

## SHORT REPORTS

## Human microbiome variation associated with race and ethnicity emerges as early as 3 months of age

Elizabeth K. Mallott<sup>1,2,3\*</sup>, Alexandra R. Sitarik<sup>4</sup>, Leslie D. Leve<sup>5</sup>, Camille Cioffi<sup>5</sup>, Carlos A. Camargo, Jr<sup>6</sup>, Kohei Hasegawa<sup>6</sup>, Seth R. Bordenstein<sup>1,2,7,8,9,10,11\*</sup>

**1** Vanderbilt Microbiome Innovation Center, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, United States of America, **3** Department of Biology, Washington University in St. Louis, St. Louis, Missouri, United States of America, **4** Department of Public Health Sciences, Henry Ford Health, Detroit, Michigan, United States of America, **5** Prevention Science Institute, University of Oregon, Eugene, Oregon, United States of America, **6** Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **7** Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **8** Vanderbilt Institute for Infection, Immunology, and Inflammation, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **9** Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, School of Medicine, Nashville, Tennessee, United States of America, **10** Departments of Biology and Entomology, Pennsylvania State University, University Park, Pennsylvania, United States of America, **11** The One Health Microbiome Center, Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, United States of America

\* mallott@wustl.edu (EKM); s.bordenstein@psu.edu (SRB)



## OPEN ACCESS

**Citation:** Mallott EK, Sitarik AR, Leve LD, Cioffi C, Camargo CA, Jr, Hasegawa K, et al. (2023) Human microbiome variation associated with race and ethnicity emerges as early as 3 months of age. *PLoS Biol* 21(8): e3002230. <https://doi.org/10.1371/journal.pbio.3002230>

**Academic Editor:** Jotham Suez, Johns Hopkins University Bloomberg School of Public Health, UNITED STATES

**Received:** October 18, 2022

**Accepted:** July 3, 2023

**Published:** August 17, 2023

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3002230>

**Copyright:** © 2023 Mallott et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Sequencing data and metadata included in this study was downloaded from NCBI's Sequence Read Archive (accessions:

## Abstract

Human microbiome variation is linked to the incidence, prevalence, and mortality of many diseases and associates with race and ethnicity in the United States. However, the age at which microbiome variability emerges between these groups remains a central gap in knowledge. Here, we identify that gut microbiome variation associated with race and ethnicity arises after 3 months of age and persists through childhood. One-third of the bacterial taxa that vary across caregiver-identified racial categories in children are taxa reported to also vary between adults. Machine learning modeling of childhood microbiomes from 8 cohort studies (2,756 samples from 729 children) distinguishes racial and ethnic categories with 87% accuracy. Importantly, predictive genera are also among the top 30 most important taxa when childhood microbiomes are used to predict adult self-identified race and ethnicity. Our results highlight a critical developmental window at or shortly after 3 months of age when social and environmental factors drive race and ethnicity-associated microbiome variation and may contribute to adult health and health disparities.

## Introduction

Two major goals of the human microbiome sciences include increasing the representation of undersampled groups in microbiome datasets [1–3] and understanding the tempo by which inequitable experiences, intergenerational inequality, and structural racism impact

PRJNA322554, PRJEB11697, and PRJEB13896), QIITA (studies 11129 and 10894), and FigShare (<https://doi.org/10.6084/m9.figshare.7011272.v3>). Additional sequencing data and metadata for included studies are available as outlined in the original publications, which are listed in [S1 Table](#). All data necessary to reproduce main text and supplementary figures are included in [S1–S15](#) Data files. Code for all analyses can be found on GitHub ([https://github.com/BordensteinLaboratory/Childhood\\_micro\\_metaanalysis](https://github.com/BordensteinLaboratory/Childhood_micro_metaanalysis)) and are archived on Zenodo (<https://doi.org/10.5281/zenodo.8063024>).

**Funding:** The WHEALS study was supported by P01 AI089473-01 from the National Institutes of Health (Bethesda, MD) (ARS). The MARC-35 study was supported by UG3/UH3 OD-023253 from the National Institutes of Health (Bethesda, MD) (CAC and KH). The Early Growth and Development Study microbiome data was supported by UH3 OD023389, P50GM098911, R01 DA035062 from the National Institutes of Health (Bethesda, MD), and a Faculty Alumni Award from the College of Education, University of Oregon (LDL and CC). EKM was supported by the Vanderbilt Microbiome Innovation Center. SRB was supported by the One Health Microbiome Center in the Huck Institutes at The Pennsylvania State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AGP, American Gut Project; ANCOM-BC, analysis of compositions of microbiomes with bias correction; ASV, amplicon sequence variant; AUC, area under the curve; IBS, irritable bowel syndrome; PERMANOVA, permutational multivariate analysis of variance; prAUC, precision recall AUC; SES, socioeconomic status.

microbiome variation and health outcomes [4–8]. Early-life social and environmental exposures can have large and lasting effects on child development and adult health, and perturbations to the gut microbiome may be important to future disease risk [9–19]. In the United States, adult gut microbiome diversity correlates with self-identified race and ethnicity [1,3]. However, socioeconomic status (SES)—neighborhood deprivation index, individual and parental education, or household income—is both correlated with adult gut microbiome diversity and is associated with race and ethnicity [20–24]. We emphasize that race and ethnicity are proxies for inequitable exposure to social and environmental determinants of health due to structural racism [6–8,25,26]. When human microbiome differences arise during development and whether or not distinguishing gut taxa overlap between childhood and adulthood are key questions that have implications for long-term effects of early life experiences, including structural racism, on microbiome variation.

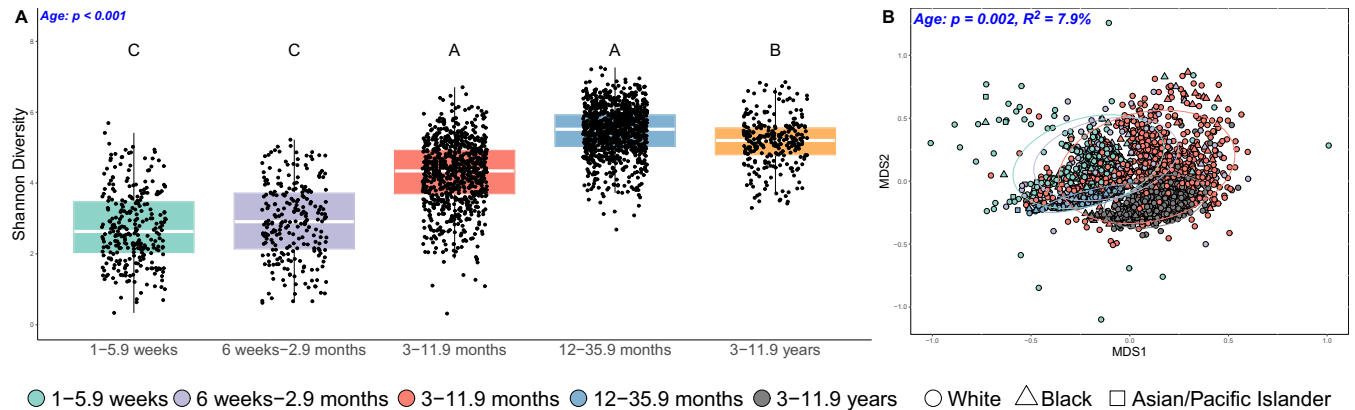
To identify the developmental window when microbiome variation emerges, how long it persists during childhood, and which distinguishing taxa overlap between children and adults, we combined 8 gut microbiome composition datasets from 2,756 samples spanning 729 children between birth and 12 years of age throughout the US ([S1 Table](#)). We used caregiver-identified race (Asian/Pacific Islander, Black, White) and ethnicity (Hispanic, non-Hispanic) to capture complex interactions of multiple biosocial factors that influence gut microbiome composition, even though race and ethnicity are not biological categories that directly influence microbiome variation [5–7,26]. We used a diverse dataset of childhood microbiome samples to identify features of the gut microbiome that are potential markers of the inequitable experiences underlying health disparities. We selected samples from multiple 16S rRNA gene sequencing studies that represent a higher diversity of children than is commonly present in large analyses of the gut microbiome [1–3]. In the present study, 17.2% of samples were from non-White individuals, and 14.3% of samples were from Hispanic individuals. While the majority of samples from Hispanic individuals are from Hispanic White children, some Hispanic Black children are present in the dataset.

## Results

### Microbiome variation emerges at or shortly after 3 months of age

Subject explained the greatest proportion of variation, consistent with other studies of the gut microbiome ([S1 Fig](#)). As age had the second strongest association with gut microbiome composition of the variables tested (Figs 1 and [S1–S9](#) and [S2–S4](#) Tables), we stratified samples by age and analyzed each age category separately while controlling for study differences to disentangle when in development race and ethnicity-associated microbiome variation originates. Delivery route and infant diet were not included in the age-stratified analysis, as they covaried with race and ethnicity ([S10](#) and [S11](#) Figs and [S5 Table](#)).

