

SCIENTIFIC REPORTS



OPEN

Deep sequencing of the 16S ribosomal RNA of the neonatal oral microbiome: a comparison of breast-fed and formula-fed infants

Received: 15 March 2016
Accepted: 08 November 2016
Published: 06 December 2016

S. S. Al-Shehri^{1,2}, E. L. Sweeney³, D. M. Cowley⁴, H. G. Liley⁴, P. D. Ranasinghe⁵, B. G. Charles¹, P. N. Shaw¹, D. Vagenas³, J. A. Duley^{1,4,*} & C. L. Knox^{3,*}

In utero and upon delivery, neonates are exposed to a wide array of microorganisms from various sources, including maternal bacteria. Prior studies have proposed that the mode of feeding shapes the gut microbiota and, subsequently the child's health. However, the effect of the mode of feeding and its influence on the development of the neonatal oral microbiota in early infancy has not yet been reported. The aim of this study was to compare the oral microbiota of healthy infants that were exclusively breast-fed or formula-fed using 16S-rRNA gene sequencing. We demonstrated that the oral bacterial communities were dominated by the phylum *Firmicutes*, in both groups. There was a higher prevalence of the phylum *Bacteroidetes* in the mouths of formula-fed infants than in breast-fed infants ($p = 0.01$), but in contrast *Actinobacteria* were more prevalent in breast-fed babies; *Proteobacteria* was more prevalent in saliva of breast-fed babies than in formula-fed neonates ($p = 0.04$). We also found evidence suggesting that the oral microbiota composition changed over time, particularly *Streptococcus* species, which had an increasing trend between 4–8 weeks in both groups. This study findings confirmed that the mode of feeding influences the development of oral microbiota, and this may have implications for long-term human health.

The fetus is first exposed to bacteria *in utero*¹, and following delivery is colonized with diverse microorganisms at different bodily sites (termed the microbiota). This microbial colonization initiates maturation of the infant immune system and alterations to the microbiota (dysbiosis) may result in illness or increased risk of infection, particularly during infancy. Within the first few hours of life, the most important sources of neonatal microbiota are derived from the mother's vaginal, fecal and skin microbiota^{2,3}. The presence of these microorganisms, and those from different environmental sources are likely to contribute to the development and regulation of the oral bacterial microbiota in the first few months of life⁴.

Within the neonatal oral cavity, microorganisms may be regulated by other mechanisms. In a recent study, our group discovered that neonatal saliva contains high levels of the metabolites xanthine and hypoxanthine; these are substrates of the enzyme xanthine oxidase (XO), which is highly abundant within breast milk^{5,6}. During breast-feeding, the mixing of neonatal saliva with breast milk generates hydrogen peroxide, a reactive oxygen species (ROS) that in turn activates the 'lactoperoxidase system' to produce additional bactericidal ROS, as well as hypothiocyanate and nitrite. These metabolites provide a unique antibacterial activity within the neonatal mouth at a time when other immune mechanisms are not yet fully developed. Furthermore, we demonstrated that some pathogenic bacteria, including *Staphylococcus aureus* and *Salmonella* species, were preferentially inhibited by the production of ROS during *in vitro* experiments when simulated neonatal saliva and expressed human breast milk were combined with these bacteria⁶.

¹School of Pharmacy, The University of Queensland, St Lucia, 4102, Australia. ²School of Applied Medical Sciences, Taif University, Taif, 21974, Saudi Arabia. ³The Institute of Health and Biomedical Innovation, Faculty of Health, School of Biomedical Sciences, Queensland University of Technology, Brisbane, 4059, Australia. ⁴Mater Research Institute, The University of Queensland, Woolloongabba, 4102, Australia. ⁵Queensland University of Technology, Science and Engineering Faculty, School of Earth, Environmental and Biological Sciences, 4001, Australia. ⁶These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.S.A.-S. (email: sa.alshehri@tu.edu.sa)

	Breast-fed			Formula-fed			p value
	Male	Female	Total	Male	Female	Total	
	12	14	26	5	7	12	
Gestational age (w) ¹							
Mean ± SD	40 ± 0.5	40 ± 1	40 ± 1	40 ± 2	39 ± 2	40 ± 2	0.3
Range	39–41	38–42		38–41	37–41		
Birth weight (g)							
Mean ± SD	3488 ± 316	3584 ± 570	3539 ± 464	3916 ± 599	3664 ± 227	3779 ± 432	0.2
Range	3054–4110	2700–4492		3182–4720	3490–4092		
Age (w) at collection ²							
Mean ± SD	6 ± 1	6 ± 2	6 ± 2	7 ± 1	6 ± 2	6 ± 1	0.2
Range	4–8	4–8		6–8	4–8		

Table 1. A comparison of the gestational age, birth weight, and age at the time of oral specimen collection, for breast-fed and formula-fed infants. ¹w = weeks, at the time of delivery; term >37 weeks gestation. ²w = weeks, at the time of collection of oral specimen.

