Bacterial microbiome of breast milk and child saliva from low-income Mexican-American women and children

Veronica Davé[1], Kelly Street[1], Stephen Francis[1], Asa Bradman[1,2], Lee Riley[1], Brenda Eskenazi[1,2] and Nina Holland[1,2]

BACKGROUND: The childhood salivary microbiome, which plays an important role in healthy development, may be influenced by breast milk consumption. The composition of the milk microbiome and the role it plays in the establishment of the infant microbiome are not well understood.

METHODS: Here, we sequenced the bacterial 16S rRNA gene to characterize microbial communities in breast milk and 5-year-old child saliva from 10 low-income, Mexican-American mother–child pairs with a high prevalence of obesity.

RESULTS: Members of the genus *Streptococcus* dominated both milk and salivary microbial communities in most subjects. *Staphylococcus* was observed predominately in milk samples while *Prevotella* was more prevalent in child saliva. No statistically significant relationships were observed between maternal and child microbiomes or between child microbiome and BMI. However, prepregnancy BMI was correlated with both lower *Streptococcus* abundance ($r = -0.67$) and higher microbial diversity ($r = 0.77$) in breast milk ($P < 0.05$ for both). Diversity estimates were notably similar to data from other low-income cohorts or children.

CONCLUSION: These findings contribute to the currently limited state of knowledge regarding the breast milk and salivary microbiomes in mother–child pairs and may inform future studies seeking to elucidate the relationship between early-life microbial exposures and pediatric health.

The human microbiome, often called the "second genome", plays an important role in many aspects of health and disease (1). Perturbations in the gut microbiota may be associated with the promotion of atherosclerosis (2), celiac disease (3), and adult and childhood obesity (4). Likewise, changes in the oral microbiome have been linked to periodontal diseases and dental caries (5), pediatric inflammatory bowel disease (6), and infant birth weight (7). In recent years, next-generation sequencing of the hypervariable region of the bacterial 16S rRNA gene has become a powerful tool for the assessment of human microbial communities, and large-scale efforts such as those carried out by the Human Microbiome Project Consortium (8) have begun to characterize the diversity of the microbiome in healthy adults. Despite the growing importance of the microbiome in epidemiologic and environmental health research, it remains poorly understood how the human microbiome is first established and subsequently maintained throughout childhood and adulthood.

The salivary microbiome is likely to play an important role in children's health through seeding the infant gut and preventing or participating in the development of infection (5). The early oral microbiome may also dictate the composition of the long-term stable adult oral microbiome (9). Therefore, understanding the process of oral microbiome establishment in infants and young children may shed light on molecular mechanisms linking early life exposures and microbiome-related health outcomes in later life. A variety of factors are already known to affect the development of the infant salivary microbiome, including mode of delivery (10) and interaction with the primary caregiver (11–13). Though microbial colonization of the oral cavity may begin *in utero* (14), it has recently been established that human milk is home to a diverse community of bacterial species (15) and may also contribute to the establishment of healthy infant oral and gut microbiomes, as indicated by observable differences between the microbiota of breast- and formula-fed infants (14–16). Likewise, there is evidence to suggest that early life exposures can indeed exert long-term effects on gut microbiome composition (17,18). Despite this potential for breast milk consumption to modulate the microbiome, a limited number of published studies have examined the bacterial composition of breast milk, and few have assessed both breast milk and salivary microbial communities in mother–child pairs.

In this pilot study, we employed next-generation sequencing of the bacterial 16S rRNA gene isolated from breast milk samples from 10 mothers and saliva samples from their young children 5 y later. Mother–child pairs were participants in the Center for the Health Assessment of Mothers and Children of Salinas, CA (CHAMACOS) longitudinal birth cohort study (19). The CHAMACOS study, which investigates the health of low-income, Mexican-American women and children in an agricultural community, affords a unique opportunity to compare the maternal milk and child salivary microbiomes across

a period of several years through the use of banked specimens. The high prevalence of obesity observed in the CHAMACOS cohort represents an additional research question with relevance to studies of the microbiome. Here, we describe the breast milk and salivary microbiomes in a randomly selected subset of CHAMACOS mothers and their children, seek preliminary evidence of similarities in the microbiome between mother–child pairs, and explore whether prepregnancy or childhood obesity may be related to microbial community composition.

RESULTS

Subject Characteristics

All subjects were members of the CHAMACOS, a longitudinal birth cohort of low-income Mexican-American families (19) among which there is high prevalence of obesity in both mothers and children. Mothers (*N* = 10) included in this analysis were mostly young (mean age = 25.4 ± 3.4 y) at the time of pregnancy, and 8 were overweight or obese before pregnancy (27.3 ± 2.8 kg/m²) based on BMI >25 kg/m². One maternal sample was excluded from analysis of the breast milk microbiome after quality filtering of 16S rRNA gene sequencing results (see *Methods*). Saliva was collected from paired children (*N* = 10) 5 y later (mean age = 5.1 ± 0.2 y). All children were breastfed for at least 6 mo (mean duration = 14.4 ± 8.3 mo). Half of children were normal weight, and half were obese based on the sex- and age-specific BMI cutoff (95th percentile) provided by the 2000 Centers for Disease Control and Prevention child growth data. This pilot study included five female and five male children.

Bacterial Community Characteristics in Breast Milk and Child Saliva

We detected a total of 241 distinct operational taxonomic units (OTUs) in breast milk samples and 268 in child saliva samples (**Figure 1**). *Streptococcus*, *Staphylococcus*, and *Neisseria* were the most highly abundant phylogenetic groups overall, with the remaining groups each representing <10% of the bacterial communities observed in each sample type (**Table 1**). Members of the genus *Streptococcus* were by far the most abundant in both breast milk (mean relative abundance = 73.8%) and child saliva samples (60.4%); no statistically significant difference was noted between *Streptococcus* abundances in the two sample types. However, abundances of other highly prevalent groups differed between mothers and children. *Staphylococcus* was observed almost exclusively in breast milk at relatively high levels (10.9% in breast milk vs. 0% in saliva), while *Neisseria* was primarily found only in child saliva (8.8% in saliva vs. 0.1% in breast milk). *Prevotella* was also substantially more common in child saliva (6.0% in saliva vs. 0.5% in breast milk).

