American Journal of Student Research

Original Research Article

A Quantitative Seeding Score Analysis of Early-Life Microbiome Colonization in Infant Feeding Modes

Shion Choe

Winchester High School, 80 Highland Ave, Winchester, MA 01890, United States

ABSTRACT

The early life microbiome of an individual is a critical factor for long term health conditions, yet the influence of various feeding modes on microbial transmission patterns remains as a complex area for study. This study presents a secondary analysis of publicly available infant microbiome data, while proposing a novel quantitative 'microbial seeding score' to evaluate the direction and magnitude of early-life microbiome colonization across feeding modes. Using genus-level microbiome data from a total of 67 infants obtained from HE_INFANTS_MFGM_2019 dataset, a microbial seeding score (S) was calculated to examine maternally associated taxa (Bacteroides, Bifidobacterium) with environmentally associated taxa (Enterococcus, Staphylococcus, Streptococcus, Klebsiella). Infants were grouped into breast milk (n=16), experimental infant formula (n=25), and standard infant formula (n=26) by feeding mode. Breast milk-fed infants indicated the highest mean seeding score ($S = 0.391 \pm 0.319$), and experimental formula-fed infants showed the intermediate mean seeding score ($S = 0.352 \pm 0.352$). Standard formula-fed infants exhibited the lowest mean seeding score ($S= 0.301 \pm 0.309$). Differences were not statistically significant (one-way ANOVA, $p=0.676$). However, there was a consistent decreasing trend in seeding scores observed across feeding modes. The main contribution of this work is methodological for introducing a simplified quantitative metric instead of reporting new empirical microbiome measurements. These findings suggest that the proposed seeding score in this study may capture exploratory patterns of variation in early microbial colonization, although differences were not statistically significant.

Keywords: Infant microbiome; microbial seeding; feeding mode; breastfeeding; quantitative modeling; early-life colonization

INTRODUCTION

The human body is a complex, co-evolving ecosystem inhabited by a vast microbiome, a diverse community of microorganisms including bacteria, viruses, eukaryotes and prokaryotes. Recent scientific definitions define the human microbiome as a collection of genomes of all microorganisms that inhabit the human body, serving as a role of an integrated biological system instead of isolated taxa (1). Depending on anatomical sites, these microbial communities vary, and they play a role in metabolic, immunological, and developmental processes that are needed for human health (2). While every organism develops from a single sterile cell, their microbial colonization is critical as they play essential roles in maintaining host homeostasis (3).

The transfer of these microbes to develop into an individual's biota is classified into phased steps. Vertical transmission occurs before birth, where microorganisms enter the fetus through the mother's blood, the placenta,

---

Corresponding author: Shion Choe, E-mail: shionchoe84@gmail.com.

Accepted February 12, 2026
https://doi.org/10.70251/HYJR2348.41731735

and amniotic fluids surrounding the fetus (4). After birth, the expansion of the microbiome results from horizontal transmission. Environmental exposures due to contact with microbe-laden materials, diet, medication, behavior of children, pets, and numerous other possibilities significantly influences the newborns' microbiome (5).

Early microbial exposure and biota development are influenced by multiple factors, such as delivery context and early feeding. Following the two microbial transmissions, the birth canal is another major seeding event. During vaginal birth, the newborn's skin comes in contact with the mother's vaginal and gut microbiota. This exposure is essential for establishing initial, and balanced microbial diversity throughout multiple body sites. However, infants born through the cesarean section (C-section) are surgically delivered and skip passing through the birth canal. Due to this, they experience side effects, lacking necessary organisms and instead resemble the initial microbial composition of the hospital environment and the mother's skin (6). This altered seeding shows subsequent differences in the development of the infant's microbiota.