Breast milk has long been considered a superior food for infants, particularly as breast-feeding has been associated with a reduced incidence of infection^{7,8}. Formula feeding has been shown to induce the most rapid weight gain in preterm infants, while breast milk has a lower protein and calorie content than formula milk it is rich in immunoglobulins, enzymes and growth factors. Recent research has demonstrated that neonatal immune mechanisms may be inadequate for regulating the gastrointestinal tract (GIT) microbiota, in particular with formula-fed babies who are at an increased risk of developing necrotising enterocolitis - the most common and lethal GIT emergency in preterm infants⁹. In contrast, preterm babies who are fed breast milk are three times less likely to develop necrotising enterocolitis than formula-fed babies^{10,11}. Infants thus appear to rely on 'innate' immune mechanisms, which may be present in breast milk, to regulate their GIT microbiota until the development of the more mature cellular and antibody-mediated immune system present in children and adults¹².

Previous studies have investigated infant microbiota, in particular the development of the infant GIT microbiota, demonstrating key differences depending on the mode of delivery^{4,13,14}. More recently, differences in the oral microbiota of three-month-old infants were reported depending on the mode of feeding, by Holgerson *et al.*¹⁵, who found that *Lactobacillus* species were isolated from breast-fed infants but not from formula-fed infants, and there were also major differences in 14 taxa (detected by HOMIM microarray) depending on the mode of feeding¹⁵. Another smaller study of six infants mapped the changes in the distribution of microorganisms within the skin, oral and GIT microbiota between 8–21 days of life¹⁶. However, the infants enrolled in this study were low-birth weight, predominantly preterm and all of these infants were treated with antibiotics during the study period. As a result, it is not known if these findings are consistent with the development of the microbiota within a healthy, term-born population. Whilst these two studies have extended our knowledge and understanding of the infant microbiota, it remains unclear how the microbiota of the oral cavity alters over time in healthy infants in the early neonatal period.

Studies of the human microbiota have not been possible without the use of next-generation sequencing (NGS). This useful tool has allowed us to better understand the relationship between microbiota and its association to human health and disease^{17,18}. The results obtained by molecular techniques, including NGS, are far superior to that of culture-based technologies and a large number of samples may be tested within a single assay¹⁹. For this current study we hypothesised that the oral microbiota of infants varies depending on the mode of feeding (i.e. breast-feeding *versus* formula-feeding) and with age. We utilised 16S rRNA gene deep sequencing to characterize the oral microbiota of healthy 4–8 week-old infants who were exclusively breast-fed or formula-fed.

Results

Participants for *in vivo* studies of oral bacteria. Among the 38 neonates recruited (delivered vaginally; 26 breast-fed *versus* 12 formula-fed), there were no significant differences in gestational age ($p = 0.3$), birth weight ($p = 0.2$), or age ($p = 0.2$) (Table 1). The mothers and infants enrolled in this study did not receive any antibiotics prior to or at the time of sample collection.

Overview of 16S rRNA sequencing on breast-fed and formula-fed infants: sequences, quality analyses and taxa identified. The microbiota swabbed from the inner cheek surfaces of breast-fed and formula-fed infants (age: 4 or 6 or 8 weeks), yielded 496,303 curated high quality reads of 250–550 bp with an average length of 492.97 ± 8.33 bp (see Supplementary Info Table 1).

Following removal of low quality sequences, which comprised 11.5% of the original dataset, the bacterial richness was assessed as the Shannon index, an indicator of mean richness in bacterial diversity. There were similarities among breast-fed and formula-fed infants, with formula-fed infants demonstrating a slightly higher (mean) diversity within their oral microbiota (Fig. 1).

Oral bacterial community structure revealed differences in the relative abundance of bacteria phyla among breast-fed and formula-fed infants. Analyses of operational taxonomic units (OTUs) from the oral cavity of breast-fed and formula-fed infants clustered within four main bacterial phyla: *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. The *Firmicutes* represented the most predominant

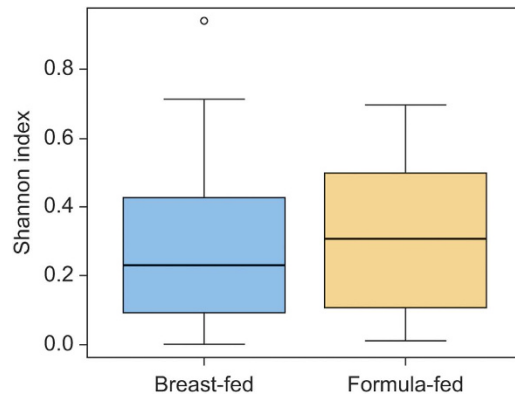


Figure 1. Oral bacterial community richness analyses. Alpha diversity measurements using a Shannon index analysis indicated samples with higher and lower Genus diversity in the mouths of breast-fed and formula-fed neonates. The depth of sequence reads was drawn against the richness of OTU clusters at different taxonomic assignments.

phylum in the oral cavity of both breast-fed and formula-fed infants, accounting for (mean \pm SD) $96.3\% \pm 0.04$ of the OTUs in breast-fed infants and $95.3\% \pm 0.05$ in formula-fed infants ($p = 0.5$) (Fig. 2a).