Diversity Analyses

We considered two possible methods for assessing community diversity in individual samples. We first ranked individuals according to the number of distinct OTUs present in each sample (**Figure 2**). On average, we observed fewer unique OTUs in breast milk samples than in child saliva samples (125.2 distinct OTUs in saliva samples vs. 81.9 in breast milk, *P* < 0.005). This ranking system demonstrated that communities that appear extremely uniform due to high abundance of *Streptococcus* (samples BM3, BM9, and BM7) may actually contain a greater number of unique taxa than communities that appear significantly more diverse (samples BM8, BM5, and S7).

Secondly, we ranked individuals according to the level of entropy (i.e., unpredictability, measured in bits, see *Methods*) present in their bacterial community data, which resulted in an ordering in which *Streptococcus*-dominated communities ranked the lowest (**Figure 2**). Again, breast milk samples tended to exhibit lower diversity than child saliva samples (2.2 bits in saliva vs. 1.4 bits in breast milk, *P* = 0.024), but ordering within mother and child groups changed substantially compared to ordering by number of OTUs detected. Both methods of estimating within-individual community diversity indicated that breast milk samples were less diverse than child saliva samples, but the entropy-based method is more likely to assign lower diversity scores to samples that are dominated by a single highly abundant taxon.

Comparisons to Other Populations

We observed high levels of *Streptococcus* spp. (73.8% in breast milk and 60.4% in saliva) in both mothers and children. To see whether these levels of *Streptococcus* were consistent with those observed in other populations, we compared average profiles in our subjects to those from eight other groups of subjects described in six published studies (**Table 2**; **Figure 3**). Average *Streptococcus* levels in saliva samples varied substantially between studies, from 11.9% in adult saliva samples from the Human Microbiome Project Consortium study (8,20) to 62.2% in saliva samples from infants in the Cephas *et al.* study (11). All three studies of infant or child saliva and three out of four studies of adult saliva reported communities where *Streptococcus* was the most abundant taxon. Diversity of the average profiles, measured as entropy, also varied considerably between studies on the salivary microbiome. Entropy of averaged community data was lowest in infant saliva in the Cephas *et al.* study (2.2 bits) and comparably low in child saliva in the CHAMACOS study (2.6 bits). The remaining studies of the salivary microbiome reported considerably higher levels of entropy (3.5 bits in Ling *et al.* (27) adults, 3.7 bits in Ling *et al.* children, 3.9 bits in Cephas *et al.* (11) adults, and 4.0 bits in Human Microbiome Project Consortium adults).

We identified another study with publicly available data on microbiome community composition in human breast milk. This study, performed by Cabrera-Rubio *et al.* (21), reported results that differed substantially from those observed in CHAMACOS mothers. *Streptococcus* was only a small component of breast milk (4.2%) in the Finnish Caucasian women enrolled in this study, whereas it was highly abundant in the Mexican-American women in the CHAMACOS study (73.4%). Conversely, *Lactobacillus* was the most abundant taxon (15.1%) reported in the Finnish study, but was nearly
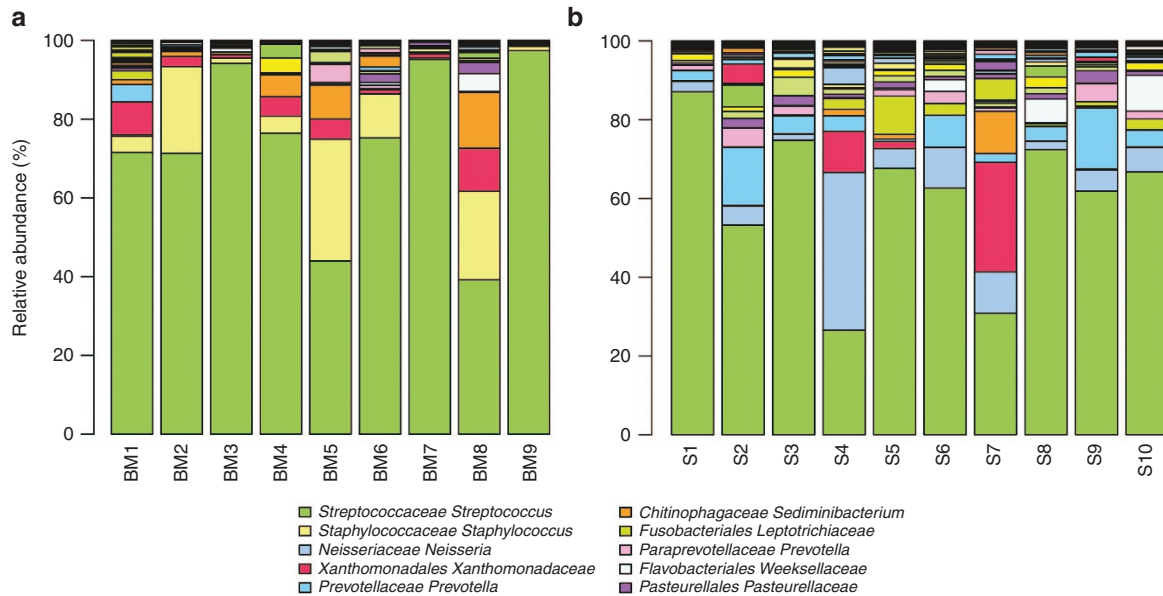
Figure 1. Relative abundances of bacterial taxa in CHAMACOS samples. (**a**) Relative abundances of taxa found in maternal milk samples. (**b**) Relative abundances of taxa found in child saliva samples. *Streptococcus* was the most abundant taxon in the majority of samples from both groups. CHAMACOS, Center for the Health Assessment of Mothers and Children of Salinas, CA.