Notably, race and ethnicity did not significantly vary with gut microbiome alpha diversity (within-individual diversity) or beta diversity (between-individual diversity) in the early weeks and months of life, including the first week, 1 to 5.9 weeks, and 6 weeks to 2.9 months (permutational multivariate analysis of variance (PERMANOVA), all  $p > 0.05$ ) (Figs 2, [S2](#), [S12](#), and [S13](#) and [S2 Table](#)). However, at 3 to 11.9 and 12 to 35.9 months, gut microbiome composition based on UniFrac distances varied slightly but significantly by both race and ethnicity (PERMANOVA, all  $p < 0.05$ ) (Figs 2B, [S2](#), [S12](#), and [S13](#) and [S2 Table](#)). Additionally, most measures of alpha diversity varied across racial categories at 3 to 11.9 months and across both racial and ethnic categories at 12 to 35.9 months (LME,  $p < 0.05$ ) ([Fig 2A](#) and [S4 Table](#)). Pairwise comparisons confirmed that Black individuals had higher within-sample diversity than White individuals at 3 to 11.9 and 12 to 35.9 months for at least one of the 5 measures of diversity ([Fig 2A](#)



**Fig 1. Age structures variation in the gut microbiome.** (A) Boxplots show increases in Shannon diversity with age, and (B) nonmetric multidimensional scaling (NMS2) plots show a significant association of age with weighted UniFrac distances. Colors and 95% confidence ellipses denote age, and shape denotes race. Blue text in the panels highlights significant  $p$ -values. Data underlying this figure can be found in [S1](#) and [S2](#) Data.

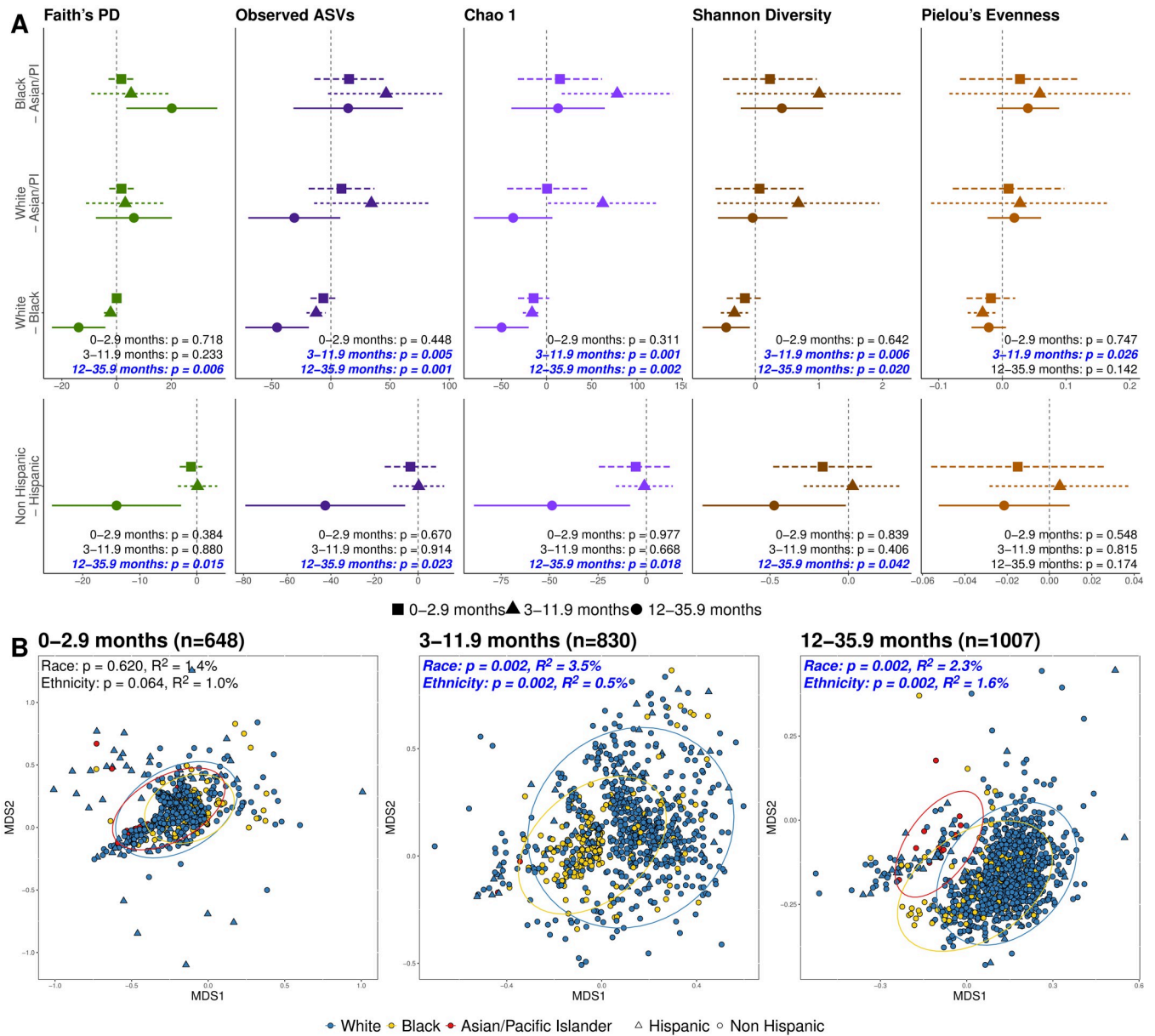
<https://doi.org/10.1371/journal.pbio.3002230.g001>

and [S4 Table](#)) [27]. While higher alpha diversity is consistently associated with better cardio-metabolic health and lower incidence of inflammatory disease in adults [28–30], studies have found mixed results in children. For example, studies of associations between alpha diversity and risk of allergic disease have found negative [31], positive [32], and no [33] association. From 3 to 11.9 years, race associated with gut microbiome composition using only unweighted UniFrac distances (PERMANOVA, all  $p < 0.05$ ) ([S12](#) and [S13](#) Figs and [S2 Table](#)). Collectively, these results reveal that race and ethnicity associate with microbial diversity after 3 months of age, and, notably, this variation persists through childhood years.

### Child gut microbiome variation recapitulates that of adults

To identify differentially abundant taxa, we used analysis of compositions of microbiomes with bias correction (ANCOM-BC) for each variable of interest across all age categories. Age was included as a factor in the models, and numerous taxa were differentially abundant across age categories ([S6–S9 Tables](#)). The abundances of several taxa significantly were associated with race and/or ethnicity in all samples combined ([S5–S9 Tables](#)), including several that varied in abundance between age categories ([S14](#) and [S15](#) Figs). Taxa positively associated with breastfeeding (*Bifidobacterium*, *Lactobacillus*, and *Staphylococcus*) [34,35] were significantly negatively correlated with age, as expected ([S14](#) and [S15](#) Figs and [S9 Table](#)). These taxa were differentially abundant between racial or ethnic categories, likely due to differences in rates of breastfeeding across these groups ([S10](#) and [S11](#) Figs and [S5 Table](#)). Delivery route also differed between racial and ethnic categories—vaginal delivery was more likely than expected in White, Asian/Pacific Islander, and non-Hispanic children and less likely than expected in Black and Hispanic children ([S10](#) and [S11](#) Figs and [S5 Table](#)). However, some individual species within *Bacteroides*, which is often more abundant in vaginally delivered children [34,35], were more enriched in Black and Hispanic children ([S9 Table](#)), contrary to our expectations.

Notably, there was moderate overlap between studies for differentially abundant taxa ([S10 Table](#)). Of the 57 gut microbial taxa that varied in abundance between children of differing self-identified racial categories, 19 were previously identified as differentially abundant between Black and White adult individuals in a recent controlled study of gut microbiome variation [3] ([Fig 3A](#) and [S9 Table](#)). Four of the 19 overlapping taxa were higher in abundance in both Black children and adults compared with White children and adults, and 4 of the

