha diversity. Means of Shannon diversity per group with 95% CI are shown with the black point range. The observed values are shown as black points within the blue bar. The Bayesian posterior predictive intervals are shown in blue. The predictions were made using the median for average number of breast-feedings and median age at PRE and POST.

Random Forest analysis of non-linear relationships between childcare entry and microbiota composition. Finally, we used the random forests (RF) algorithm to determine if we could accurately classify whether an infant belonged to the HOME or CC group after one month. The latter would indicate a characteristic effect of CC on the infant gut microbiota. The benefit RF has over the linear models is its ability to detect non-linear associations. The prediction accuracy using repeated cross validation was close to random classification with 53.5%, suggesting that CC did not produce a strong uniform shift in microbiota composition.

Discussion

This study examined the effect of entry into center-based childcare (CC) on gut microbial composition of 3–4-month-old infants by comparing the microbial composition between infants that entered CC at 3 months and infants that were cared for at home (HOME) at that age. For all infants, we assessed microbial composition at two time points: At 10 weeks of age (PRE), which was before CC entrance for the CC group, and 4 weeks later, or 4 weeks after entrance for the CC group (POST). We accounted for known covariates including age, presence of siblings (yes vs no), delivery mode (natural vs C-section) and the average number of breast-feedings per day in the period before measurement of the microbiota. We combined multivariate (Redundancy analysis and Random Forest algorithm) with Bayesian univariate statistical methods to test our hypothesis that CC entrance is associated with changes in microbial composition over time.

The microbiota of all the infants exhibited a low microbial diversity as it was dominated by only a few typical bacterial groups from the phylum Actinobacteria (*Bifidobacterium* spp and *Collinsella*), facultative anaerobes from the Firmicutes (such as *Streptococcus* spp, *Lactobacillus* spp and *Enterococcus* spp), Proteobacteria (*E. coli* and bacteria related to *Enterobacter aerogenes*) and Bacteroidetes (*Bacteroides* spp). These findings are in line with findings from previous studies in the US and Europe (Germany, Finland, Sweden and the Netherlands)^{38,59,60}.

In contrast to breastfeeding, birth mode, age, and the presence of siblings, CC was not associated with gut microbiota composition according to Redundancy analysis (RDA). In line with that, we could not achieve higher accuracy than by chance using the Random Forest algorithm to classify CC vs HOME using the POST samples. Bayesian univariate analyses, that also take the individuality of the starting microbiota into account, did show a few taxa to be differently distributed between the two groups in the POST samples. In the HOME group the relative abundances of Proteobacteria related to *Enterobacter aerogenes* and *Klebsiella pneumoniae* were lower while that of *Streptococcus intermedius* was higher compared to the CC group at time POST. Except for bacteria related to *Streptococcus intermedius*, these bacterial groups were also heavily influenced by other environmental variables. Using the same Bayesian approach, we observed a lower Shannon alpha diversity in the CC group compared to the HOME group at time POST. All in all, our results show that entrance to CC does not result in a complete and homogeneous ‘disruption’ or ‘dysbiosis’ of the microbiota as reported for rodents subjected to early life stress.

There are several possible explanations for the lack of a general effect of the entrance to childcare on infants’ gut microbiota. First, although CC entrance may be considered a major stressor in a young infant’s life that is accompanied by substantial rises in stress hormones^{33–36}, it might not be comparable in severity to maternal separations in rodents. E.g. in our CC sample infants were taken care of by surrogate caregivers, while in maternal separation paradigms with rodent pups, there is no surrogate caregiver during the separation²⁷. Also, Dutch infants typically attend childcare for around 2 days a week⁶¹; this contrasts with practices elsewhere in the world and may not be enough exposure to produce major effects on the microbiota. Second, rodents lack the genotypical variation found in human populations and are kept and studied in controlled laboratory environments in which the individual variation in gut microbiota is small^{62,63}. Hence, in rodent models the effects of environmental pressures can be studied in complete isolation. This may lead to a much greater response to an individual stressor than it normally would when other (stronger) environmental drivers of microbiota composition are present. Finally, the infant gut microbiota at 3–4 months of age is still in an unstable, dynamic, highly individual developmental stage^{17,24–26,38}. This was confirmed by a very large intra and inter-individual microbiota variability in our population. Microbiota at these young ages appears to show large fluctuations, probably resulting from a myriad of environmental influences. Additionally, infants of the present study went to different childcare centers, which exposed them to different built environments, caregivers and other infants with different microbiomes, thereby increasing the variety of environmental influences on the microbiota. Nevertheless, a generalizable stress-related effect across centers would most likely have been detectable, despite these (unknown) and potential center-specific confounding variables. Other factors than those we controlled for are e.g. fever, contact with animals, and/or genetics^{26,38}. The instability of the infant microbiota might also possibly be an intrinsic property of the dynamics of the gut microbial colonization and may obscure the effects of individual environmental factors that may each have only a modest influence on gut microbiota composition. In other words, an environmental factor would need to be very strong to generate a universal disruption in the microbial composition that overrides the normally occurring fluctuations in infant gut microbiota in the first months of life.