Table 1. Relative abundances of common phylogenetic groups in breast milk and saliva

Taxon	Breast milk Mean ± SE (%)	Child saliva Mean ± SE (%)
Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	73.8 ± 14.7	60.4 ± 15.5
Firmicutes Bacilli Bacillales Staphylococcaceae Staphylococcus	10.9 ± 10.4	0.0 ± 0.6
Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae Neisseria	0.1 ± 0.8	8.8 ± 9.0
Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae	3.9 ± 6.4	4.1 ± 6.3
Bacteroidetes Bacteroidia Bacteroidales Prevotellaceae Prevotella	0.5 ± 2.4	6.0 ± 7.5
Bacteroidetes Saprospirae Saprospirales Chitinophagaceae Sediminibacterium	3.6 ± 6.2	1.5 ± 3.8
Fusobacteria Fusobacteriia Fusobacteriales Leptotrichiaceae	0.3 ± 1.7	2.0 ± 4.5
Bacteroidetes Bacteroidia Bacteroidales Paraprevotellaceae Prevotella	0.1 ± 1.1	2.1 ± 4.5
Bacteroidetes Flavobacteriia Flavobacteriales Weeksellaceae	0.2 ± 1.5	1.9 ± 4.3
Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae Other	0.3 ± 1.9	1.4 ± 3.8

undetected (0.008%) in CHAMACOS breast milk. Both entropy and number of taxa were also notably higher in the Finnish study than in the CHAMACOS study (5.2 bits and 421 OTUs in Cabrera-Rubio *et al.* (21) mothers vs. 1.7 bits and 241 OTUs in CHAMACOS mothers). Two other studies on the breast milk microbiome, performed by Hunt *et al.* (22)

and Jost *et al.* (23), reported that *Streptococcus* was the second most abundant genus (8.2 and 17.6%, respectively) after *Staphylococcus* (15.8 and 60.1%, respectively), but were not included in the comparison figure because full microbiome profile data were not available.

Similarities in relative abundances of major taxa among our subjects indicated that microbiome profiles of subjects within our study were more related to each other than to profiles reported by other studies. However, availability of individual data and differences in taxonomical resolution between studies precluded statistical comparisons of variability in different studies of the salivary or breast milk microbiome.

Relationships Between Obesity Parameters and Microbiome Characteristics in Mother–Child Pairs

Prior studies have demonstrated that differences in microbiome characteristics may be associated with maternal or child factors such as obesity (17,21,24,25). To assess whether maternal prepregnancy BMI and child BMI at age 5 were associated with microbiome parameters in breast milk or saliva, we explored correlations between these factors, microbial diversity, and relative abundances of major groups (**Figure 4**). Maternal prepregnancy BMI was correlated with lower relative abundance of *Streptococcus* ($r = -0.67$, $P = 0.048$) and higher diversity ($r = 0.77$, $P = 0.016$) of the breast milk microbiome, measured in bits. No statistically significant relationships were observed between child sex, breastfeeding duration, child BMI Z-score at age 5, or maternal prepregnancy BMI and any child salivary microbiome parameters.

In order to assess relationships between individuals, we computed a Jensen–Shannon distance matrix (26) between all samples and used principal components analysis to visualize
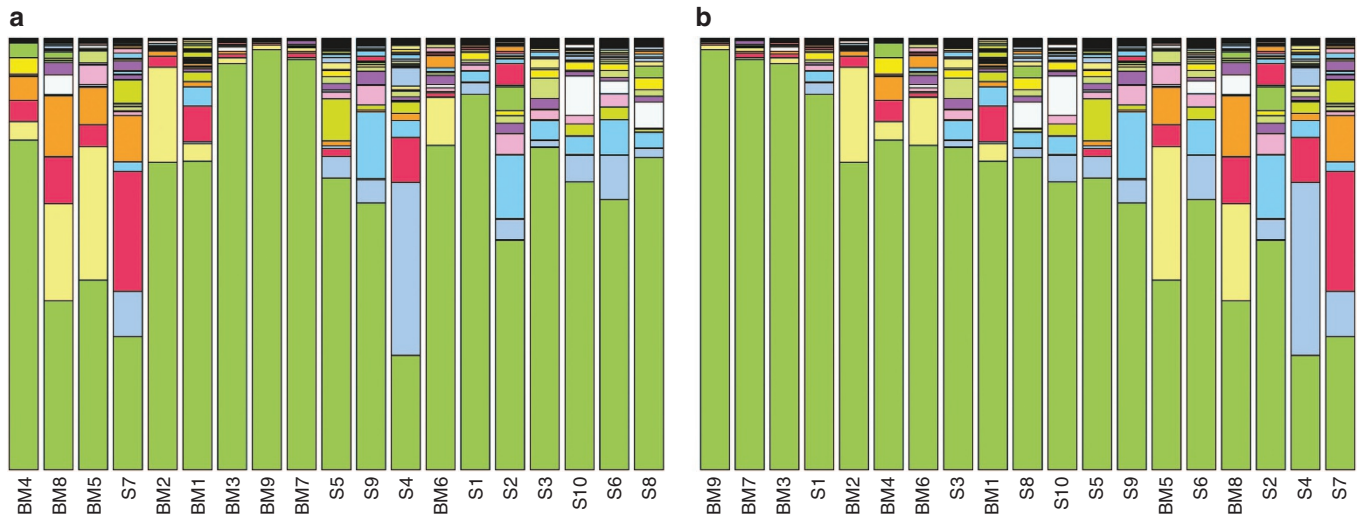
Figure 2. Ranking of samples based on diversity depends on choice of diversity metric. (**a**) Subjects ranked according to number of unique taxa with nonzero abundances. (**b**) Subjects ranked according to entropy, measured in bits. Both methods demonstrate that breast milk bacterial communities tended to be less diverse than communities in child saliva.