The relative distributions of the remaining bacterial phyla were re-analyzed after excluding the *Firmicutes*, to highlight the prevalence of the other bacterial communities present (Fig. 2b). These analyses revealed that some of the phylum-level distribution patterns were significantly different between the mouths of breast-fed and formula-fed babies. The phylum *Actinobacteria* was the second most abundant in breast-fed neonates, representing $2.1\% \pm 0.03$ of the OTUs in oral samples, versus $1.6\% \pm 0.03$ in formula-fed samples ($p = 0.8$). In contrast, the phylum *Bacteroidetes* was the next most abundant in formula-fed infants ($1.8\% \pm 0.6$ of the OTUs) compared with breast-fed neonates ($0.2\% \pm 0.1$ of the OTUs) ($p = 0.01$) (Fig. 2c). For breast-fed neonates, the phylum *Proteobacteria* was also abundant, but the prevalence of this phylum was very low in formula-fed neonates ($p = 0.04$).

Furthermore, the bacteria collected from the oral cavity of breast-fed and formula-fed infants were analysed for genus-level mean distributions of the three major phyla within all samples: results are shown as box plots in Fig. 3. The *Streptococcus* spp. (Phylum *Firmicutes*) was the predominant genus identified in the oral cavities of both breast-fed and formula-fed infants. Less than 1% of these reads could not be classified to the genus level (shown as family “*Streptococcaceae_unclassified*”; (Fig. 3a), which may have been as a consequence of genetic similarities in the gene portion amplified within this study. There was no significant difference in *Streptococcus* spp. prevalence between breast- and bottle-fed babies.

There were no differences in the prevalence of the genera *Lactobacillus* (*Firmicutes*), *Rothia* (*Actinobacteria*) and *Veillonella* (*Firmicutes*) in the oral cavity of breast-fed and formula-fed infants (Fig. 3b). However, *Prevotella* spp. (Phylum *Bacteroidetes*) were significantly more abundant in the oral cavity of formula-fed infants ($p = 0.02$), when compared to breast-fed infants (Fig. 3c). The relative abundance of the bacteria shown in Fig. 3 were further plotted using a Heat Map (Fig. 4), which demonstrates the higher abundance of *Prevotella* spp. within the oral cavity of formula-fed infants when compared to breast-fed infants.

Interestingly, the abundance of both *Streptococcus* spp. and ‘*Streptococcaceae_unclassified*’ for each collection time cohort (4, 6, 8 weeks for both breast-fed and formula-fed infants) demonstrated apparent changes over time (Fig. 5). The analysis showed that while there was a trend of increasing relative abundance of *Streptococcus* spp. over time, there was a corresponding decrease in the relative abundance of unclassified Family *Streptococcaceae*.

For those reads identified as ‘Bacteria_others’, we further investigated the closest known identities by comparing these sequence data against databases of known bacterial sequences. Figure 6 shows the accession numbers and the sources of the most closely related sequences that we were able to extract, from <http://www.ncbi.nlm.nih.gov/gene/>. We found that the majority of sequences of the ‘Bacteria_others’ were most closely related (10–18% homology) to bacterial sequences originally isolated from either the skin surfaces on the human arm (47%), the human mouth (40%), the human lower respiratory tract (7%), or the human upper respiratory tract (2%). Only one sequence was related closely to organisms isolated from the GIT, while four sequences were environmental isolates. Furthermore the sequences of ‘Bacteria_others’ were most closely related to the Phylum *Firmicutes* (Class *Bacilli* 45%; *Negativecutes* 5%), Phylum *Actinobacteria* (35%), Phylum *Bacteroidetes* (7.5%), Phylum *Fusobacteria* (3.75%) and Phylum *Proteobacteria* (3.75%).

Discussion

An important priority for studying the human microbiome is to improve our knowledge and understanding of the bacteria implicated in human health and disease states. The microbiota has been suggested as a potential target for therapeutic interventions for a range of diseases^{20,21}. While it is known that the oral microbiome of adults is relatively stable and may harbour up to 700 different bacterial species²², it is likely that the infant oral microbiome is more dynamic and may be influenced by a range of factors, including diet and the infant’s developing immune competency. Our current research investigated the composition of the oral bacterial microbiota in breast-fed and