Although many prior studies have described differences in infant microbiome composition by delivery mode and feeding type, they often relied on complex multivariate or compositional approaches. Those approaches may be difficult to interpret in the perspective of transmission source. A simple and transparent metric directly contrasting maternally associated versus environmentally associated taxa has not been specifically established. Direct mechanism of maternal microbial transmission is provided by breastfeeding through contact and milk-derived bacteria. However, formula feeding may increase reliance on environmental exposures. Therefore, a practical and biologically relevant test case is represented by feeding mode for assessing whether a simplified metric may distinguish patterns of early microbial seeding. Human milk consists of not only balanced nutrition, but also contains essential components. In addition, breast milk delivers immune factors and symbiotic bacteria, directly contributing to the diversification of early life gut microbiome (7). Microbial and nutritional exposure from delivery mode and breastfeeding are both powerful mechanisms which the maternal, and early environment shapes the infant's internal ecosystem. These factors are hypothesized to influence the biological composition and abundance of colonizing microbes. This study conducts a secondary analysis of an existing infant microbiome dataset, introducing and evaluating a novel quantitative metric as the microbial seeding score that was specifically designed to contrast maternally correlated and environmentally associated microbial taxa. This study seeks to answer the research question about whether early feeding mode (breastfeeding versus formula feeding) influences the direction and magnitude of early-life microbiome seeding in infants. This study hypothesized that infants who are breastfed will show higher seeding scores than infants fed standard or experimental infant formula.

While prior studies have described differences in infant microbiome composition quantitatively by delivery mode and feeding type, only a few studies have suggested a simplified quantitative model that directly contrasted maternal and environmental microbial transmission. This study addresses this gap through a seeding score that mathematically captures the magnitude and direction of early-life microbial colonization.

METHODS AND MATERIALS

Data Source

The study is derived from the HE_INFANTS_MFGM_2019 dataset from the Bornstein Lab's curated collection of microbiome and metabolome, and all subjects included infants sampled at a two-month baseline after birth. Data consisted of two main variables; genus-level bacterial abundances and the infant metadata, including their feeding type. The two variables were used to demonstrate how the microbial "seeding score" was affected by bacterial abundance and feeding type, and show the difference between maternal and environmental microbial colonization patterns.

Microbial Seeding Score Construction

The seeding score (S) was calculated by the relative abundance (%) of key bacteria types of each infant (Bacteroides, Bifidobacterium, Enterococcus, Staphylococcus, Streptococcus, Klebsiella). Bacteroides and Bifidobacterium were chosen as maternally associated taxa because of their documented enrichment in breastfed infants and also their reported roles in early-life gut colonization. These taxa have been specifically repeatedly to be enriched in breastfed infants and correlated with maternal microbial transmission through breast milk and maternal contact. However, Enterococcus, Staphylococcus, Streptococcus, and Klebsiella were frequently reported in hospital-related environments and formula-fed infant microbiomes. Therefore, they were useful indicators of environmental exposure. However, this binary classification was not intended to imply exclusive transmission pathways since some of these genera may also be maternally derived. It rather functions as a heuristic framework to contrast dominant transmission pattern in early-life microbial colonization.

S = (Bacteroides + Bifidobacterium) – (Enterococcus + Staphylococcus + Streptococcus + Klebsiella)

All bacterial abundances were indicated as relative abundance percentages. This ensured comparability across samples, while preventing bias from differences in sequencing depth.

A higher seeding score (S) indicated stronger maternal seeding, and lower seeding score (S) indicated greater possibility of environmental influence (seeding). Infants were categorized into feeding groups into 3 different diets, regular breast milk, standard infant formula (Manufactured nutritional substitute for human milk), and experimental infant formula (Modified nutritional substitute for human milk). Each feeding mode was compared as individual variables to group samples due to their importance in shaping early microbial systems. Final data included only baseline samples (Infants 2 months after birth) to minimize possibilities of confounding effects due to age difference.

Statistical Analysis

Group-level differences in the microbial seeding score (S) in feeding modes were assessed by a one-way analysis of variance (ANOVA). Statistical significance was evaluated at an α level of 0.05. Summarized group-level comparisons were used for all analyses as a baseline sample at two months of age. Samples with missing genus-level abundance values for any taxa included in the seeding score were excluded from data analysis. Therefore, no imputation was performed. All bacterial abundances were then normalized as relative percentages to explain differences in sequencing depth. At this time, there was no additional transformation applied. Distributional plots were used to visually evaluate assumptions of normality and homogeneity of variance. In addition, with the exploratory nature of the analysis and comparative group variances, one-way ANOVA was applied as it was appropriate for group-level comparison.