Table 2. Summary of comparison populations

Population description	Study group	N	Female (%)	Sample type	Age range	Methodology
CHAMACOS, Mexican-American, low-income, agriculture region in USA	Children	10	50%	Saliva, collected by Salivette swab	5.0 ± 0.1 y	Barcoded Illumina sequencing of V4 region
	Mothers	10	100%	Breast milk, self-collected near delivery	25.4 ± 3.4 y	
Cephas *et al.* (11), White or African-American, enrolled in Women, Infants, and Children program in USA	Infants	5	20%	Saliva, collected by sterile pipette	3–6 mo	Barcoded 454-pyrosequencing of V4–V6 region
	Adults	5	100%	Saliva, expectorated	30.8 ± 9.5 y	
Ling *et al.* (27), Chinese, boarding school residents in China	Children	10	50%	Saliva, expectorated	3–6 y	Barcoded 454-pyrosequencing of V3 region
	Adults	10	50%	Saliva, expectorated	22–24 y	
Human Microbiome Project Consortium, mostly White (77.3%), primarily born in USA or Canada	Adults	300	50%	Saliva, collected by sterile pipette	27 ± 5 y	Barcoded 454-pyrosequencing of V3–V5 region
Cabrera-Rubio *et al.* (21), Caucasian women in Finland	Mothers	18	100%	Breast milk, self-collected near delivery	32.0 ± 5.1 y	Barcoded 454-pyrosequencing of V1–V3 region
Hunt *et al.* (22), Healthy women in USA	Mothers	16	100%	Breast milk, collected by kit and pump postdelivery	20–40 y	Barcoded 454-pyrosequencing of V1–V2 region
Jost *et al.* (23), Healthy women in Switzerland	Mothers	7	100%	Breast milk, collected by pump postdelivery	Not given	Barcoded 454-pyrosequencing of V5–V6 region

differences (**Figure 5**). As expected, we observed a clear separation between maternal and child samples along the first principal component (PC-1), which is most likely attributable to differences in both sample matrix (breast milk vs. saliva) and subject age (adult vs. child). We also observed an equally clear division along the second principal component (PC-2), which separates four individuals from the other mothers and children in the study. Interestingly, two of the five normal-weight children in the study were among the group that separated along PC-2.

Using principal components analysis to assess relationships between nine mothers and their children, we observed no consistent connection between paired samples (**Figure 5**). We investigated whether any statistically significant correlations existed between mother and child *Streptococcus* abundance ($r = -0.32$, $P = 0.41$), *Staphylococcus* abundance ($r = -0.02$, $P = 0.97$), and within-individual entropy score ($r = -0.33$, $P = 0.38$), but observed no clear relationships. However, statistical power to assess these relationships was limited due to the sample size of this study.
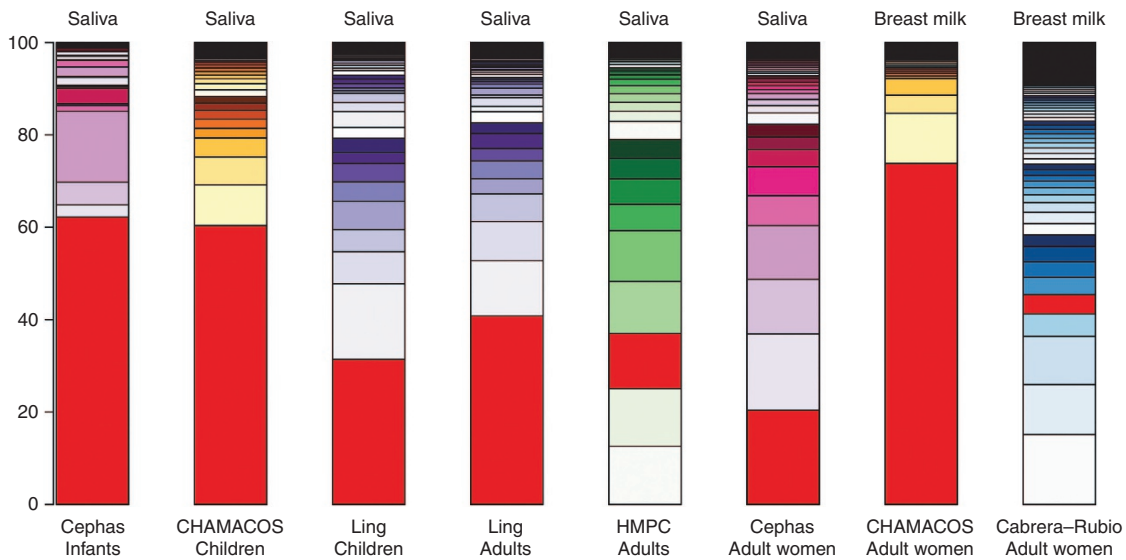
Figure 3. Average relative abundances across studies show unique patterns. Comparisons between studies showed that microbial diversity varied substantially between populations. *Streptococcus* relative abundance shown in red. Studies of child or infant saliva (CHAMACOS, Cephas *et al.* (11), Ling *et al.* (27)) tended to show lower levels of diversity and higher abundances of *Streptococcus* compared to studies of adult saliva. Microbial community composition in breast milk from Mexican-American CHAMACOS women differed substantially from breast milk results reported by Cabrera-Rubio *et al.* (21) in Caucasian women in Finland. CHAMACOS, Center for the Health Assessment of Mothers and Children of Salinas, CA. HMPC, Human Microbiome Project Consortium.
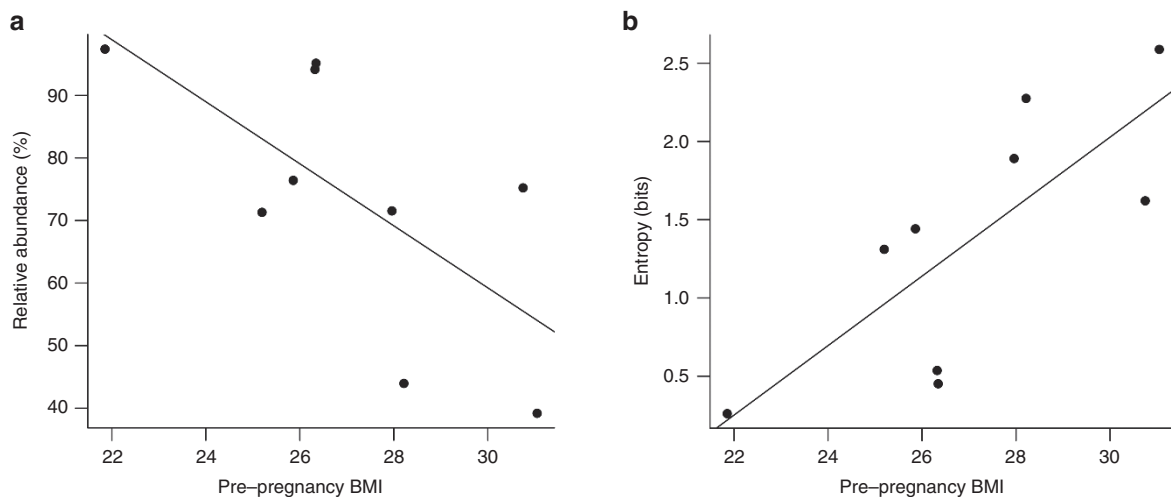