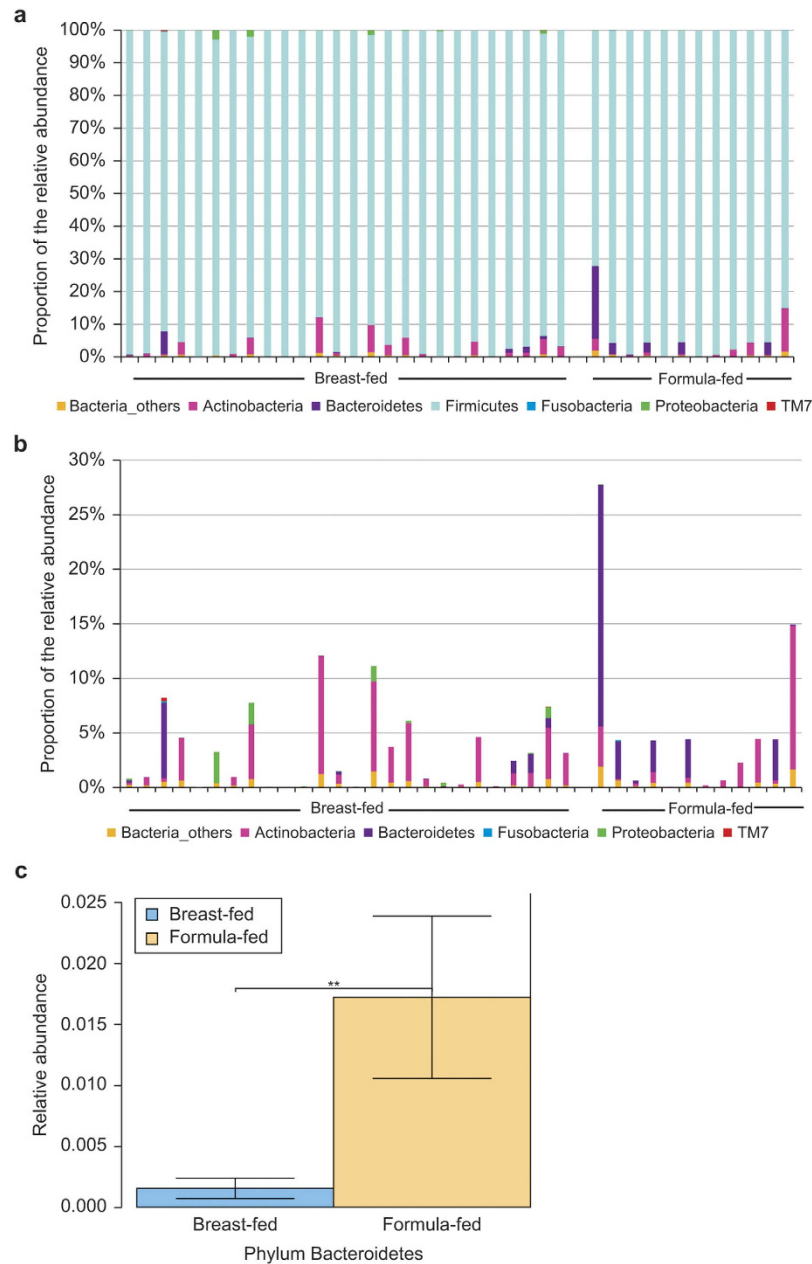


Figure 2. Proportion of the relative abundance of the major oral bacterial phyla. Taxa were detected by 16S rRNA analysis using Roche-454 GS FLX deep sequencing. Each column within the bar graph represents the percentages of the bacterial OTUs detected within each infant oral swab sample. **(a)** Bar graph showing the proportion of the relative abundance (%) of the major oral phyla operational taxonomic units (OTUs) for breast-fed samples ($n = 26$) and formula-fed samples ($n = 12$). **(b)** Bar graph showing the relative abundance (%) of oral phyla excluding the *Firmicutes*, of both breast-fed and formula-fed oral samples. There were key differences in the abundance of some phyla among breast-fed and formula-fed infants. **(c)** Relative abundance of the Phylum *Bacteroidetes* in oral swab samples from breast-fed and formula-fed infants (** $p = 0.01$).

formula-fed infants at 4–8 weeks of life, using 16S rRNA gene deep sequencing. We demonstrated distinct differences in the oral bacterial communities associated with breast-feeding and formula-feeding.

The most abundant bacterial phylum within the oral cavity of both breast-fed and formula-fed infants was the *Firmicutes*, and this was consistent with reports of others^{14,23}. Interestingly, *Actinobacteria* was the second most prevalent oral phylum in breast-fed babies, while the phylum *Bacteroidetes*, which includes the genus *Bacteroides*, was the second most prevalent in the mouths of formula-fed neonates but was infrequently detected in infants who were breast-fed. These findings were consistent with previous studies of infant GIT microbiota, in which *Bacteroidetes* was the phylum most commonly found in association with formula-feeding⁴ and these microorganisms were also highly abundant within the faeces of formula-fed infants²⁴. In contrast, the phylum *Proteobacteria* (which include the ubiquitous GIT bacteria *Escherichia coli*) was more abundant in the mouths of