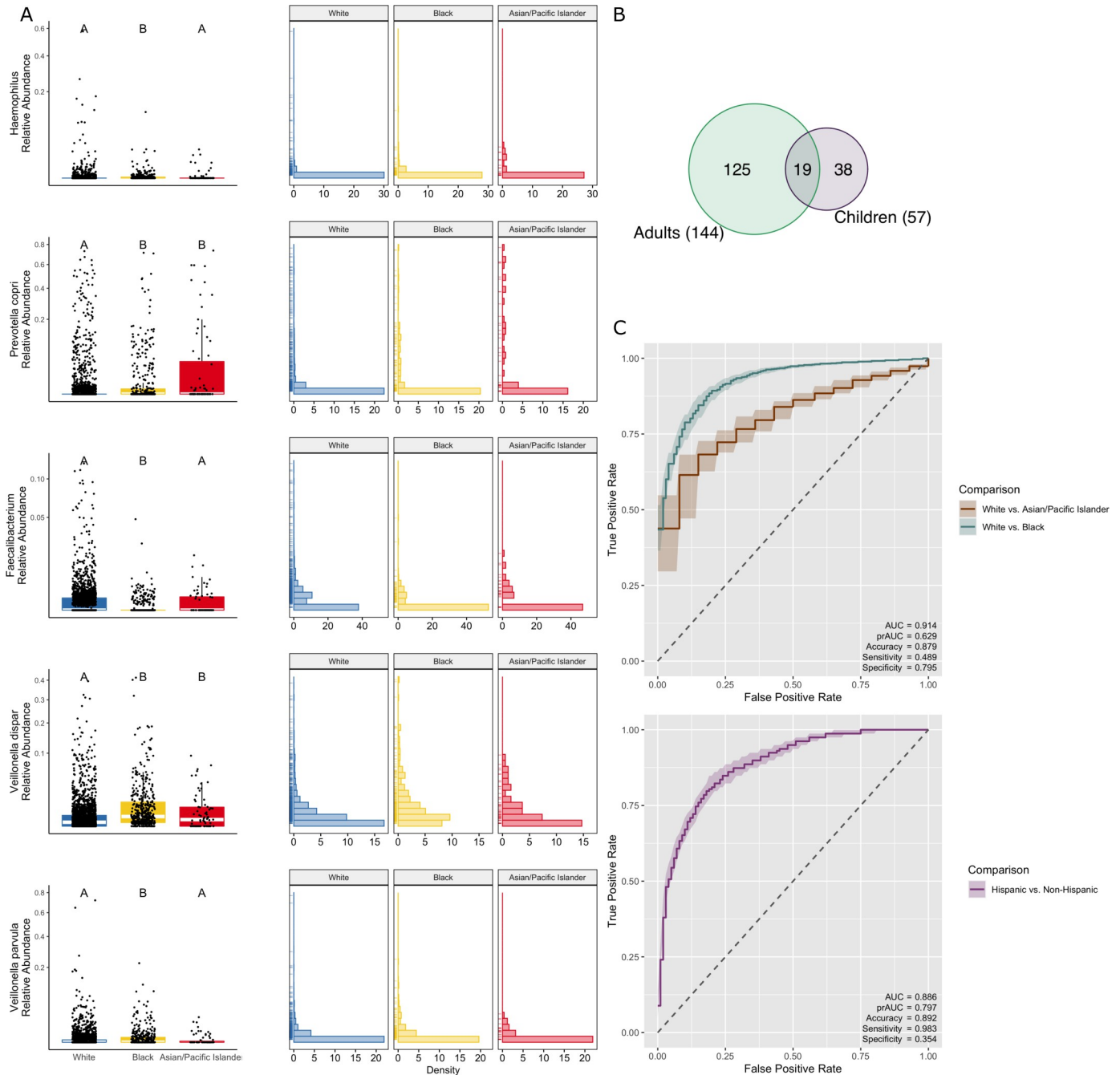


**Fig 2. Microbiome variation emerges at or shortly after 3 months of age.** (A) Dot and whisker plots show estimates for Tukey pairwise comparisons in the alpha diversity linear mixed effects models. Dots indicate the estimated difference in alpha diversity when accounting for other covariates in the model, whiskers denote 95% confidence intervals, and the dashed line indicates zero or no difference. Comparisons with whiskers that do not cross zero indicate a significant difference in alpha diversity between those 2 categories. Colors in the dot whisker plots denote alpha diversity metric, and dot shape and line type denote age category. (B) NMDS plots show weighted UniFrac distances between by race and ethnicity at 0-2.9 months, 3-11.9 months, and 12-35.9 months. Colors and 95% confidence ellipses in the NMDS plots denote race, and shape denotes ethnicity. Blue text in the panels highlights significant *p*-values. NMDS plots for additional age categories and unweighted UniFrac distances can be found in the Supporting information (S12 and S13 Figs). Data underlying this figure can be found in S1, S2, and S4 Data.

<https://doi.org/10.1371/journal.pbio.3002230.g002>

overlapping taxa were lower in abundance in both Black children and adults. The remaining 11 overlapping taxa were either differentially abundant between either Asian/Pacific Islander children and Black children or Asian/Pacific Islander children and White children, or the direction of effect differed between Black and White adults and children. Among the 8 taxa that overlapped and had the same effect in children and adults, *Haemophilus* spp. and





**Fig 3. Child gut microbiome variation recapitulates that of adults.** (A) Boxplots showing the relative abundance of select taxa identified as differentially abundant using ANCOM in the current study that overlap with taxa identified as differentially abundant in adults [3]. All boxplots show the median and interquartile range (IQR), and whiskers extend to 1.5\*IQR. Relative abundances for boxplots and histograms are square root transformed. (B) Venn diagram showing overlapping taxa that are differentially abundant in the gut microbiome between Black individuals and White individuals in the present study in children and in previously published work in adults. (C) Receiver operating characteristic (ROC) curves for a random forest model classifying race and ethnicity metadata based on the gut microbiome. Shading represents a 50% confidence interval around the median. Overall model accuracy for race and ethnicity was >87% (the percentage of samples correctly classified as Asian/Pacific Islander, Black, or White and Hispanic and non-Hispanic). Data underlying this figure can be found in S5, S6, and S7 Data and S9 Table.

<https://doi.org/10.1371/journal.pbio.3002230.g003>

*Prevotella copri* are higher in abundance in both Black children and adults compared to White individuals (*Haemophilus* spp.:  $\log_2\text{FC}_{\text{adults}} = 0.712$ ,  $\log_2\text{FC}_{\text{children}} = 0.739$ ; *P. copri*:  $\log_2\text{FC}_{\text{adults}} = 5.110$ ,  $\log_2\text{FC}_{\text{children}} = 2.513$ ) (ANCOM-BC, all  $q < 0.05$ ) (Fig 3C). These taxa have been associated with an increased risk of autoimmune and allergic diseases, asthma, and obesity across humans in Europe and North America [28,36–39]. *Faecalibacterium*, which is generally considered to be protective against inflammation [33,40], is lower in abundance in Black children and adults compared to White individuals ( $\log_2\text{FC}_{\text{adults}} = -1.356$ ,  $\log_2\text{FC}_{\text{children}} = -0.230$ ) (ANCOM-BC,  $q < 0.05$ ) (Fig 3C). Conversely, *Veillonella*, which is associated with a decreased risk of asthma and allergic disease [33,36], is consistently lower in abundance in White children (*Veillonella dispar*:  $\log_2\text{FC}_{\text{adults}} = 1.295$ ,  $\log_2\text{FC}_{\text{children}} = 0.550$ ; *Veillonella parvula*:  $\log_2\text{FC}_{\text{adults}} = 3.321$ ,  $\log_2\text{FC}_{\text{children}} = 1.010$ ) (ANCOM-BC, both  $q < 0.05$ ) (Fig 3C). Thus, we are finding higher relative abundances of at least one taxon that is positively associated with health in Asian/Pacific Islander, Black, and White children, highlighting the complexity of linking the relative abundance of individual gut microbial taxa to health as a whole. We do note, however, that several of the 19 differentially abundant taxa that overlap between adults and children (S8 Table) have also been found to be associated with SES and unfavorable social and environmental exposures [10,23,41,42].

To detect differentially abundant taxa within each age category, we used generalized linear mixed models with a negative binomial distribution (ANCOM-BC requires more samples per group than we had within each age category). However, few taxa were identified as differentially abundant within each age category (S6–S9 Tables). No phyla or families were differentially abundant between racial and ethnic categories within any age category, and only one genus differed between White and Asian/Pacific Islander children (S6–S9 Tables). Of the 6 species that differed in abundance between racial categories and 4 species that differed in abundance between ethnic categories, none were found in more than one age group (S9 Table). *Coprococcus*, one of the differentially abundant taxa within a specific age group (12 to 35.9 months), was more abundant in non-Hispanic children and has been previously associated both with obesity and a high-fiber diet [43]. The other differentially abundant taxa within specific age groups did not have clear links to health-related outcomes in the literature. Overall, taxa with age-associated variation did not systematically vary by race or ethnicity.

We next used a machine learning approach to identify additional characteristics of the microbiome that may be markers of inequitable exposure to social and environmental determinants of health. A random forest classifier based on the abundance of genera spanning all childhood samples distinguished Black versus White versus Asian/Pacific Islander categories and Hispanic versus non-Hispanic categories with 87% accuracy. Notably, 13 amplicon sequence variants (ASVs) among the top 30 most important genera that increased classification accuracy in the model (S16 and S17 Figs and S11 Table) are taxa identified as differentially abundant between self-identified racial categories in both children in the current study and adults in previous work [3] (Fig 3B and S9 Table). For race, we used a 3-part model, and model performance estimated as area under the curve (AUC; values above 0.5 indicate the classifier is performing better than chance) was 0.914 (Fig 3B). For ethnicity, we used a binary model, and AUC was 0.886 (Fig 3B).

Additionally, we used the childhood microbiome data in a random forest model to assess if childhood microbiome variation predicts that of healthy adults in the American Gut Project (AGP) dataset. As expected, compositional data from children did not reliably distinguish adults of differing racial categories (S18 and S19 Figs), with an AUC of 0.570. Twenty-six of the top 30 taxa identified as important microbiome characteristics in the model using data from children to predict adult metadata were also identified as important taxa in the random forest model that only used data from children (S16 and S19 Figs). However, the taxa with the

highest importance differed with respect to the magnitude and direction of the differences between adults and children (S20 Fig).

Specifically, Enterobacteriaceae and *Prevotella* are highly important in child–child models but are of modest importance in child–adult models (S16 and S19 Figs), and their relative abundances are lowest in White children but highest in White adults (S20 Fig). Other studies have similarly found that specific taxa can be used to differentiate the gut microbiome of groups of people but that the direction of effect can differ between adults and children. *Prevotella* was highly important in both adult and child random forest models used to detect taxa that distinguish the gut microbiome across geographic regions, but the direction of the differences in relative abundance differed [44]. In children, *Prevotella* was more abundant in the US, but *Prevotella* was more abundant in adults outside of the US [44]. *Alistipes* was found to be protective against irritable bowel syndrome (IBS) in adults, but predictive of IBS in children [45].