Figure 4. Maternal prepregnancy BMI is associated with breast milk microbiome parameters. (**a**) Correlation between maternal prepregnancy BMI and *Streptococcus* abundance ($r = -0.67$, $P = 0.048$). (**b**) Correlation between maternal prepregnancy BMI and diversity ($r = 0.77$, $P = 0.016$).

DISCUSSION

In this pilot study, we employed barcoded 16S rRNA gene sequencing to characterize the bacterial community composition of breast milk and saliva in mothers and their 5-year-old children, respectively. To our knowledge, our results represent the first study of the oral and breast milk bacterial communities in a low-income Mexican-American agricultural population. Through comparison with available data from several other cohorts, we illustrate significant differences in microbiome composition and diversity between geographically and socioeconomically disparate populations.

The very high abundance of *Streptococcus* spp. in both breast milk and child saliva samples was one of the most notable features of our results. In child saliva samples, this is a relatively unsurprising finding; many other studies have reported similarly high *Streptococcus* abundances in the oral cavity. The Human Microbiome Project Consortium, which has generated the largest collection of healthy adult microbiome community composition data to date, reported that *Prevotellaceae* was, on average, the most abundant family in saliva, but also noted that *Streptococcus* spp. dominated most oral habitats (8). Other 16S rRNA gene sequencing-based studies of the bacterial microbiome have also reported that *Streptococcus* is one of the most abundant taxa in saliva in both children and adults (9,11,27,28), though relative abundances still tended to be lower than those observed in CHAMACOS children. Interestingly, the salivary microbiome profiles of low-income CHAMACOS children most closely resembled those reported by Cephas *et al.* (11)
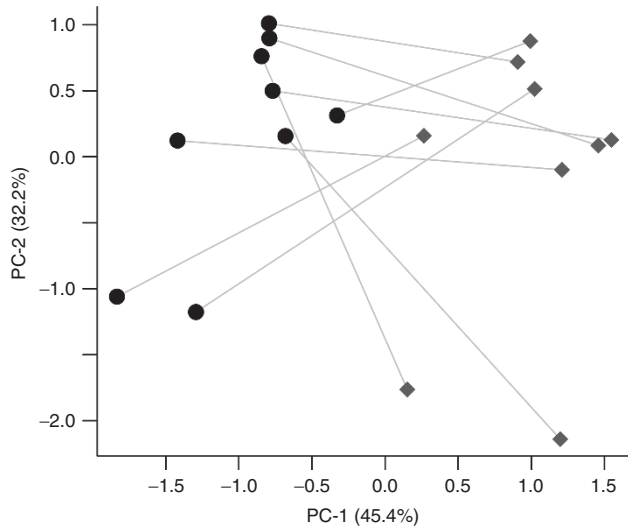
Figure 5. Principal component analysis shows clear separation between sample types. Black lines connect nine mother–child pairs. Maternal (black circles) and child (grey diamonds) samples show clear separation along the first principal component, which is likely attributable to differences between milk and salivary communities. The clear separation along the second PC, however, may be indicative of higher order structure and merits further study. Results are derived from the symmetric distance matrix obtained using Jensen–Shannon distance as a measure of dissimilarity between samples. PC, principal component.

among infants enrolled in the Women, Infants, and Children program, another low-income population.

The very high relative abundance of *Streptococcus* (39.1–97.3%) observed in breast milk from CHAMACOS mothers appears to be more unusual than that observed in child saliva. Although *Streptococcus* is known to be a common member of the human milk microbiome, its relative abundance appears to differ substantially between groups, between individuals, and even between samples from the same individual (15,21–23). In our study, *Streptococcus* was significantly more common than in other published metagenomic analyses of breast milk (21,22), and therefore community diversity in our samples was correspondingly lower. These dissimilarities in diversity and *Streptococcus* abundance may be attributable to differences among study populations. For example, the population of women enrolled in the CHAMACOS study is notably distinct from the mostly educated, Finnish Caucasian participants in the Cabrera-Rubio *et al.* (21,29) study in terms of geographical location, socioeconomic status, diet, and ethnicity; any of these factors may have contributed to the significant differences observed between these two studies of the human milk microbiome.