RESULTS

This analysis evaluated differences in the microbial seeding score (S) in infant feeding modes at a two-month baseline. A total of 67 independent infant samples at a two-month baseline were analyzed, consisting of breast milk-fed infants (n=16), experimental infant formula-fed infants (n=25), and standard infant formula-fed infants (n=26) (Table 1).

Breast milk-fed infants showed the highest mean seeding score, with an average value of S = 0.391 ± 0.319 (mean ± SD). This value was higher than those reported in the experimental and standard infant formula groups. As shown in Figure 1, standard infant formula-led infants indicated the lowest mean seeding score (S = 0.301 ± 0.309), supporting the observed monotonic decrease in average seeding scores among feeding modes.

One group was fed with an experimental infant formula, which was modified by testers as a way to compare gaps of current literature. Experimental infant formula-fed infants indicated an intermediate mean seeding score of S = 0.352 ± 0.352. This value fell between the breast milk and standard infant formula group. While higher than the standard formula group, the mean seeding score remained lower than that of the breast milk group. The mean seeding score of the experimental infant formula group turned out to remain lower than that of the breast milk group.

A one-way analysis of variance (ANOVA) indicated no statistically significant differences in seeding scores in feeding modes (F(2,64)=0.395, p=0.676). Although group differences were not statistically significant, there

Table 1. Descriptive statistics of microbial seeding scores (S) by infant feeding model at two months of age. *Values in the table represent mean ± standard deviation, and S is defined as the difference between bacterial genera that are maternally associated and environmentally associated.*

Feeding Mode	Sample Size (n)	Mean Seeding Score (S)	Standard Deviation (SD)
Breast Milk	16	0.391	0.319
Experimental Infant Formula	25	0.352	0.352
Standard Infant Formula	26	0.301	0.309

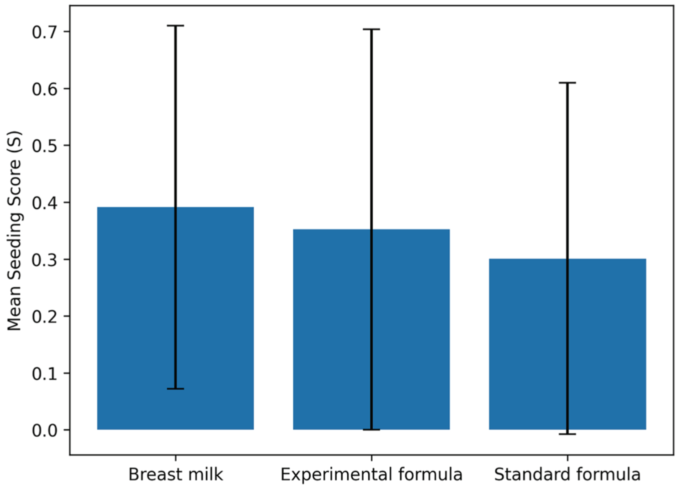

Figure 1. *Mean microbial seeding score (S) by infant feeding mode at two months of age. S is defined as the difference between genera that are maternally associated (Bacteroides and Bifidobacterium) and environmentally associated (Enterococcus, Staphylococcus, Streptococcus, Klebsiella) according to relative abundance percentages. Bars show mean ± standard deviation for breast milk (n=16), experimental formula (n=25), and standard formula (n=26) groups. One-way ANOVA (p=0.676) was used to assess group differences.*

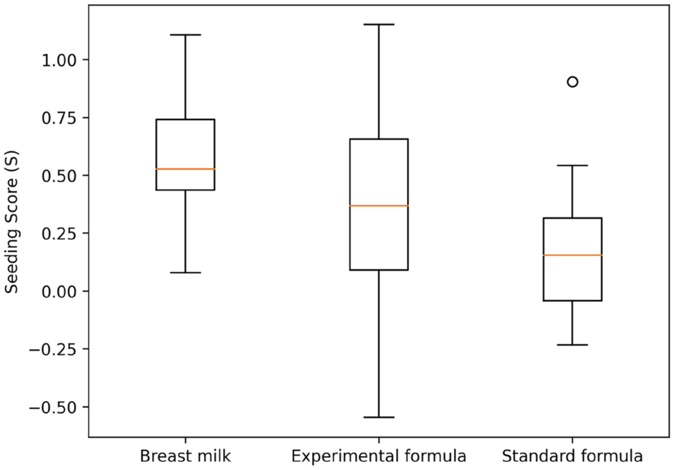

Figure 1. *Distribution of microbial seeding scores (S) among infant feeding modes at two months of age. S is computed as the relative abundance difference between bacterial genera that are maternally associated and environmentally associated. Box plots indicate median, interquartile range, and overall variability within each feeding group (breast milk, experimental formula, and standard formula). There was a substantial inter-individual variability observed that was consistent with non-significant group differences identified by one-way ANOVA (p=0.676).*