In contrast, other taxa have a similar direction of effect in both children and adults. *Ruminococcus* is specifically important in the child–adult models, likely due to similar variation in abundance between racial categories in both children and adults (S20 Fig). Higher abundances of *Ruminococcus* are linked with an increased risk of colorectal cancer [46], a disease for which there is a known racial health disparity [47,48]; however, we find that *Ruminococcus* is most abundant in White individuals, a group whose colorectal cancer risk is lower than that of Black individuals but higher than that of Asian/Pacific Islander individuals. Race-associated variation in the relative abundance of *Ruminococcus* across adult guts is not universal, is likely due to a subset of *Ruminococcus* species, and may interact with other factors such as stress or BMI [1,49]. Thus, it is difficult to know how or if the differences observed in the microbiome here contribute directly to health disparities.

## Discussion

Race and ethnicity associate with gut microbiome composition and diversity beginning at 3 months of age, indicative of a narrow window of time (at or shortly after 3 months) and tempo when this variation emerges. Specifically, we found both race and ethnicity account for small but statistically significant proportions of the variation in gut microbiome composition, multiple taxa were differentially abundant between self-reported racial and ethnic categories, several of which were previously identified as differentially abundant in adults [3], and a random forest classifier reliably distinguishes caregiver-identified race and ethnicity. Notably, our findings do not support race- or ethnicity-associated variation appearing at birth or shortly after, when mother-to-infant and other mechanisms of vertical microbial transmission are expected to be strongest [50,51]. None of the differentially abundant taxa identified in the current study are known to be vaginally acquired by infants, and only 2 species are known to be vertically transmitted from the mother [51]. Instead, external factors are most likely shaping race- and ethnicity-associated microbiome variation at or shortly after 3 months. Our results highlight the impetus to increase the diversity of individuals included in studies in the microbiome sciences [1–3] and support the call for studies investigating how structural racism and other structural inequities affect microbiome variation and health [4–7].

The race- and ethnicity-associated differences in the gut microbiome likely reflect differences in environmental and social factors [6–8,25,26]. In the US, there are clear racial and ethnic disparities in health that are tied to differences in these same factors—psychosocial stressors, socioeconomic differences, culture, diet and access to food, access to healthcare and education, interactions with the built environment, and environmental pollutants [6,25,49,52,53]. These factors are important social and environmental determinants of health that have tangible impacts through the modification of human physiology [52,53]. In addition,

there is evidence that the developmental trajectory of the gut microbiome is associated with immune system development, metabolic programming, antibiotic resistance, and risk of asthma, allergic, and autoimmune disease [17,33,36,54–60]. Thus, variation in social and environmental determinants of health that is associated with race and ethnicity may not only shape microbiome variation and impact health but also contribute to health disparities [6,7,20,25]. The tempo and types of factors contributing significantly to race- and ethnicity-associated gut microbiome variation are a priority for research.

Previous studies have identified race- and ethnicity-associated variation in the gut microbiome of children [27,61–64], though they did not pinpoint when in development variation appears and the association is not consistent across studies [36,41,65–73]. In particular, previous work demonstrated that sociodemographic factors related to rates of exposure to stress, access to grocery stores and healthcare, and environmental exposure risk are correlated with race-associated variation in the gut microbiome and that the effect of some of these factors, such as household income, are stronger in infants compared with neonates [27]. Due to the limitations of available metadata for all studies, we were not able to include all factors known to be important in our analysis, such as antibiotic exposure [10,27,74,75], environmental microbial exposures [27,34,56,76], childhood diet [54,70], and various measures of maternal health during pregnancy [9,27,54,63,66,72,77–79]. Many of the studies did not measure potentially important factors that are associated with race and ethnicity, including SES, discrimination or stress, and detrimental environmental exposures. Factors that are known to impact gut microbiome composition and were included in our models—age, sex, delivery route, and infant diet—were not independent of race and/or ethnicity (S10 and S11 Figs and S5 Table). While our study included a relatively high proportions of non-Hispanic Black and Hispanic White children, our inferences were limited by low numbers of Asian American/Pacific Islander children. The datasets used in the current study did not have a sufficient number of Middle Eastern, Native American, and Alaskan Native children to include those individuals in the analysis.

Self-identified race and ethnicity are complex concepts and have limitations. Self-identification varies over time, may not be reflected by predetermined categories used in surveys, and may not capture all aspects of race and ethnicity [80–82]. An additional limitation is that the majority of included studies were conducted in urban areas in distinct geographic locations. The data may not be representative of children from rural areas or the entirety of the US. The results of our study are also not generalizable to other countries due to cultural variation in definitions of racial and ethnic categories. These limitations highlight the necessity of future efforts to recruit a far greater diversity of participants for understanding human microbiome diversity [1–3].

During the first 3 months of age, typically high inter- and intraindividual variability in the infant gut microbiome may contribute to the effect of race and ethnicity, in addition to other maternal, environmental, and social factors that associate with the gut microbiome during this developmental period [35,83,84]. Additionally, the rapid development and marked variation in abundance of microbial taxa within and between individuals continues for at least the first year of life [34,85,86]. Differences in social exposures through childcare, dietary variation due to differential rates of breastfeeding and methods of starting solid food, and environmental exposures through time spent in green spaces may be especially impactful starting at 3 months of age and continuing throughout the first year [9–19,87,88]. Many studies of early life and external factor associations with gut microbiome variation have had limited power to detect the effects of multiple factors, finding few or inconsistent relationships between early life determinants and gut microbiome diversity and composition [10,17,76]. Our findings underscore the need for well-powered, longitudinal studies of diverse cohorts that comprehensively assess



all internal and external factors known to affect the developmental trajectory of the microbiome [5–7,25,89–92]. Other studies have found that the development of the gut microbiome appears to be particularly sensitive to environmental factors and early life events during the first 3 years of life [14,34,93,94]. Additional work is now needed to assess if social and environmental determinants of health begin to influence variation in the microbiome at or near 3 months of age in a way that is potentially important for understanding health disparities in adults, providing a relatively narrow window of time in which to identify potentially impactful factors.

## Materials and methods

Eight datasets with 16S rRNA sequencing data and available race and ethnicity metadata were used in this study [27,66,67,70,72,95,96] (S21 Fig and S1 Table). Individuals between birth and 12 years of age, living in the US, with a caregiver-reported race of Black, White, or Asian/Pacific Islander, and with a caregiver-reported ethnicity of Hispanic or non-Hispanic were included in the analysis. Individuals were not selected based on a known disease phenotype (e.g., type 1 diabetes). Study was included in all models as strata to control for the effects of different study parameters, and individual identity was included as a factor in all models to assess the impact of individual differences on microbiome communities. While sequencing method, primer choice, and sequencing depth did have a significant association with microbial community composition when included in models, including study as strata removed the effect of these study-specific parameters (S2 Table). As some of the included studies had multiple participants from the same family, we also tested if individual identity or family had a larger effect size. In all cases, individual identity explained a larger proportion of the variation than family (S2 Table).

Sequence analyses were carried out in QIIME2 (v.2021.4) [97]. Each study was individually imported into QIIME, and the DADA2 algorithm was used to denoise each study separately to allow us to use appropriate trimming and truncation parameters for each dataset. Feature tables and representative sequences from all studies were then merged using the fragment insertion method [98] to control for differences in amplification and sequencing methodologies between studies. The merged table was filtered to remove sequences absent from the insertion tree. Taxonomy was assigned using a Naïve-Bayesian classifier trained on the Greengenes 13\_8 99% OTU full-length 16S rRNA gene sequence database. Mitochondria and chloroplast sequences were filtered from the merged feature table prior to downstream analysis.

Alpha and beta diversity indices were calculated in QIIME and exported for statistical analysis in R [99]. Linear mixed effects models as implemented in the *lme4* package [100] were used to detect significant associations between race, ethnicity, age, sex, delivery route, and infant diet on multiple measures of within-sample diversity (Faith's PD, observed ASVs, Chao 1, Shannon diversity, and Pielou's evenness). Study and individual identity were included as random effects in all linear models to control for the effects of different study parameters and repeatedly sampling individuals. PERMANOVA, as implemented in the *vegan* package [101], was used to examine associations between race, ethnicity, age, sex, delivery route, and infant diet on unweighted and weighted UniFrac distances (example model:  $\text{WeightedUniFrac} \sim \text{Race} + \text{Ethnicity} + \text{Age} + \text{Sex} + \text{Delivery route} + \text{Infant diet} + \text{SubjectID}$ , strata = Study). Study was included as the strata in the PERMANOVA models to constrain permutations within each study and control for study-specific methodological differences in sample collection and processing. For both the alpha and beta diversity analyses, we additionally examined the effect of sequencing technology, primer set, and sequencing depth (S2 and S4 Tables) (S7–S9 and S21 Figs). Analysis of composition of microbiomes was used to identify differentially abundant

phyla, families, genera, and species across all samples using the *ANCOM-BC* package [102]. Generalized linear models using a negative binomial distribution were used to detect differentially abundant phyla, families, genera, and species within each age category using the *glmmTMB* package [103]. Random forest classification was performed using the *mikropml* package [104] in R. A total of 100 training/test data splits were used for each model, and 5-fold cross-validation was repeated 100 times for each of the 100 training/test data splits using the default settings of the *run\_ml()* command. Median AUC, precision recall AUC (prAUC), accuracy, sensitivity, and specificity are reported for each model.