We also assessed whether maternal or child obesity parameters were associated with characteristics of the breast milk or salivary microbiomes. We observed strong correlations between prepregnancy BMI, decreased *Streptococcus* relative abundance, and increased microbial diversity in breast milk. It is unclear why prepregnancy BMI may affect diversity of the breast milk microbiome, but existing lines of evidence have demonstrated relationships between prepregnancy BMI and differences in the gut microbiome in infants (24) and young

children (25). It is possible that these effects may be mediated in part by BMI-related variation in the breast milk microbiome.

This pilot study has several limitations. For example, it is possible that the differences we observed between our study and other similar published analyses may be due to differences in sample collection. Additionally, the use of samples from only one time point, though common in many existing studies of the microbiome, may not provide an accurate representation of the average long-term microbiome composition, which has been shown to be variable in the short term (22). Because *Streptococcus* was so highly abundant in both sample types, another important question is whether species distribution within this genus is similar between mothers and children. Published studies based on both bacterial isolation and culture-independent methods indicate that numerous *Streptococcus* species, including *S. mitis*, *S. salivarius*, *S. oris*, *S. parasanguis*, *S. lactarius*, *S. australis*, *S. gallolyticus*, and *S. vestibularis*, can be found in breast milk (15), but even full-length 16S sequences often do not provide enough information to resolve species within the *Streptococcus* genus (5), representing an important disadvantage of 16S rRNA gene sequencing-based metagenomics. To gain further insight into the role these species play in saliva and breast milk, future studies of the human milk and oral microbiomes should aim to improve resolution within *Streptococcus* and other highly abundant genera.

The longitudinal nature of the CHAMACOS birth cohort study enabled us to simultaneously characterize a mother's breast milk microbiome composition at delivery and the oral microbiome of her child 5 y later. To our knowledge, this is the first study to assess the relationship between the breast milk and salivary microbiomes in mother–child pairs over a period of 5 y. Likely due to the relatively small sample size and limited statistical power of this pilot study combined with the length of time between sample collection, we were unable to observe any clear similarities between bacterial communities in breast milk and 5-year-old child saliva. While it is likely that environmental factors such as diet and dental hygiene play a significant role in defining the oral microbiome in 5-year-old children, existing evidence suggests that certain early life exposures may indeed affect long-term health outcomes such as obesity, most likely through modulation of the infant microbiome (17,18,24,30). Since breast milk consumption is known to affect infant microbiome composition (9,16) and breast milk is an early source of bacteria in the infant oral cavity (14), it is possible that the breast milk microbiome itself may influence the long-term composition of both the oral and gut bacterial communities in children. Further analyses that include additional CHAMACOS mother–child pairs may help to increase our understanding of this potentially important aspect of microbiome establishment during early childhood.

MATERIALS AND METHODS
Study Participants and Subject Selection
For this study, 10 mother–child pairs were randomly selected from the CHAMACOS longitudinal birth cohort, described previously (19). CHAMACOS mothers were enrolled during pregnancy in 1999–2000, and interviews and sample collection were conducted repeatedly for

the next 5 y. This pilot analysis used breast milk samples that were collected 2–4 days after delivery from CHAMACOS mothers who had initiated breastfeeding. Saliva samples were collected from their children 5 y later. All infants were breastfed for ≥6 mo, and only healthy infants delivered vaginally after full-term pregnancy were considered for inclusion due to the known effects of cesarean delivery on microbiome composition (10,21). Developmental assessments of children, including anthropometrics, were conducted at birth and at the time of each maternal interview. Signed informed consent was obtained from all mothers in the study. The Committee for the Protection of Human Subjects at the University of California, Berkeley, approved all study procedures. After sequencing data quality assessment and control, 9 mothers and 10 children remained in the analysis (see *Quality Filtering of Reads*).

Sample Collection
Breast milk samples were self-collected by CHAMACOS mothers using a breast milk pump, according to the manufacturer's instructions. Participants were advised to wash hands and the breast area with soap and water prior to collection. Milk was collected into a sterile breast milk collection bottle and care was taken to ensure the inside of the bottle did not come into contact with skin, clothes, or other external surfaces. Exposure of the sample to air was minimized. Breast milk samples were aliquotted and stored at −20 °C until shipment to the University of California, Berkeley School of Public Health Biorepository (Director: Holland), where they were transferred to long-term storage at −80 °C.

Saliva samples were collected from children at 5 y of age using Salivette saliva collection swabs (Sarstedt AG & Co, Nümbrecht, Germany). After the child rinsed his or her mouth with water, Salivette cotton plugs were gently tipped into the mouth to prevent contact with the examiner's or child's hands. Children were encouraged to hold the cotton plug in their mouths for up to five minutes to promote maximal saliva absorption and to obtain a representative oral sample, at which point they were instructed to spit the plug back into the Salivette tube. Salivette samples were stored at −20 °C until transfer to long-term storage at −80°C.

DNA Extraction and Purification
DNA was isolated from 1.5 ml of breast milk from each mother and from one Salivette saliva collection swab from each child. Before DNA isolation, Salivette cotton swabs were transferred to ClotSpin Baskets (Qiagen, Valencia, CA), thawed at room temperature, and centrifuged for 3 min at 2,095×*g* to withdraw saliva into the ClotSpin Basket reservoir. Recovered saliva volumes ranged from 0.2 ml to over 1.5 ml. For those samples with greater than 1.5 ml recovered volume, only 1.5 ml was used for DNA isolation. Samples were subjected to a mechanical "bead beating" pretreatment using Pathogen Lysis Tubes (Qiagen, Valencia, CA) to disrupt Gram-positive bacterial cell walls according to the manufacturer's protocol. Fat from breast milk samples was removed along with liquid supernatant after pelleting by centrifugation and before bead beating. DNA isolation from mechanically lysed breast milk and saliva samples was completed using QIAamp Ultraclean Production Pathogen Mini Kits (Qiagen, Valencia, CA) according to the manufacturer's protocol.