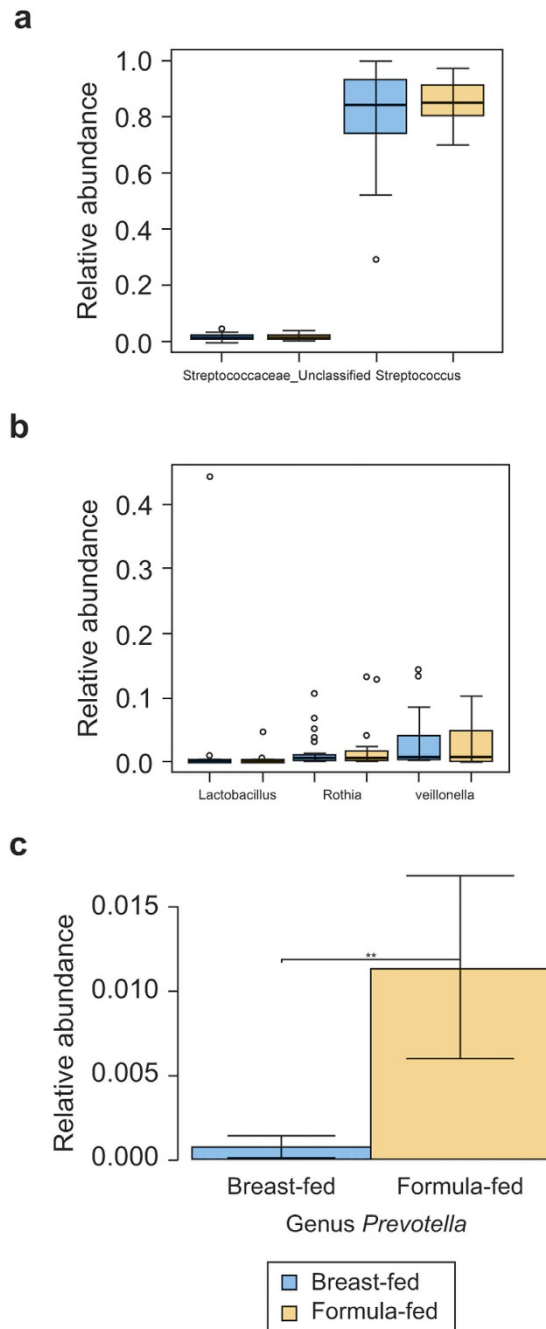


Figure 3. Relative abundance of 16S rRNA Genus-level sequences. (a) Genus level sequences were detected in oral swabs from breast-fed (red) and formula-fed (blue) infants using deep sequencing. Values are expressed as box plot showing the mean, percentiles, and SD. (b) The relative abundance of sequences that were classified only to the Family level ‘*Streptococcaceae_unclassified*’, compared to known *Streptococcus* spp. sequences. (c) The relative abundance of the major Genera sequences detected in the oral samples: *Lactobacillus* spp. (Phylum *Firmicutes*), *Rothia* spp. (Phylum *Actinobacteria*) and *Veillonella* spp. (Phylum *Firmicutes*). (d) *Prevotella* spp. (Phylum *Bacteroidetes*) sequences were more abundant the oral samples from formula-fed compared to breast-fed infants ($p = 0.02$).

breast-fed infants when compared to formula-fed infants. The *Proteobacteria* are common bacteria in the GIT of neonates²⁵ and reported as the predominant phylum (65%) in human breast milk²⁴. A recent study has suggested the existence of an entero-mammary pathway for the transfer of these bacteria from the GIT of the mother to the neonate²⁶. We suggest therefore, that the breast-fed neonates in our current study may have acquired such *Proteobacteria* from breast-feeding; this hypothesis is further strengthened by the very low abundance of these bacteria within the oral cavity of formula-fed infants.

We also found that the genus *Prevotella* was significantly more abundant in the mouths of formula-fed infants. *Prevotella* spp. are part of the normal oral and vaginal microflora; an association has been demonstrated

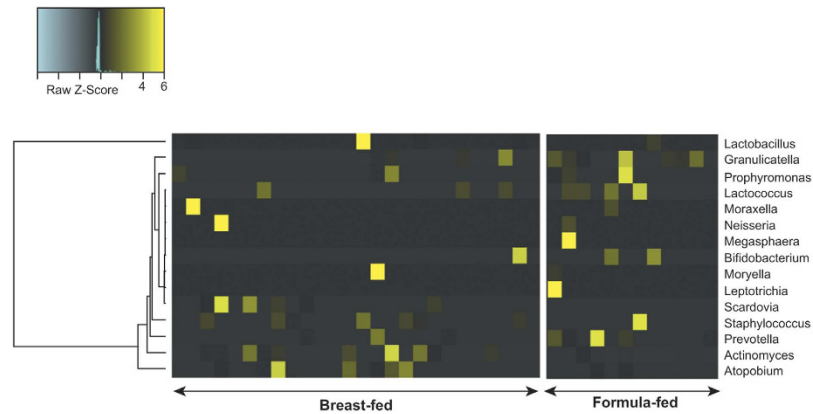


Figure 4. Heat map of oral bacterial community compositions within breast-fed and formula-fed infants and their hierarchical clustering. The 16S rRNA gene sequencing revealed differences in the abundance of oral bacterial communities. The bacterial genera are represented as heat maps corresponding to a phylogenetic tree. The heat map demonstrates the higher abundance of *Prevotella* spp. within the oral samples from breast-fed compared to formula-fed infants.