## Supporting information

### **S1 Text. Impact of age, delivery mode, and infant diet on gut microbiome composition and diversity.**

(DOCX)

**S1 Fig.** Nonmetric multidimensional scaling plots showing the effect of race on weighted (A) and unweighted (B) UniFrac distances in all samples combined. Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S2 Fig. Nonmetric multidimensional scaling plots showing the effect of age on unweighted UniFrac distances.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(EPS)

**S3 Fig.** Nonmetric multidimensional scaling plots showing the effect of sex on weighted (A) and unweighted (B) UniFrac distances. Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S4 Fig.** Nonmetric multidimensional scaling plots showing the effect of delivery mode on weighted (A) and unweighted (B) UniFrac distances. Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S5 Fig. Nonmetric multidimensional scaling plots showing the effect of infant diet on unweighted and weighted UniFrac distances.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S6 Fig. Nonmetric multidimensional scaling plots showing the effect of study on unweighted and weighted UniFrac distances.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S7 Fig. Nonmetric multidimensional scaling plots showing the effect of sequencing technology on unweighted and weighted UniFrac distances.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S8 Fig. Nonmetric multidimensional scaling plots showing the effect of primer set on unweighted and weighted UniFrac distances.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S9 Fig. Nonmetric multidimensional scaling plots showing the effect of sequencing depth on unweighted and weighted UniFrac distances.** Low depth is <20,000 reads, medium depth is 20,000–49,999 reads, and high depth is  $\geq 50,000$  reads. Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(TIF)

**S10 Fig. Barplots of observed and expected numbers of individuals by age category, sex, delivery mode, and infant diet for each racial category.** Data underlying this figure can be found in [S5 Table](#).

(EPS)

**S11 Fig. Barplots of observed and expected numbers of individuals by age category, sex, delivery mode, and infant diet for each ethnicity category.** Data underlying this figure can be found in [S5 Table](#).

(EPS)

**S12 Fig. Nonmetric multidimensional scaling plots showing the effect of race on weighted UniFrac distances for additional age categories.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(EPS)

**S13 Fig. Nonmetric multidimensional scaling plots showing the effect of race on unweighted UniFrac distances within age categories.** Data underlying this figure can be found in [S2](#) and [S3](#) Data.

(EPS)

**S14 Fig. Correlation plots showing the association between age and species relative abundance by race.** Taxa that are differentially abundant across age categories according to ANCOM-BC results and were identified as differentially abundant between racial categories are included. Data underlying this figure can be found in [S8 Data](#).

(EPS)

**S15 Fig. Correlation plots showing the association between age and species relative abundance by ethnicity.** Taxa that are differentially abundant across age categories according to ANCOM-BC results and were identified as differentially abundant between ethnicity categories are included. Data underlying this figure can be found in [S9 Data](#).

(EPS)

**S16 Fig. Feature importance from a random forest model used to identify taxa distinguishing children of different self-identified racial categories.** Dots denote the median importance, and whiskers denote 95% confidence intervals. Data underlying this figure can be found in [S10 Data](#).

(TIFF)

**S17 Fig.** Relative abundances across White (blue), Black (yellow), and Asian/Pacific Islander (red) children of the 13 taxa identified as (1) important features in the random forest model; (2) differentially abundant in the ANCOM analysis; and (3) differentially abundant in a previous study of adult gut microbiomes. All boxplots show the median and interquartile range (IQR), and whiskers extend to  $1.5 \times \text{IQR}$ . Relative abundances for boxplots are square root transformed. Data underlying this figure can be found in [S11 Data](#).

(EPS)

**S18 Fig. Receiver operating characteristic (ROC) curves for a random forest model classifying adult gut microbiome samples by race using samples from children as a training dataset.** Shading represents a 50% confidence interval around the median. Data underlying this figure can be found in [S12 Data](#).

(TIFF)

**S19 Fig. Feature importance from a random forest model used to identify taxa distinguishing adults of different self-identified racial categories based on data from children.** Dots denote the median importance, and whiskers denote 95% confidence intervals. Data underlying this figure can be found in [S13 Data](#).

(TIFF)

**S20 Fig. Relative abundance of highly important features from the random forest models using data from multiple child microbiome studies and adults from the American Gut Project.** Enterobacteriaceae and *Prevotella* (A and B) were highly important in the child–child models and *Ruminococcus* (C) was highly important in the child–adult models. All boxplots show the median and interquartile range (IQR), and whiskers extend to 1.5\*IQR. Relative abundances for boxplots are square root transformed. Data underlying this figure can be found in [S14 Data](#).

(EPS)

**S21 Fig. Box plots showing sequencing depth (number of forward reads prior to filtering for each sample) by study.** Data underlying this figure can be found in [S15 Data](#).

(EPS)

**S22 Fig. Box plots showing Shannon diversity and observed ASV alpha diversity metrics by age.** Data underlying this figure can be found in [S1 Data](#).

(TIF)

**S23 Fig. Box plots showing Shannon diversity and observed ASV alpha diversity metrics by race.** Data underlying this figure can be found in [S1 Data](#).

(TIF)

**S24 Fig. Box plots showing Shannon diversity and observed ASV alpha diversity metrics by ethnicity.** Data underlying this figure can be found in [S1 Data](#).

(TIF)

**S25 Fig. Box plots showing Shannon diversity and observed ASV alpha diversity metrics by sex.** Data underlying this figure can be found in [S1 Data](#).

(TIF)

**S26 Fig. Box plots showing Shannon diversity and observed ASV alpha diversity metrics by infant diet.** Data underlying this figure can be found in [S1 Data](#).

(TIF)

**S27 Fig. Box plots showing Shannon diversity and observed ASV alpha diversity metrics by delivery mode.** Data underlying this figure can be found in [S1 Data](#).

(TIF)

**S1 Table. Characteristics of studies included in the analysis.**

(XLSX)

**S2 Table. Permutational multivariate analysis of variance (PERMANOVA) and homogeneity of variance (Beta dispersion) test statistics.**

(XLSX)

**S3 Table. Pairwise PERMANOVAs statistics for race, ethnicity, and study in the full dataset, as well as race, ethnicity, age, sex, delivery mode, and infant diet for samples where all variables were available.**

(XLSX)

**S4 Table. Linear mixed effects model statistics for alpha diversity comparisons.** Model statistics are reported on the table on the left, and pairwise comparison statistics are presented in the table on the right for variables that were significant.

(XLSX)

**S5 Table. Observed vs. expected numbers of samples for each metadata variable of interest between race and ethnicity categories.**

(XLSX)

**S6 Table. Test statistics for differential abundance analyses at the phyla level.**

(XLSX)

**S7 Table. Test statistics for differential abundance analyses at the family level.**

(XLSX)

**S8 Table. Test statistics for differential abundance analyses at the genus level.**

(XLSX)

**S9 Table. Test statistics for differential abundance analyses at the species level.**

(XLSX)

**S10 Table. Genera identified as differentially abundant between self-identified racial categories across studies.**

(XLSX)

**S11 Table. Important features identified with the random forest classifiers.** Both child-child and child-adult models are listed.

(XLSX)

**S1 Data. Alpha diversity values (Faith's PD, Observed features, Shannon diversity, Pielou's evenness, Chao1) for all samples along with metadata shown in Figs 1A and S22–S27.**

(XLSX)

**S2 Data. MDS1 and MDS2 values for weighted UniFrac distances along with metadata shown in Figs 1B, 2B, S1–S9, and S12.**

(XLSX)

**S3 Data. MDS1 and MDS2 values for weighted UniFrac distances along with metadata shown in Figs S1–S9 and S13.**

(XLSX)

**S4 Data. Confidence intervals for Tukey contrasts from linear mixed effects models of the effect of race and ethnicity on alpha diversity for the 0–2.9 month, 3–11.9 month, and 12–35.9 month age categories.** Tukey contrasts were performed using the multcomp package in R after running linear mixed effects models using the lme4 package in R. The values below are from the summary output of those contrasts.

(XLSX)



**S5 Data. Relative abundance of taxa plotted in Fig 3A along with race.**

(XLSX)

**S6 Data. Taxa that are differentially abundant in children (this study) and adults [3].**

(XLSX)

**S7 Data. Sensitivity, specificity, and false positive rates output from the child-only random forest model.** These data were used to construct the ROC curve in Fig 3C.

(XLSX)

**S8 Data. Relative abundance of taxa in S14 Fig along with race and age metadata.**

(XLSX)

**S9 Data. Relative abundance of taxa in S14 Fig along with ethnicity and age metadata.**

(XLSX)

**S10 Data. Feature importance values for the child-only random forest model.**

(XLSX)

**S11 Data. Relative abundance of taxa plotted in S17 Fig along with race.**

(XLSX)

**S12 Data. Sensitivity, specificity, and false positive rates output from the child-only random forest model.** These data were used to construct the ROC curve in S18 Fig.

(XLSX)

**S13 Data. Feature importance values for the child-adult random forest model.**

(XLSX)

**S14 Data. Relative abundance of taxa plotted in S21 Fig along with race and age group (adults or children).**

(XLSX)

**S15 Data. Sequencing depth for all samples included in the analysis along with study.**

(XLSX)

## Acknowledgments

Computational resources were supported by the Vanderbilt Microbiome Innovation Center. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.