Amplification of 16S rRNA Genes
Amplification of the V4 region of the 16S rRNA gene was performed using PCR. A reverse primer (806R) containing a unique 12-nucleotide error-correcting Golay code for sample identification and a universal forward primer (515F) sequence were taken from a list of 2,167 barcoded primers published by Caporaso *et al.* (31). Forward and reverse primer sequences included the 5' and 3' reverse complement Illumina adapters, respectively, to allow sequencing on the HiSeq 2500 (Illumina, San Diego, CA). Isolated breast milk and saliva DNA samples were amplified using 5 PRIME HotMasterMix (5 PRIME, Hilden, Germany) as described previously (32). Reactions were performed in triplicates for each sample and consisted of 10 μl HotMasterMix, 0.5 μl each of 10 μmol/l 515F and 806R primers, 1 μl of template DNA, and 13 μl of nuclease-free water. Reaction conditions consisted of the following steps: a 3-min hold at 94 °C to denature template DNA, 35 cycles of 45 s at 94 °C, 1 min at 50 °C, and 1.5 min at 72 °C, and a final 10-min hold at 70 °C.

Triplicate reaction products were pooled and gel electrophoresis was used to verify presence of a band of the correct size (~300 bp). Amplicon pools were purified using UltraClean PCR Clean-Up Kits (MO BIO, Carlsbad, CA) to remove unincorporated primer and PCR contaminants, according to the manufacturer's protocol. Cleaned amplicon pools were quantified using Quant-iT PicoGreen dsDNA Assay Kits (Life Technologies, Carlsbad, CA) and concentrations were normalized before barcoded samples were pooled in equal amounts. The final pool of samples was purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA) and sent to the Vincent J. Coates Genomics Sequencing Laboratory (University of California, Berkeley, CA) for quality verification and sequencing via 2100 Bioanalyzer (Agilent, Santa Clara, CA) and HiSeq 2500, respectively.

Quality Filtering of Reads
An aggressive quality filtering strategy was adopted to remove low-quality reads. Raw reads were first filtered by the CASAVA 5.1 (Illumina, San Diego, CA) quality check using default settings. Reads which passed the CASAVA 5.1 filter were split by subject by the Quantitative Insights Into Microbial Ecology (QIIME) Version 1.8 open source software package (33). Demultiplexing was performed using the Quantitative Insights Into Microbial Ecology split_libraries_fastq.py script. Of the raw reads which passed the CASAVA 5.1 quality filter, 1,576,042 reads did not pass the secondary Quantitative Insights Into Microbial Ecology filter, which removes short and low-quality reads (34), resulting in a total of 5,259,716 quality-filtered reads (12,267–518,969 per subject). Reads which failed to pass all quality control steps either contained barcode errors or were too short after quality filtering.

Assignment of OTUs
We employed QIIME's subsampled open-reference OTU picking protocol (35) with default parameters to assign OTUs based on the Greengenes Version 13.8 database reference set (36,37). Chimeric sequences were checked against the reference database and then filtered. OTU assignments were made based on a minimum of 97% sequence identity, corresponding to genus-level similarity; any sequences that did not match the reference database were clustered *de novo*. Taxonomy was then assigned to the resulting OTUs using the UCLUST consensus taxonomy assigner (38). At this step, any reads that failed to align using PyNAST (36) were excluded and any taxa with fewer than two reads across all samples were removed from the final OTU table. 2,061,324 reads remained in the analysis (3,439–248,421 per subject). One maternal breast milk sample (BM10) was dropped from the dataset due to low read count (<12,000 reads after alignment), which may have been indicative of poor PCR performance. Among the remaining subjects, we attained an average of $108,310 \pm 72,766$ reads per sample. Because read counts were sufficiently large for all subjects that passed quality filtering steps, rarefaction was not deemed necessary (39).

Data Analyses
In order to estimate within-sample diversity, we utilized an entropy-like measurement, expressed in bits, that calculates the negative average $\log_2$ relative abundance across all present taxa (40), thus providing an estimate of the unpredictability of each subject's community profile data. The diversity of the *j*th sample is calculated as follows:

$$D_j = -\sum_{i=1}^{n_j} p_{i,j} \log_2 p_{i,j}$$

where n_j denotes the number of taxa present in the *j*th sample and $p_{i,j}$ denotes the relative abundance of the *i*th taxon in the *j*th sample. All downstream analyses related to within-sample diversity utilized this metric, unless otherwise noted. Similar measures have been used as metrics of community diversity in previous studies of the microbiome (6).

For further analyses, it was useful to define a distance metric between pairs of samples for applications such as principal components analysis. We used the Jensen–Shannon distance (26), which calculates the distance between two samples as follows:

$$D_{i,j} = \sqrt{\frac{1}{2}D_{KL}(i,j) + \frac{1}{2}D_{KL}(j,i)}$$

where $D_{KL}(i,j)$ denotes the Kullback–Leibler divergence of sample j from sample i, restricted to the set of taxa found to be present in both samples. With the resulting matrix of Jensen–Shannon distances, we used principal components analysis to visualize the differences between samples. Student's *t*-tests were used to assess differences in entropy and natural log-transformed OTU counts for the most abundant taxa. Pearson correlation was used to assess whether statistically significant relationships existed between mother and child *Streptococcus* and *Staphylococcus* abundance, mother and child diversity estimates, and mother and child obesity metrics.