The results of our study showed significant associations with known important environmental factors. Their effects were often partly overlapping and impacted similar bacterial groups (Fig. 2A,B), possibly indicating an accelerated colonization process in infants born by C-section. For instance, being older and being born by C-section were both positively associated with the abundance of the highly variable facultative anaerobes from the Bacilli and related to *S. bovis* and *S. mitis* L. *plantarum* and *Granulicatella* at the expense of *Bifidobacterium* and several Proteobacteria. Contrarily, a younger age and being born vaginally were both associated with higher relative abundances of *Bifidobacterium*, *Collinsella* and several Proteobacteria in line with previous research of (Dutch) infants^{59,64}. The latter pattern, except for *Collinsella*, was also associated with having no siblings. C-section delivery was associated with higher Shannon diversity. This increase in diversity may be the result of the lower proportions of the generally dominant *Bifidobacterium* in this group of infants, as has been previously reported⁶⁵.

Breastfeeding showed the strongest association with infant gut microbiota composition. However, surprisingly, breastfeeding was only weakly related to *Bifidobacterium*, but was rather more associated with increases in mainly *Staphylococcus* and to a lesser extent, Proteobacteria (*Serratia*, *E. coli*, *Klebsiella pneumoniae* and *Enterobacter aerogenes* et rel) and *Bacteroides vulgatus* et rel. Breastfeeding was also strongly associated with a decrease in *Enterococcus* and *Collinsella*. In recent years in European countries, infant formulae have often been supplemented with prebiotics such as short chain galacto-oligosaccharides (scGOS) alone, or in a mixture with a chicory root derived inulin containing long chain fructo-oligosaccharides (lFOS)⁶⁶. Prebiotics mimic the bifidogenic effect of oligosaccharides found in human milk thus preventing the difference in *Bifidobacterium* abundance that was previously associated with formula feeding⁶⁰. Furthermore, human milk itself has been found to

contain bacteria, including Proteobacteria, as well as *Staphylococcus*, which is generally associated with the skin⁶⁷. The latter has also been previously found to be depleted in formula-fed infants in concordance with our earlier findings⁶⁰. In total, the environmental variables explained a non-redundant ~5.9% of the microbiota composition. This is in the same order of magnitude as previously reported in other human studies⁶⁸. This means that in infants generally only a relatively small proportion of the gut microbial composition can be explained by the factors that are most commonly accounted for.

Despite the fact that we did not find an effect of CC entrance on the microbiota at the group level even after using different statistical approaches, it is not possible to conclude that there was no disrupting effect on the microbiota at the individual level. Individual infants did show large changes in the microbiota between the two assessments, but these changes were not uniform across individuals. Although temporal microbial dynamics at the population level, with regards to bacterial succession patterns, have been shown to be universal over different cultures and geography⁵⁹ it is known that this process is very variable between individuals^{17,24–26}. Future studies including larger populations and especially repeated measurements from before and after life changing events are necessary to determine whether the resulting bacterial shift is typical or deviant for a specific individual. Subsequently, potential sub-groups of infants displaying specific signatures of dysbiosis can be determined, and their microbial signatures related to possible adverse health outcomes later in life. However, given the importance of gut microbial colonization for the development of the microbiota-gut-brain axis and the future health of the individual¹¹, it is also possible that the colonization process might be robust against perturbations such as entrance to CC, hence explaining the lack of microbial change at the group level.