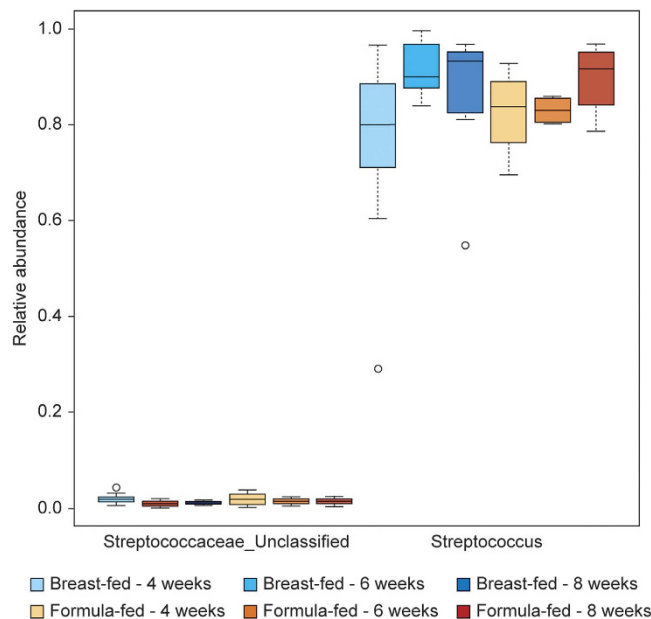


Figure 5. Temporal changes in streptococcal relative abundance in oral samples from breast-fed and formula-fed infants. The relative abundance of oral *Streptococcus* Genus-level sequences increased from weeks 4 to 8, for both breast-fed and formula-fed infants. Paralleling this increase, there was a decrease in the number of Genus-level sequences that could not be classified (*Streptococcaceae_unclassified*) from 4, 6 and 8 weeks of life. Each box plot corresponds to a sample time point and the mode of infant feeding versus the abundance of the OTUs. Box plots showing the mean, percentiles and SD.

between the presence of *Prevotella* spp. in the mouths of mothers and the isolation of these microorganisms in their infants' saliva²⁷. Subsequent studies identified unique strains of *Prevotella* spp. within the mouths of infants that did not match the isolates obtained from the mother's saliva²⁸. This suggests that there may be alternative reservoirs or influences allowing these microorganisms to colonise the infant mouth: the results of our study and those of Kononen *et al.*, suggested that the mode of feeding may influence the colonisation of the oral cavity with *Prevotella* spp.

While studies have demonstrated an association between infant oral and GIT microbiota in relation to mode of delivery^{13,23}, very few studies have investigated the infant oral microbiome in relation to the mode of feeding. Holgerson *et al.*¹⁵ studied 73 infants at three months of age and reported differences within the oral microbiota of breast-fed versus formula-fed infants who were exclusively/partially breast-fed, or formula-fed¹⁵. Their results and the findings of our current study provide further confirmation that the mode of feeding affects the oral microbiota of infants from the early neonatal period (4 weeks of age) and up to 3 months of age. Human breast