To compare our results to those from other studies, we accessed available data from several recent publications on the human salivary or breast milk microbiome (8,11,21–23,27). These studies were selected based on several criteria, including similarity with our study in terms of age and sample type, use of a 16S rRNA gene sequencing-based methodology to assess whole bacterial community composition, and data availability. Where possible, we calculated an average profile based on the mean relative abundances of all identified taxa for each group of subjects. When data from more than one time point were reported, we selected the time point most comparable to those used in our study. Differences in resolution or naming conventions between datasets precluded statistical comparisons between our data and other published results. Therefore, we limited our analysis to visual and numerical comparisons and focused on the *Streptococcus* genus, which was highly abundant in our samples. All analyses were conducted using the statistical software package R Version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria).

REFERENCES
1. Grice EA, Segre JA. The human microbiome: our second genome. Annu Rev Genomics Hum Genet 2012;13:151–70.
2. Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of ʟ-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 2013;19:576–85.
3. de Sousa Moraes LF, Grzeskowiak LM, de Sales Teixeira TF, Gouveia Peluzio Mdo C. Intestinal microbiota and probiotics in celiac disease. Clin Microbiol Rev 2014;27:482–9.
4. Ley RE. Obesity and the human microbiome. Curr Opin Gastroenterol 2010;26:5–11.
5. Wade WG. The oral microbiome in health and disease. Pharmacol Res 2013;69:137–43.
6. Docktor MJ, Paster BJ, Abramowicz S, et al. Alterations in diversity of the oral microbiome in pediatric inflammatory bowel disease. Inflamm Bowel Dis 2012;18:935–42.
7. Groer MW, Luciano AA, Dishaw LJ, Ashmeade TL, Miller E, Gilbert JA. Development of the preterm infant gut microbiome: a research priority. Microbiome 2014;2:38.
8. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;486:207–14.
9. Zaura E, Nicu EA, Krom BP, Keijser BJ. Acquiring and maintaining a normal oral microbiome: current perspective. Front Cell Infect Microbiol 2014;4:85.
10. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci USA 2010;107:11971–5.
11. Cephas KD, Kim J, Mathai RA, et al. Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. PLoS One 2011;6:e23503.
12. Douglass JM, Li Y, Tinanoff N. Association of mutans streptococci between caregivers and their children. Pediatr Dent 2008;30:375–87.
13. Li Y, Ismail AI, Ge Y, Tellez M, Sohn W. Similarity of bacterial populations in saliva from African-American mother–child dyads. J Clin Microbiol 2007;45:3082–5.
14. Sampaio-Maia B, Monteiro-Silva F. Acquisition and maturation of oral microbiome throughout childhood: an update. Dent Res J (Isfahan) 2014;11:291–301.
15. Fernández L, Langa S, Martín V, et al. The human milk microbiota: origin and potential roles in health and disease. Pharmacol Res 2013;69:1–10.
16. Holgerson PL, Vestman NR, Claesson R, et al. Oral microbial profile discriminates breast-fed from formula-fed infants. J Pediatr Gastroenterol Nutr 2013;56:127–36.
17. Ajslev TA, Andersen CS, Gamborg M, Sørensen TI, Jess T. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. Int J Obes (Lond) 2011;35:522–9.
18. Kalliomäki M, Collado MC, Salminen S, Isolauri E. Early differences in fecal microbiota composition in children may predict overweight. Am J Clin Nutr 2008;87:534–8.
19. Eskenazi B, Harley K, Bradman A, et al. Association of in utero organophosphate pesticide exposure and fetal growth and length of gestation in an agricultural population. Environ Health Perspect 2004;112:1116–24.
20. Aagaard K, Petrosino J, Keitel W, et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. FASEB J 2013;27:1012–22.
21. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr 2012;96:544–51.
22. Hunt KM, Foster JA, Forney LJ, et al. Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS One 2011;6:e21313.
23. Jost T, Lacroix C, Braegger C, Chassard C. Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. Br J Nutr 2013;110:1253–62.
24. Collado MC, Isolauri E, Laitinen K, Salminen S. Effect of mother's weight on infant's microbiota acquisition, composition, and activity during early infancy: a prospective follow-up study initiated in early pregnancy. Am J Clin Nutr 2010;92:1023–30.
25. Galley JD, Bailey M, Kamp Dush C, Schoppe-Sullivan S, Christian LM. Maternal obesity is associated with alterations in the gut microbiome in toddlers. PLoS One 2014;9:e113026.
26. Lin J. Divergence measures based on the Shannon entropy. IEEE Trans Inform Theory 1991;37:145–51.
27. Ling Z, Liu X, Wang Y, Li L, Xiang C. Pyrosequencing analysis of the salivary microbiota of healthy Chinese children and adults. Microb Ecol 2013;65:487–95.
28. Zaura E, Keijser BJ, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral microbial communities. BMC Microbiol 2009;9:259.
29. Laitinen K, Poussa T, Isolauri E; Nutrition, Allergy, Mucosal Immunology and Intestinal Microbiota Group. Probiotics and dietary counselling

contribute to glucose regulation during and after pregnancy: a randomised controlled trial. Br J Nutr 2009;101:1679–87.

30. Dogra S, Sakwinska O, Soh SE, et al.; GUSTO Study Group. Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. mBio 2015;6:e02419–14.

31. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 2012;6:1621–4.

32. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 2011;108:Suppl 1:4516–22.

33. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–6.

34. Bokulich NA, Subramanian S, Faith JJ, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 2013;10:57–9.

35. Rideout JR, He Y, Navas-Molina JA, et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ 2014;2:e545.

36. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 2010;26:266–7.

37. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069–72.

38. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460–1.

39. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol 2014;10:e1003531.

40. Hill MO. Diversity and evenness: a unifying notation and its consequences. Ecology 1973;54:427–32.