This is the first human study that examined the effect of entrance to center-based childcare on microbiome development. The strengths are that it included a relatively large population of healthy infants with a natural variation of environmental variables and an ecologically valid stressor as opposed to typical rodent studies that generally include low number of individuals under strictly controlled laboratory conditions. Another strength is that we used a combination of different multivariate and Bayesian univariate statistical methods to analyze the data. However, there are also some weaknesses with regards to this study. Our population was ethnically and socio-economically uniform⁶⁹. This can be seen as an advantage, but it precludes generalization to the larger population. Also, to conclude whether the microbiota is disrupted at an individual level, more sample time points than just PRE and POST would be needed. For example, it would be interesting to follow infants entering childcare for a longer period of time, as elevated infant cortisol has been observed throughout the first months in childcare³⁴. Larger and more stable differences in the gut microbiota of infants attending childcare and those being taken care of at home, may only appear after a longer period of time in childcare.

Conclusion

Entering center-based childcare has been shown to produce large increases in stress hormones in infants and can therefore be considered a significant stressor in early life. Childcare includes maternal separation and in animal models early life maternal separation has been found to lead to large shifts in gut microbial composition. However, in the present study infant gut microbiota was not impacted in a uniform way by entering childcare at the age of 3 months. Large shifts in gut microbiota were observed, but were idiosyncratic to individual infants and were also observed in infants not attending center-based childcare. In general, the infants' gut microbiota was found to be intrinsically very dynamic. Other environmental variables, namely breastfeeding, birth mode, age, and the presence of siblings, were shown to significantly impact the microbial composition, with effects that were largely overlapping and typically included the most abundant and variable taxa. Our results suggest that in infants the stress-inducing effects of childcare entry might not be as strong as the maternal separation paradigms of animal models. Alternatively, general effects may potentially only become visible after longer periods of childcare entry and when infants are older and their gut microbiota has become more stable.

Received: 2 December 2019; Accepted: 13 May 2020;

Published online: 24 June 2020

References

1. Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* <https://doi.org/10.1136/gutjnl-2015-309990> (2015).
2. Cryan, J. F. *et al.* The Microbiota-Gut-Brain Axis. *Physiol Rev* **99**, 1877–2013, <https://doi.org/10.1152/physrev.00018.2018> (2019).
3. Adlerberth, I. & Wold, A. E. Establishment of the gut microbiota in Western infants. *Acta Paediatrica* **98**, 229–238, <https://doi.org/10.1111/j.1651-2227.2008.01060.x> (2009).
4. Gonzalez, A. *et al.* The mind-body-microbial continuum. *Dialogues Clin Neurosci* **13**, 55–62 (2011).
5. Leslie, J. L., Vendrov, K. C., Jenior, M. L. & Young, V. B. The Gut Microbiota Is Associated with Clearance of *Clostridium difficile* Infection Independent of Adaptive Immunity. *MSphere* **4**, <https://doi.org/10.1128/mSphereDirect.00698-18> (2019).
6. de Weerth, C. Do bacteria shape our development? Crosstalk between intestinal microbiota and HPA axis. *Neurosci Biobehav R* **83**, 458–471, <https://doi.org/10.1016/j.neubiorev.2017.09.016> (2017).
7. Mayer, E. A. Gut feelings: the emerging biology of gut-brain communication. *Nat Rev Neurosci* **12**, 453–466, <https://doi.org/10.1038/nrn3071> (2011).
8. Derrien, M., Alvarez, A. S. & de Vos, W. M. The Gut Microbiota in the First Decade of Life. *Trends Microbiol*, <https://doi.org/10.1016/j.tim.2019.08.001> (2019).
9. Robertson, R. C., Manges, A. R., Finlay, B. B. & Prendergast, A. J. The Human Microbiome and Child Growth - First 1000 Days and Beyond. *Trends in Microbiology* **27**, 131–147, <https://doi.org/10.1016/j.tim.2018.09.008> (2019).
10. Rodriguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis* **26**, 26050, <https://doi.org/10.3402/mehd.v26.26050> (2015).
11. Borre, Y. E., Moloney, R. D., Clarke, G., Dinan, T. G. & Cryan, J. F. The impact of microbiota on brain and behavior: mechanisms & therapeutic potential. *Advances in experimental medicine and biology* **817**, 373–403, https://doi.org/10.1007/978-1-4939-0897-4_17 (2014).

12. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* **107**, 11971–11975, <https://doi.org/10.1073/pnas.1002601107> (2010).
13. Dominguez-Bello, M. G. *et al.* Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature medicine* **22**, 250–253, <https://doi.org/10.1038/nm.4039> (2016).
14. Korpela, K. & de Vos, W. M. Early life colonization of the human gut: microbes matter everywhere. *Current opinion in microbiology* **44**, 70–78, <https://doi.org/10.1016/j.mib.2018.06.003> (2018).
15. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nature medicine* **22**, 713–722, <https://doi.org/10.1038/nm.4142> (2016).
16. Sevelsted, A., Stokholm, J., Bonnelykke, K. & Bisgaard, H. Cesarean Section and Chronic Immune Disorders. *Obstet Gynecol Surv* **70**, 303–305, <https://doi.org/10.1097/01.ogx.0000466336.81671.9f> (2015).
17. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* **108**(Suppl 1), 4578–4585, <https://doi.org/10.1073/pnas.1000081107> (2011).
18. Sela, D. A. *et al.* The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **105**, 18964–18969, <https://doi.org/10.1073/pnas.0809584105> (2008).
19. Marcobal, A. *et al.* Consumption of human milk oligosaccharides by gut-related microbes. *Journal of agricultural and food chemistry* **58**, 5334–5340, <https://doi.org/10.1021/jf9044205> (2010).
20. Smith, M. I. *et al.* Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* **339**, 548–554, <https://doi.org/10.1126/science.1229000> (2013).
21. Blanton, L. V. *et al.* Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* **351**, <https://doi.org/10.1126/science.aad3311> (2016).
22. Subramanian, S. *et al.* Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**, 417–421, <https://doi.org/10.1038/nature13421> (2014).
23. Planer, J. D. *et al.* Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* **534**, 263–266, <https://doi.org/10.1038/nature17940> (2016).
24. Trosvik, P., Stenseth, N. C. & Rudi, K. Convergent temporal dynamics of the human infant gut microbiota. *ISME J* **4**, 151–158, <https://doi.org/10.1038/ismej.2009.96> (2010).
25. Favier, C. F., Vaughan, E. E., De Vos, W. M. & Akkermans, A. D. Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microbiol* **68**, 219–226 (2002).
26. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol* **5**, e177, <https://doi.org/10.1371/journal.pbio.0050177> (2007).
27. O'Mahony, S. M., Hyland, N. P., Dinan, T. G. & Cryan, J. F. Maternal separation as a model of brain-gut axis dysfunction. *Psychopharmacology* **214**, 71–88, <https://doi.org/10.1007/s00213-010-2010-9> (2011).
28. O'Mahony, S. M. *et al.* Early life stress alters behavior, immunity, and microbiota in rats: implications for irritable bowel syndrome and psychiatric illnesses. *Biological psychiatry* **65**, 263–267, <https://doi.org/10.1016/j.biopsych.2008.06.026> (2009).
29. Desbonnet, L. *et al.* Effects of the probiotic *Bifidobacterium infantis* in the maternal separation model of depression. *Neuroscience* **170**, 1179–1188, <https://doi.org/10.1016/j.neuroscience.2010.08.005> (2010).
30. Foster, J. A., Rinaman, L. & Cryan, J. F. Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiol Stress* **7**, 124–136, <https://doi.org/10.1016/j.ynstr.2017.03.001> (2017).
31. Bailey, M. T. & Coe, C. L. Maternal separation disrupts the integrity of the intestinal microflora in infant rhesus monkeys. *Developmental psychobiology* **35**, 146–155, doi: 10.1002/(Sici)1098-2302(199909)35:2<146::Aid-Dev7>3.0.Co;2-G (1999).
32. De Palma, G. *et al.* Microbiota and host determinants of behavioural phenotype in maternally separated mice. *Nature communications* **6**, 7735, <https://doi.org/10.1038/ncomms8735> (2015).