was a consistent directional trend in mean seeding scores as observed across feeding modes. Figure 2 exhibits substantial inter-individual variability within feeding groups. This likely contributed to the lack of statistical significance observed in spite of consistent group-level trends. Overall, seeding scores followed a monotonic decreasing pattern in feeding modes in an order from breast milk to experimental formula to standard formula.

DISCUSSION

This study examined whether a simplified microbial seeding score may identify the differences in maternal versus environmental colonization patterns among infant feeding modes at a two-month baseline. Although statistical significance was not observed, the direction of group-level difference was examined in the context of existing literature and the intended exploratory use of the proposed metric. This pattern may reflect biologically relevant variation in early microbial colonization. However, the analysis shall be interpreted as exploratory rather than confirmatory. Limited sample size, high inter-individual variability, or the early developmental stage of the infant microbiome may be reflected by the absence of statistical significance. Seen in this perspective, this study shall be viewed as a proof-of-concept evaluation of a proposed quantitative metric instead of a definitive test of feeding mode effects.

Breast milk fed infants with the highest average seeding score (S) is consistent with enrichment of maternally related taxa as observed in prior studies. Breast milk is recognized as an important source of beneficial bacteria and human milk oligosaccharides that are associated with maternal colonization patterns. Conversely, standard infant formula and experimental infant formula groups exhibited lower average seeding score (S), indicating microbial dominance of the environment, which is consistent with increased representation of environmentally associated taxa.

The experimental infant formula group showed intermediate seeding scores. This suggests partial alignment with maternal colonization patterns. This finding specifically suggests that modifying formula composition may influence early microbial structure. However, the complexity of human milk and the difficulty of replicating its microbial and biochemical properties may be reflected by the persistence of lower seeding scores compared to breastfed infants.

Several limitations need to be considered when interpreting the findings in this study. First, the analysis in this study was restricted to a single two-month baseline timepoint, while microbiome maturation typically happens over the first three years of life. Second, a limited set of bacterial genera was aggregated by using the simplified seeding score, not capturing broader taxonomic or functional diversity. Lastly, the analysis of this study was based on modest sample size and high inter-individual variability that may have limited statistical power to identify group differences.

CONCLUSION

This study introduced a simplified microbial seeding score as a quantitative framework to contrast microbial taxa that were maternally and environmentally associated through the use of secondary infant microbiome data. However, due to the limitations of this study to a single early timepoint and a restricted set of bacterial genera, it is suggested that future studies focus on applying this modeling approach to longitudinal datasets in multiple developmental stages to evaluate trajectories of microbiome maturation over time. It is also recommended to expand this model to functional or metabolomics features to improve the ability of the model to characterize early-life microbial dynamics and nutritional interventions.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486 (7402): 207-14. https://doi.org/10.1038/nature11234
2. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Med*. 2016; 8: 51. https://doi.org/10.1186/s13073-016-0307-y
3. Blum WEH. The role of the microbiome in developmental programming of immune and metabolic health. *Pediatrics*. 2017; 141 (4): e20172437.
4. Arora P, Sharma C, Batra S. Vertical transmission of microbiota during pregnancy. *Front Public Health*. 2020; 8: 225.
5. Dong T, Liu Y, Li S, *et al*. Environmental factors shaping early-life gut microbiome. *Rev Med Microbiol*. 2019; 30 (3): 123-32.
6. Zhou P, Wang Y, Li X, *et al*. Microbiome seeding through the birth canal and its impact on infant development. *Cell Host Microbe*. 2023; 31 (3): 396-410.
7. Davis EC, Wang M, Donovan SM. Breastfeeding shapes the developing infant gut microbiome. *J Allergy Clin Immunol*. 2022; 150 (4): 1175-89.