## Author Contributions

**Conceptualization:** Elizabeth K. Mallott, Seth R. Bordenstein.

**Data curation:** Alexandra R. Sitarik, Leslie D. Leve, Camille Cioffi, Carlos A. Camargo, Jr, Kohei Hasegawa.

**Formal analysis:** Elizabeth K. Mallott.

**Funding acquisition:** Alexandra R. Sitarik, Leslie D. Leve, Camille Cioffi, Carlos A. Camargo, Jr, Kohei Hasegawa, Seth R. Bordenstein.

**Investigation:** Alexandra R. Sitarik, Leslie D. Leve, Camille Cioffi, Carlos A. Camargo, Jr, Kohei Hasegawa.

**Methodology:** Elizabeth K. Mallott, Alexandra R. Sitarik, Leslie D. Leve, Camille Cioffi, Carlos A. Camargo, Jr, Kohei Hasegawa.

**Supervision:** Seth R. Bordenstein.

**Visualization:** Elizabeth K. Mallott.

**Writing – original draft:** Elizabeth K. Mallott.

**Writing – review & editing:** Elizabeth K. Mallott, Alexandra R. Sitarik, Leslie D. Leve, Camille Cioffi, Carlos A. Camargo, Jr, Kohei Hasegawa, Seth R. Bordenstein.

## References

1. Brooks AW, Priya S, Blekhman R, Bordenstein SR. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol.* 2018; 16(18):e2006842. <https://doi.org/10.1371/journal.pbio.2006842> PMID: 30513082
2. Abdill RJ, Adamowicz EM, Blekhman R. Public human microbiome data are dominated by highly developed countries. *PLoS Biol.* 2022 Feb 15; 20(2):e3001536. <https://doi.org/10.1371/journal.pbio.3001536> PMID: 35167588
3. Li J, Markowitz RHG, Brooks AW, Mallott EK, Leigh BA, Olszewski T, et al. Individuality and ethnicity eclipse a short-term dietary intervention in shaping microbiomes and viromes. *PLoS Biol.* 2022 Aug 23; 20(8):e3001758. <https://doi.org/10.1371/journal.pbio.3001758> PMID: 35998206
4. De Wolfe TJ, Arefin MR, Benezra A, Rebolledo Gómez M. Chasing Ghosts: Race, Racism, and the Future of Microbiome Research. *mSystems.* 6(5):e00604–21. <https://doi.org/10.1128/mSystems.00604-21> PMID: 34636673
5. Benezra A. Race in the microbiome. *Sci Technol Hum Values.* 2020 Sep 1; 45(5):877–902.
6. Byrd DA, Carson TL, Williams F, Vogtmann E. Elucidating the role of the gastrointestinal microbiota in racial and ethnic health disparities. *Genome Biol.* 2020 Aug 3; 21(1):192. <https://doi.org/10.1186/s13059-020-02117-w> PMID: 32746938
7. Kozik AJ. Frameshift—a vision for human microbiome research. *mSphere.* 2020 Oct 28; 5(5):e00944–20.
8. Ishaq SL, Rapp M, Byerly R, McClellan LS, O’Boyle MR, Nykanen A, et al. Framing the discussion of microorganisms as a facet of social equity in human health. *PLoS Biol.* 2019 Nov 26; 17(11):e3000536. <https://doi.org/10.1371/journal.pbio.3000536> PMID: 31770370
9. Grech A, Collins CE, Holmes A, Lal R, Duncanson K, Taylor R, et al. Maternal exposures and the infant gut microbiome: a systematic review with meta-analysis. *Gut Microbes.* 2021 Jan 1; 13(1):1–30. <https://doi.org/10.1080/19490976.2021.1897210> PMID: 33978558
10. Gschwendtner S, Kang H, Thiering E, Kublik S, Fösel B, Schulz H, et al. Early life determinants induce sustainable changes in the gut microbiome of six-year-old children. *Sci Rep.* 2019; 9:12675. <https://doi.org/10.1038/s41598-019-49160-7> PMID: 31481742
11. Stiemsma LT, Michels KB. The role of the microbiome in the developmental origins of health and disease. *Pediatrics.* 2018; 141(4):e20172437. <https://doi.org/10.1542/peds.2017-2437> PMID: 29519955
12. Subramanian S, Huq S, Yatsunenkov T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature.* 2014; 510(7505):417–421. <https://doi.org/10.1038/nature13421> PMID: 24896187
13. Vallès Y, Francino MP. Air pollution, early life microbiome, and development. *Curr Envir Health Rpt.* 2018 Dec 1; 5(4):512–21. <https://doi.org/10.1007/s40572-018-0215-y> PMID: 30269309
14. Arrieta MC, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. The Intestinal Microbiome in Early Life: Health and Disease. *Front Immunol.* 2014; 5(5):427. <https://doi.org/10.3389/fimmu.2014.00427> PMID: 25250028
15. Stinson LF. Establishment of the early-life microbiome: a DOHaD perspective. *J Dev Orig Health Dis.* 2020 Jun; 11(3):201–210. <https://doi.org/10.1017/S2040174419000588> PMID: 31601287
16. Montoya-Williams D, Lemas DJ, Spiryda L, Patel K, Carney OO, Neu J, et al. The Neonatal Microbiome and Its Partial Role in Mediating the Association between Birth by Cesarean Section and Adverse Pediatric Outcomes. *Neonatology.* 2018; 114(2):103–111. <https://doi.org/10.1159/000487102> PMID: 29788027

17. Sbihi H, Boutin RCT, Cutler C, Suen M, Finlay BB, Turvey SE. Thinking bigger: How early-life environmental exposures shape the gut microbiome and influence the development of asthma and allergic disease. *Allergy*. 2019; 74(11):2103–2115. <https://doi.org/10.1111/all.13812> PMID: 30964945
18. Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. *Nat Med*. 2016 Jul; 22(7):713–722. <https://doi.org/10.1038/nm.4142> PMID: 27387886
19. Sarkar A, Yoo JY, Valeria Ozorio Dutra S, Morgan KH, Groer M. The Association between Early-Life Gut Microbiota and Long-Term Health and Diseases. *J Clin Med*. 2021 Jan 25; 10(3):459. <https://doi.org/10.3390/jcm10030459> PMID: 33504109
20. Amato KR, Arrieta MC, Azad MB, Bailey MT, Broussard JL, Bruggeling CE, et al. The human gut microbiome and health inequities. *Proc Natl Acad Sci U S A*. 2021 Jun 22 [cited 2021 Jun 14]; 118(25). Available from: <http://www.pnas.org/content/118/25/e2017947118> <https://doi.org/10.1073/pnas.2017947118> PMID: 34161260
21. Bowyer R, Jackson M, Le Roy C, Ni Lochlainn M, Spector T, Dowd J, et al. Socioeconomic status and the gut microbiome: A TwinsUK cohort study. *Microorganisms*. 2019; 7(1):17. <https://doi.org/10.3390/microorganisms7010017> PMID: 30641975
22. Harrison CA, Taren D. How poverty affects diet to shape the microbiota and chronic disease. *Nat Rev Immunol*. 2018 Apr; 18(4):279–287. <https://doi.org/10.1038/nri.2017.121> PMID: 29109542
23. Lewis CR, Bonham KS, McCann SH, Volpe AR, D'Sa V, Naymik M, et al. Family SES Is Associated with the Gut Microbiome in Infants and Children. *Microorganisms*. 2021 Aug; 9(8):1608. <https://doi.org/10.3390/microorganisms9081608> PMID: 34442687
24. Miller GE, Engen PA, Gillevet PM, Shaikh M, Sikaroodi M, Forsyth CB, et al. Lower neighborhood socioeconomic status associated with reduced diversity of the colonic microbiota in healthy adults. *PLoS ONE*. 2016; 11(2):e0148952. <https://doi.org/10.1371/journal.pone.0148952> PMID: 26859894
25. Findley K, Williams DR, Grice EA, Bonham VL. Health disparities and the microbiome. *Trends Microbiol*. 2016 Nov 1; 24(11):847–850. <https://doi.org/10.1016/j.tim.2016.08.001> PMID: 27712950
26. Fortenberry JD. The uses of race and ethnicity in human microbiome research. *Trends Microbiol*. 2013 Apr 1; 21(4):165–166. <https://doi.org/10.1016/j.tim.2013.01.001> PMID: 23566633
27. Levin AM, Sitarik AR, Havstad SL, Fujimura KE, Wegienka G, Cassidy-Bushrow AE, et al. Joint effects of pregnancy, sociocultural, and environmental factors on early life gut microbiome structure and diversity. *Sci Rep*. 2016 Aug 25; 6(1):31775. <https://doi.org/10.1038/srep31775> PMID: 27558272
28. Stanislawski MA, Dabelea D, Lange LA, Wagner BD, Lozupone CA. Gut microbiota phenotypes of obesity. *NPJ Biofilms Microbiomes*. 2019 Jul 1; 5(1):18. <https://doi.org/10.1038/s41522-019-0091-8> PMID: 31285833
29. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009; 457:480–484. <https://doi.org/10.1038/nature07540> PMID: 19043404
30. Forbes JD, Chen C Knox NC, Marrie R-A, El-Gabalawy H, de Kievit T, et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases—does a common dysbiosis exist? *Microbiome*. 2018; 6:221. <https://doi.org/10.1186/s40168-018-0603-4> PMID: 30545401
31. Moroishi Y, Gui J, Hoen AG, Morrison HG, Baker ER, Nadeau KC, et al. The relationship between the gut microbiome and the risk of respiratory infections among newborns. *Commun Med*. 2022 Jul 14; 2(1):1–8. <https://doi.org/10.1038/s43856-022-00152-1> PMID: 35847562
32. Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin Exp Allergy*. 2014; 44(6):842–850. <https://doi.org/10.1111/cea.12253> PMID: 24330256
33. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med*. 2015 Sep 30; 7(307):307ra152–307ra152. <https://doi.org/10.1126/scitranslmed.aab2271> PMID: 26424567
34. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 2018; 562(7728):583–588. <https://doi.org/10.1038/s41586-018-0617-x> PMID: 30356187
35. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe*. 2015; 17(5):690–703. <https://doi.org/10.1016/j.chom.2015.04.004> PMID: 25974306
36. Savage JH, Lee-Sarwar KA, Sordillo J, Bunyavanich S, Zhou Y, O'Connor G, et al. A prospective microbiome-wide association study of food sensitization and food allergy in early childhood. *Allergy*. 2018; 73(1):145–152. <https://doi.org/10.1111/all.13232> PMID: 28632934

37. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, et al. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *Mathis D, editor. Elife*. 2013 Nov 5; 2: e01202. <https://doi.org/10.7554/eLife.01202> PMID: 24192039
38. Michail S, Lin M, Frey MR, Fanter R, Paliy O, Hilbush B, et al. Altered gut microbial energy and metabolism in children with non-alcoholic fatty liver disease. *FEMS Microbiol Ecol*. 2015 Feb 1; 91(2):1–9. <https://doi.org/10.1093/femsec/fiu002> PMID: 25764541
39. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosch D, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med*. 2016 Oct; 22(10):1187–1191. <https://doi.org/10.1038/nm.4176> PMID: 27618652
40. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux JJ, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A*. 2008 Oct 28; 105(43):16731–16736. <https://doi.org/10.1073/pnas.0804812105> PMID: 18936492
41. Flannery JE, Stagaman K, Burns AR, Hickey RJ, Roos LE, Giuliano RJ, et al. Gut Feelings Begin in Childhood: the Gut Metagenome Correlates with Early Environment, Caregiving, and Behavior. *mBio*. 2020 Feb 25; 11(1):e02780–19. <https://doi.org/10.1128/mBio.02780-19> PMID: 31964729
42. Bailey MT, Dowd SE, Galley JD, Hufnagle AR, Allen RG, Lyte M. Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation. *Brain Behav Immun*. 2011 Mar 1; 25(3):397–407. <https://doi.org/10.1016/j.bbi.2010.10.023> PMID: 21040780
43. Wang Z, Usyk M, Vázquez-Baeza Y, Chen GC, Isasi CR, Williams-Nguyen JS, et al. Microbial co-occurrence complicates associations of gut microbiome with US immigration, dietary intake and obesity. *Genome Biol*. 2021 Dec 10; 22(1):336. <https://doi.org/10.1186/s13059-021-02559-w> PMID: 34893089
44. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012; 486:222–228. <https://doi.org/10.1038/nature11053> PMID: 22699611
45. Pittayanon R, Lau JT, Yuan Y, Leontiadis GI, Tse F, Surette M, et al. Gut Microbiota in Patients With Irritable Bowel Syndrome—A Systematic Review. *Gastroenterology*. 2019 Jul 1; 157(1):97–108. <https://doi.org/10.1053/j.gastro.2019.03.049> PMID: 30940523
46. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial Dysbiosis in Colorectal Cancer (CRC) Patients. *PLoS ONE*. 2011 Jan 27; 6(1):e16393. <https://doi.org/10.1371/journal.pone.0016393> PMID: 21297998
47. Ollberding NJ, Nomura AMY, Wilkens LR, Henderson BE, Kolonel LN. Racial/Ethnic Differences in Colorectal Cancer Risk: The Multiethnic Cohort Study. *Int J Cancer*. 2011 Oct 15; 129(8):1899–1906. <https://doi.org/10.1002/ijc.25822> PMID: 21128280
48. Carethers JM. Chapter Six—Racial and ethnic disparities in colorectal cancer incidence and mortality. In: Berger FG, Boland CR, editors. *Advances in Cancer Research [Internet]*. Academic Press; 2021 [cited 2023 May 9]. p. 197–229. (Novel Approaches to Colorectal Cancer; vol. 151). Available from: <https://www.sciencedirect.com/science/article/pii/S0065230X21000233>
49. Carson TL, Wang F, Cui X, Jackson BE, Van Der Pol WJ, Lefkowitz EJ, et al. Associations between Race, Perceived Psychological Stress, and the Gut Microbiota in a Sample of Generally Healthy Black and White Women: A Pilot Study on the Role of Race and Perceived Psychological Stress. *Psychosom Med*. 2018 Sep; 80(7):640–648. <https://doi.org/10.1097/PSY.0000000000000614> PMID: 29901485
50. Korpela K, Costea P, Coelho LP, Kandels-Lewis S, Willemsen G, Boomsma DI, et al. Selective maternal seeding and environment shape the human gut microbiome. *Genome Res*. 2018; 28:561–568. <https://doi.org/10.1101/gr.233940.117> PMID: 29496731
51. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe*. 2018 Jul 11; 24(1):133–145.e5. <https://doi.org/10.1016/j.chom.2018.06.005> PMID: 30001516
52. Williams DR, Lawrence JA, Davis BA. Racism and Health: Evidence and Needed Research. *Annu Rev Public Health*. 2019; 40(1):105–125. <https://doi.org/10.1146/annurev-publhealth-040218-043750> PMID: 30601726
53. Williams DR, Sternthal M. Understanding Racial-ethnic Disparities in Health: Sociological Contributions. *J Health Soc Behav*. 2010 Mar 1; 51(1\_suppl):S15–S27. <https://doi.org/10.1177/0022146510383838> PMID: 20943580
54. Zhong H, Penders J, Shi Z, Ren H, Cai K, Fang C, et al. Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. *Microbiome*. 2019; 7(1):2. <https://doi.org/10.1186/s40168-018-0608-z> PMID: 30609941

55. Busi SB, de Nies L, Habier J, Wampach L, Fritz JV, Heintz-Buschart A, et al. Persistence of birth mode-dependent effects on gut microbiome composition, immune system stimulation and antimicrobial resistance during the first year of life. *ISME Commun.* 2021 Mar 26; 1:8. <https://doi.org/10.1038/s43705-021-00003-5> PMID: 36717704
56. Depner M, Taft DH, Kirjavainen PV, Kalanetra KM, Karvonen AM, Peschel S, et al. Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma. *Nat Med.* 2020 Nov 2; 26:1766–1775. <https://doi.org/10.1038/s41591-020-1095-x> PMID: 33139948
57. Calatayud M, Koren O, Collado MC. Maternal Microbiome and Metabolic Health Program Microbiome Development and Health of the Offspring. *Trends Endocrinol Metab.* 2019; 30:735–744. <https://doi.org/10.1016/j.tem.2019.07.021> PMID: 31493988
58. Lebeaux RM, Coker MO, Dade EF, Palys TJ, Morrison HG, Ross BD, et al. The infant gut resistome is associated with *E. coli* and early-life exposures. *BMC Microbiol.* 2021 Jul 2; 21(1):201. <https://doi.org/10.1186/s12866-021-02129-x> PMID: 34215179
59. Olivares M, Walker AW, Capilla A, Benítez-Páez A, Palau F, Parkhill J, et al. Gut microbiota trajectory in early life may predict development of celiac disease. *Microbiome.* 2018; 6:36. <https://doi.org/10.1186/s40168-018-0415-6> PMID: 29458413
60. Vatanen T, Franzosa EA, Schwager R, Tripathi S, Arthur TD, Vehik K, et al. The human gut microbiome of early onset type 1 diabetes in the TEDDY study. *Nature.* 2018; 562(7728):589–594.
61. Hollister EB, Riehle K, Luna RA, Weidler EM, Rubio-Gonzales M, Mistretta TA, et al. Structure and function of the healthy pre-adolescent pediatric gut microbiome. *Microbiome.* 2015 Aug 26; 3(1):36. <https://doi.org/10.1186/s40168-015-0101-x> PMID: 26306392
62. Sordillo JE, Zhou Y, McGeachie MJ, Ziniti J, Lange N, Laranjo N, et al. Factors influencing the infant gut microbiome at age 3–6 months: Findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). *J Allergy Clin Immunol.* 2017 Feb 1; 139(2):482–491.e14.
63. Stearns JC, Zulyniak MA, de Souza RJ, Campbell NC, Fontes M, Shaikh M, et al. Ethnic and diet-related differences in the healthy infant microbiome. *Genome Med.* 2017 Mar 29; 9(1):32. <https://doi.org/10.1186/s13073-017-0421-5> PMID: 28356137
64. Balakrishnan B, Selvaraju V, Chen J, Ayine P, Yang L, Ramesh Babu J, et al. Ethnic variability associating gut and oral microbiome with obesity in children. *Gut Microbes.* 2021 Jan 1; 13(1):1882926. <https://doi.org/10.1080/19490976.2021.1882926> PMID: 33596768
65. Baumann-Dudenhofer AM, D'Souza AW, Tarr PI, Warner BB, Dantas G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat Med.* 2018; 24:1822–1829. <https://doi.org/10.1038/s41591-018-0216-2> PMID: 30374198
66. Chu DM, Antony KM, Ma J, Prince AL, Showalter L, Moller M, et al. The early infant gut microbiome varies in association with a maternal high-fat diet. *Genome Med.* 2016; 8(1):77. <https://doi.org/10.1186/s13073-016-0330-z> PMID: 27503374
67. Cioffi CC, Tavalire HF, Neiderhiser JM, Bohannon B, Leve LD. History of breastfeeding but not mode of delivery shapes the gut microbiome in childhood. *PLoS ONE.* 2020 Jul 2; 15(7):e0235223. <https://doi.org/10.1371/journal.pone.0235223> PMID: 32614839
68. Galley JD, Bailey M, Dush CK, Schoppe-Sullivan S, Christian LM. Maternal Obesity Is Associated with Alterations in the Gut Microbiome in Toddlers. *PLoS ONE.* 2014 Nov 19; 9(11):e113026. <https://doi.org/10.1371/journal.pone.0113026> PMID: 25409177
69. Grier A, McDavid A, Wang B, Qui X, Java J, Bandyopadhyay S, et al. Neonate gut and respiratory microbiota: coordinated development through time and space. *Microbiome.* 2018; 6(1):193.
70. Herman DR, Rhoades N, Mercado J, Argueta P, Lopez U, Flores GE. Dietary Habits of 2- to 9-Year-Old American Children Are Associated with Gut Microbiome Composition. *J Acad Nutr Diet.* 2019; 120(4):517–534. <https://doi.org/10.1016/j.jand.2019.07.024> PMID: 31668602
71. Gao W, Salzwedel AP, Carlson AL, Xia K, Azcarate-Peril MA, Styner MA, et al. Gut microbiome and brain functional connectivity in infants—a preliminary study focusing on the amygdala. *Psychopharmacology (Berl).* 2019 May 1; 236(5):1641–1651. <https://doi.org/10.1007/s00213-018-5161-8> PMID: 30604186
72. Robinson A, Fiechtner L, Roche B, Ajami NJ, Petrosino JF, Camargo CA, et al. Association of maternal gestational weight gain with the infant fecal microbiota. *J Pediatr Gastroenterol Nutr.* 2017 Nov; 65(5):509. <https://doi.org/10.1097/MPG.0000000000001566> PMID: 28272161
73. Zhang M, Differding MK, Benjamin-Neelon SE, Østbye T, Hoyo C, Mueller NT. Association of prenatal antibiotics with measures of infant adiposity and the gut microbiome. *Ann Clin Microbiol Antimicrob.* 2019 Jun 21; 18(1):18. <https://doi.org/10.1186/s12941-019-0318-9> PMID: 31226994



74. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. *Pediatrics*. 2006 Aug 1; 118(2):511–521. <https://doi.org/10.1542/peds.2005-2824> PMID: 16882802
75. Yassour M, Vatanen T, Siljander H, Hämäläinen AM, Härkönen T, Ryhänen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med*. 2016 Jun 15; 8(343):343ra81–343ra81. <https://doi.org/10.1126/scitranslmed.aad0917> PMID: 27306663
76. Quin C, Gibson DL. Human behavior, not race or geography, is the strongest predictor of microbial succession in the gut bacteriome of infants. *Gut Microbes*. 2020 Sep 2; 11(5):1143–1171. <https://doi.org/10.1080/19490976.2020.1736973> PMID: 32249675
77. Lundgren SN, Madan JC, Emond JA, Morrison HG, Christensen BC, Karagas MR, et al. Maternal diet during pregnancy is related with the infant stool microbiome in a delivery mode-dependent manner. *Microbiome*. 2018; 6:109. <https://doi.org/10.1186/s40168-018-0490-8> PMID: 29973274
78. Singh SB, Madan J, Coker M, Hoen A, Baker ER, Karagas MR, et al. Does birth mode modify associations of maternal pre-pregnancy BMI and gestational weight gain with the infant gut microbiome? *Int J Obes (Lond)*. 2020; 44:23–32. <https://doi.org/10.1038/s41366-018-0273-0> PMID: 30765892
79. Sugino KY, Paneth N, Comstock SS. Michigan cohorts to determine associations of maternal pre-pregnancy body mass index with pregnancy and infant gastrointestinal microbial communities: Late pregnancy and early infancy. *PLoS ONE*. 2019 Mar 18; 14(3):e0213733. <https://doi.org/10.1371/journal.pone.0213733> PMID: 30883572
80. Ford CL, Harawa NT. A new conceptualization of ethnicity for social epidemiologic and health equity research. *Soc Sci Med*. 2010 Jul 1; 71(2):251–258. <https://doi.org/10.1016/j.socscimed.2010.04.008> PMID: 20488602
81. Cobb RJ, Thomas CS, Laster Pirtle WN, Darity WA. Self-identified race, socially assigned skin tone, and adult physiological dysregulation: Assessing multiple dimensions of “race” in health disparities research. *SSM Popul Health*. 2016 Dec 1; 2:595–602. <https://doi.org/10.1016/j.ssmph.2016.06.007> PMID: 29349174
82. Roth WD. The multiple dimensions of race. *Ethn Racial Stud*. 2016 Jun 20; 39(8):1310–1338.
83. Yassour M, Jason E, Hogstrom LJ, Arthur TD, Tripathi S, Siljander H, et al. Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe*. 2018 11; 24(1):146–154.e4. <https://doi.org/10.1016/j.chom.2018.06.007> PMID: 30001517
84. Shao Y, Forster SC, Tsaliki E, Vervier K, Strang A, Simpson N, et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*. 2019; 574:117–121. <https://doi.org/10.1038/s41586-019-1560-1> PMID: 31534227
85. Vatanen T, Plichta DR, Somani J, Münch PC, Arthur TD, Hall AB, et al. Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. *Nat Microbiol*. 2019; 4:470–479. <https://doi.org/10.1038/s41564-018-0321-5> PMID: 30559407
86. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med*. 2015 Oct; 21(10):1228–1234. <https://doi.org/10.1038/nm.3950> PMID: 26366711
87. Amir A, Erez-Granat O, Braun T, Sosnovski K, Hadar R, BenShoshan M, et al. Gut microbiome development in early childhood is affected by day care attendance. *NPJ Biofilms Microbiomes*. 2022 Jan 11; 8(1):1–10.
88. Homann CM, Rossel CAJ, Dizzell S, Bervoets L, Simioni J, Li J, et al. Infants’ First Solid Foods: Impact on Gut Microbiota Development in Two Intercontinental Cohorts. *Nutrients*. 2021 Aug; 13(8):2639. <https://doi.org/10.3390/nu13082639> PMID: 34444798
89. Demmer RT. The microbiome and population health: considerations to enhance study design and data analysis in observational and interventional epidemiology. *Am J Epidemiol*. 2018; 187(6):1291–4.
90. Dowd JB, Renson A. “Under the skin” and into the gut: Social epidemiology of the microbiome. *Curr Epidemiol Rep*. 2018 Dec; 5(4):432–441. <https://doi.org/10.1007/s40471-018-0167-7> PMID: 30596004
91. Herd P, Palloni A, Rey F, Dowd JB. Social and population health science approaches to understand the human microbiome. *Nat Hum Behav*. 2018; 2(11):808–815. <https://doi.org/10.1038/s41562-018-0452-y> PMID: 31457107
92. Ishaq SL, Parada FJ, Wolf PG, Bonilla CY, Carney MA, Benezra A, et al. Introducing the Microbes and Social Equity Working Group: Considering the Microbial Components of Social, Environmental, and Health Justice. *mSystems*. 0(0):e00471–21.

93. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci*. 2011; 108(Suppl 1):4578–4585. <https://doi.org/10.1073/pnas.1000081107> PMID: 20668239
94. Stiemsma LT, Turvey SE. Asthma and the microbiome: defining the critical window in early life. *Allergy Asthma Clin Immunol*. 2017 Jan 6; 13:3. <https://doi.org/10.1186/s13223-016-0173-6> PMID: 28077947
95. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med*. 2017 Mar; 23(3):314–326. <https://doi.org/10.1038/nm.4272> PMID: 28112736
96. Planer JD, Peng Y, Kau AL, Blanton LV, Ndao IM, Tarr PI, et al. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature*. 2016 Jun; 534(7606):263–266. <https://doi.org/10.1038/nature17940> PMID: 27279225
97. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019 Aug 24; 37(8):852–857. <https://doi.org/10.1038/s41587-019-0209-9> PMID: 31341288
98. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*. 3(3): e00021–18. <https://doi.org/10.1128/mSystems.00021-18> PMID: 29719869
99. Team R. R: A language and environment for statistical computing [Internet]. 2019. Available from: <https://www.R-project.org/>
100. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. lme4: Linear Mixed-Effects Models using “Eigen” and S4 [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=lme4>
101. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community ecology package [Internet]. 2019. Available from: <https://cran.r-project.org/package=vegan>
102. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun*. 2020 Jul 14; 11(1):3514. <https://doi.org/10.1038/s41467-020-17041-7> PMID: 32665548
103. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J*. 2017; 9(2):378–400.
104. Topçuoğlu BD, Lapp Z, Sovacool KL, Snitkin E, Wiens J, Schloss PD. mikropml: User-Friendly R Package for Supervised Machine Learning Pipelines. *J Open Source Softw*. 2021; 6(61):3073. <https://doi.org/10.21105/joss.03073> PMID: 34414351