33. Ahnert, L., Gunnar, M. R., Lamb, M. E. & Barthel, M. Transition to child care: associations with infant-mother attachment, infant negative emotion, and cortisol elevations. *Child development* **75**, 639–650, <https://doi.org/10.1111/j.1467-8624.2004.00698.x> (2004).
34. Albers, E. M., Beijers, R., Riksen-Walraven, J. M., Sweep, F. C. & de Weerth, C. Cortisol levels of infants in center care across the first year of life: links with quality of care and infant temperament. *Stress* **19**, 8–17, <https://doi.org/10.3109/10253890.2015.1089230> (2016).
35. Watamura, S. E., Coe, C. L., Laudenslager, M. L. & Robertson, S. S. Child care setting affects salivary cortisol and antibody secretion in young children. *Psychoneuroendocrinology* **35**, 1156–1166, <https://doi.org/10.1016/j.psyneuen.2010.02.001> (2010).
36. Waynforth, D. The influence of parent-infant cosleeping, nursing, and childcare on cortisol and sIgA immunity in a sample of British children. *Developmental psychobiology* **49**, 640–648, <https://doi.org/10.1002/dev.20248> (2007).
37. Beijers, R., Jansen, J., Riksen-Walraven, M. & de Weerth, C. Nonparental care and infant health: do number of hours and number of concurrent arrangements matter? *Early human development* **87**, 9–15, <https://doi.org/10.1016/j.earlhumdev.2010.09.003> (2011).
38. Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588, <https://doi.org/10.1038/s41586-018-0617-x> (2018).
39. Salonen, A. *et al.* Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods* **81**, 127–134, <https://doi.org/10.1016/j.mimet.2010.02.007> (2010).
40. Claesson, M. J. *et al.* Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* **4**, e6669, <https://doi.org/10.1371/journal.pone.0006669> (2009).
41. Rajilic-Stojanovic, M. *et al.* Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* **11**, 1736–1751, <https://doi.org/10.1111/j.1462-2920.2009.01900.x> (2009).
42. Lahti, L., Elo, L. L., Aittokallio, T. & Kaski, S. Probabilistic analysis of probe reliability in differential gene expression studies with short oligonucleotide arrays. *IEEE/ACM transactions on computational biology and bioinformatics/IEEE, ACM* **8**, 217–225, <https://doi.org/10.1109/TCBB.2009.38> (2011).
43. Lahti, L., Torrente, A., Elo, L. L., Brazma, A. & Rung, J. A fully scalable online pre-processing algorithm for short oligonucleotide microarray atlases. *Nucleic Acids Res* **41**, e110, <https://doi.org/10.1093/nar/gkt229> (2013).
44. Team, R. C. R: A language and environment for statistical computing, <https://www.r-project.org/> (2018).
45. Aitchison, J. *The statistical analysis of compositional data*. (Chapman & Hall, Ltd., 1986).
46. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in microbiology* **8**, 2224, <https://doi.org/10.3389/fmicb.2017.02224> (2017).
47. Oksanen, J. F. *et al.* Eduard Soezocs, Helene Wagner. vegan: Community Ecology Package. R package version 2.5-2, <https://CRAN.R-project.org/package=vegan> (2018).
48. Kruschke, J. K. Bayesian estimation supersedes the t test. *Journal of experimental psychology. General* **142**, 573–603, <https://doi.org/10.1037/a0029146> (2013).
49. Gelman, A., Hill, J. & Yajima, M. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* **5**, 189–211, <https://doi.org/10.1080/19345747.2011.618213> (2012).
50. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *2017* **80**, 28, <https://doi.org/10.18637/jss.v080.i01> (2017).
51. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *2017* **76**, 32, <https://doi.org/10.18637/jss.v076.i01> (2017).