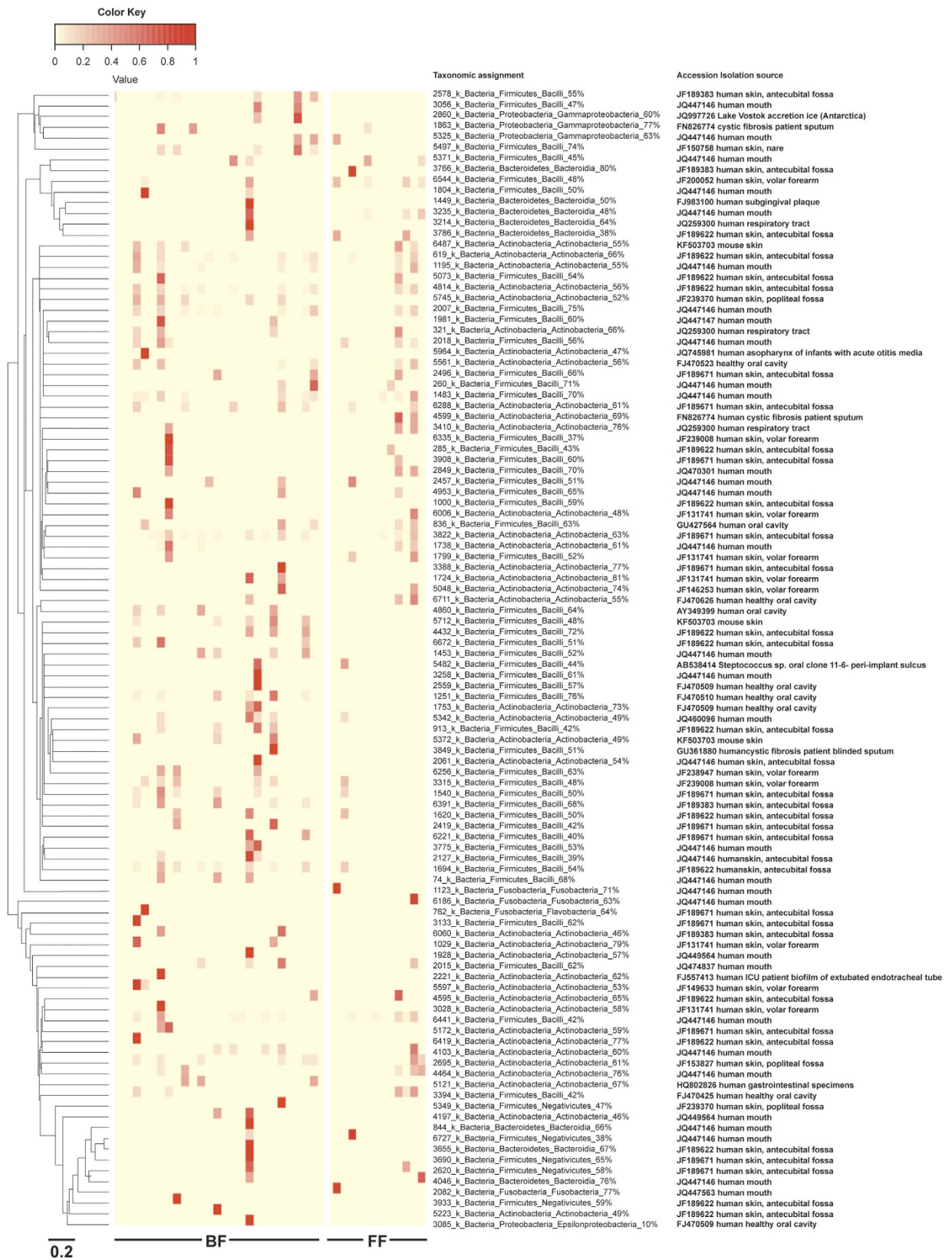


Figure 6. Heat map of sequences identified as ‘Bacteria_others’ within oral samples from infants. The heat map demonstrates the abundance of ‘Bacteria_others’ sequences and the taxonomic identities (phylogeny). The identities were explored using an RDP naïve Bayesian classifier and are shown at the Class level. The percentage values (taxonomic assignment) indicate the homology to bacteria classified to the Class level that have been isolated/identified previously, and the accession isolation source indicates the sites from which these closely-related bacterial DNA sequences have been previously isolated.

milk also has its own unique microbiome²⁹, so breast-feeding of infants may assist in the transmission of ‘healthy’ microflora from mothers to their infants. However, it is also known that pathogenic viruses (human immunodeficiency virus, Zika virus, cytomegalovirus) and bacteria (*Candida* spp., *Klebsiella* spp., *Acinetobacter* spp. and *Staphylococcus* spp.) may be acquired from an infected mother during breast feeding, predisposing to dysbiosis and disease^{30–32}.

Saliva and breast milk contain growth-enhancing and antibacterial factors, including amino acids, organic acids, proteins, fats, carbohydrates, enzymes, and immunoglobulins. Thus, the differences in oral microbiota reported here for breastfed *versus* formula-fed infants cannot be attributed solely to the production of ROS described in our previous study⁶, which provided *in vitro* evidence for the mixing of neonatal saliva and breast milk in the regulation of the infant oral microbiota⁶. Our present *in vivo* findings support the concept of microbial regulation according to the mode of feeding, and agree with previous reports that the neonatal microflora of the GIT differs significantly between breast-fed and formula-fed infants^{4,33}. Our study only included infants delivered vaginally, to minimize the effect of mode of delivery on oral bacterial communities, and excluded mothers or babies who received antimicrobial treatment. Other studies have demonstrated differences for the oral microbiota of infants delivered vaginally compared with those born by Cesarean section¹³, and for infants who have been treated with antimicrobials¹⁶.

It was demonstrated here that the infant oral microbiota is dynamic between 4–8 weeks of age. The abundance of Family *Streptococcaceae*, the most common oral bacterial family identified in both breast-fed and formula-fed infants, accounting for 55–90% of all bacteria within our study. Furthermore, the abundance of the known streptococcal genera and species appeared to increase in the mouths of both breast-fed and formula-fed babies between 4 to 8 weeks of age. This was consistent with other studies showing that *Streptococcus* spp. were abundant in the saliva of infants^{14,34} and their incidence increased with age^{16,35}. One study of oral microbiota in six low-birth weight infants (five of whom were born prematurely) sampled at different ages reported that the *Streptococcus* spp. and *Staphylococcus* spp. were the most abundant genera in the mouths of all babies tested and these increased significantly between 8–21 days of life¹⁶.