52. Gabry, J. shinystan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models. R package version 2.5.0, <https://CRAN.R-project.org/package=shinystan> (2018).
53. Kay, M. Tidybayes: Tidy Data and Geoms for Bayesian Models (Version v1.1.0). *Zenodo*, <https://doi.org/10.5281/zenodo.1308151> (2018).
54. van Buuren, S. & Groothuis-Oud, K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45**, 1–67 (2011).
55. Marshall, A., Altman, D. G. & Holder, R. L. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol* **10**, 112, <https://doi.org/10.1186/1471-2288-10-112> (2010).
56. Kleinke, K. Multiple Imputation Under Violated Distributional Assumptions: A Systematic Evaluation of the Assumed Robustness of Predictive Mean Matching. *J Educ Behav Stat* **42**, 371–404, <https://doi.org/10.3102/1076998616687084> (2017).
57. Knights, D., Costello, E. K. & Knight, R. Supervised classification of human microbiota. *FEMS Microbiol Rev* **35**, 343–359, <https://doi.org/10.1111/j.1574-6976.2010.00251.x> (2011).
58. Max Kuhn. Contributions from Jed Wing, S. W., Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan, Tyler Hunt. caret: Classification and Regression Training. R package version 6.0-79, <https://CRAN.R-project.org/package=caret> (2018).
59. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227, <https://doi.org/10.1038/nature11053> (2012).
60. Borewicz, K. *et al.* The effect of prebiotic fortified infant formulas on microbiota composition and dynamics in early life. *Scientific reports* **9**, 2434, <https://doi.org/10.1038/s41598-018-38268-x> (2019).
61. Beijers, R., Riksen-Walraven, M., Putnam, S., de Jong, M. & de Weerth, C. Early non-parental care and toddler behaviour problems: Links with temperamental negative affectivity and inhibitory control. *Early Child Res Q* **28**, 714–722, <https://doi.org/10.1016/j.ecresq.2013.06.002> (2013).
62. Hugenholtz, F. Z. J., o'Toole, P. W., Marchesi, J. & Smidt H. In *Manual of Environmental Microbiology* (ed Nakatsu CH Yates MV, Miller RV, Pillai SD) 4.4.2-1 – 10 (ASM Press, 2016).
63. Nguyen, T. L., Vieira-Silva, S., Liston, A. & Raes, J. How informative is the mouse for human gut microbiota research? *Disease models & mechanisms* **8**, 1–16, <https://doi.org/10.1242/dmm.017400> (2015).
64. Penders, J. *et al.* Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**, 511–521, <https://doi.org/10.1542/peds.2005-2824> (2006).
65. Jakobsson, H. E. *et al.* Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* **63**, 559–566, <https://doi.org/10.1136/gutjnl-2012-303249> (2014).
66. EFSA Panel on Dietetic Products, N. a. A. Scientific Opinion on the essential composition of infant and follow-on formulae. *EFSA Journal* **12**, <https://doi.org/10.2903/j.efsa.2014.3760> (2014).
67. Ward, T. L., Hosid, S., Ioshikhes, I. & Altosaar, I. Human milk metagenome: a functional capacity analysis. *BMC Microbiol* **13**, 116, <https://doi.org/10.1186/1471-2180-13-116> (2013).
68. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564, <https://doi.org/10.1126/science.aad3503> (2016).
69. Beijers, R., Jansen, J., Riksen-Walraven, M. & de Weerth, C. Maternal prenatal anxiety and stress predict infant illnesses and health complaints. *Pediatrics* **126**, e401–409, <https://doi.org/10.1542/peds.2009-3226> (2010).

Acknowledgements

This research was supported by the Netherlands Organization for Scientific Research VIDI grant (575-25-009, to C de Weerth), a Jacobs Foundation Advanced Research Fellowship (to C de Weerth), a Netherlands Organization for Scientific Research VICI grant (016.Vici.185.038, to C de Weerth) and by the Spinoza Award of the Netherlands Organization for Scientific Research to WM de Vos.

Author contributions

G.H. performed the literature search, statistical analysis and wrote the manuscript. H.E. performed the literature search, statistical analysis and wrote the manuscript. C.W. performed the literature search, designed the study and collected the data. W.V. contributed to the design and application of the HITChip analysis of the microbiota. All authors contributed to the writing of the manuscript, critically reviewed and revised the manuscript, and approved the final manuscript as submitted.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66404-z>.

Correspondence and requests for materials should be addressed to C.d.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020