Others studies have demonstrated a “protective” role for streptococci in the oral cavity. Hydrogen peroxide-producing streptococci were shown to inhibit the growth of pathogenic microorganisms, including the nosocomial pathogens methicillin-resistant *Staphylococcus aureus* (MRSA) and *Pseudomonas aeruginosa*³⁶. It has also been shown that lactic acid bacteria isolated from breast milk were able to inhibit growth of *Enterococcus faecalis*, *Salmonella enterica* subsp. *enterica*, *Listeria monocytogenes*, *Staphylococcus aureus* and *Escherichia coli*³⁷. Similarly, our recent study demonstrated that the mixing of infant saliva and breast milk during breast-feeding generates hydrogen peroxide concentrations in excess of 70 μ M, which were sufficient to inhibit the *in vitro* growth of pathogenic bacteria such as *Salmonella* spp.⁶. We also found unusually high concentrations of thiocyanate in neonatal saliva: this serves as a substrate for milk lactoperoxidase, which produces bactericidal ROS and hypothiocyanate⁶. Taken together, our studies implied an important role for ROS in regulating bacteria within the infant mouth. This raises the question of whether this primal metabolic mechanism for perinatally producing oral hydrogen peroxide may be replaced during weaning and teething by hydrogen peroxide-producing oral streptococci, thus maintaining the protective role of ROS throughout infancy and childhood?

While the findings of our current study are novel, there are some limitations that must be addressed. Over 50% of total reads were identified as members of the *Streptococcaceae* family and the majority were classified to genus *Streptococcus*. Some isolates belonging to the family *Streptococcaceae* were unable to be assigned to any known genus and this is likely as a consequence of the similarities in the gene sequence targeted in this study, and is consistent with the findings of others, that the discriminatory power of the 16S rRNA gene sequences was insufficient to successfully identify clinical isolates^{38,39}. Furthermore, we found that some 16S rRNA reads could not be assigned to any known taxonomic level beyond domain bacteria (Fig. 6). Reads belonging to ‘Bacteria_others’ were most closely related to sequences derived from human sources, e.g human forearm skin, human oral or nasal cavities and human respiratory tract. These sequences may be those of novel bacterial taxa, however we utilised stringent (85%) taxonomic classifications to identify the DNA sequences within our study, while the RDP classifier was able to provide reliable taxonomic identities at 60% taxonomic identity for fragment lengths ranging from 220 bp–550 bp. These stringent conditions, coupled with the short sequence lengths generated by the 16S rRNA gene deep sequencing, may have resulted in inability to group/identify some of these bacterial sequences to the genus level.

We anticipated a high prevalence of *Bifidobacterium* spp. within the mouths of breast-fed infants, as these bacteria are among the first anaerobic microorganisms to reach high concentrations within the GIT and are responsible for human milk oligosaccharide metabolism⁴⁰. However, we detected these bacteria in only one breast-fed and two formula-fed infants. This may be a consequence of the aerobic/microaerophilic environment of the mouth being inappropriate for their growth, or these oral microorganisms may colonise the mouth later in infancy (>8 weeks). Perhaps unsurprisingly, we identified inter-individual variations within both the breast-fed and formula-fed groups, both at taxon and phylum levels, perhaps arising from variation between host metabolisms, or the oral niches of the infants, and/or derived from maternal/environmental contact. Nonetheless, the differences were not marked and this may serve as a hallmark of stability of the microbial community and potentially act as biomarkers for infant health.

In conclusion, this current molecular study has provided novel information about the impact of breast-feeding *versus* formula-feeding on the oral microbiota of the neonate. Differences were detected in the bacterial profiles of breast-fed and formula-fed infants, contributing to our growing understanding of the infant microbiome, as well as the factors that may influence variation within the oral microbiota. We also noted alterations in the oral microbiota over time, with a trend of increasing abundance of *Streptococcus* spp. between 4, 6 and 8 weeks of life in both breast-fed and formula-fed neonates, perhaps highlighting the dynamic importance of these bacteria

	Forward fusion primer sequence	Reverse fusion primer sequence
Adaptor	5'CCATCTCATCCCTGCGTGTCTCCGACC	5'CCTATCCCCTGTGTGCCTTGGCAGTC
Key	TCAG	TCAG
MID	Barcode sample specific	—
Target primer*	AGAGTTTGATCMTGGCTCAG	GWATTACCGCGGCKGCTG

Table 2. The fusion primers design for 16S rRNA amplification, the fusion primer sequences contained [Adapter]-[Key]-[MID barcode]-[Target primers]. *Base redundancies were included within the target primers (W = A + T, K = T + G, M = A + C).

within the oral cavity. This study presents preliminary data prompting further research that will be critical to our understanding of the changing role of oral microbiota during neonatal development and ontogeny of dysbiosis.

Methods

Subjects and specimen collection. This study was approved by the Human Research Ethics Committees of the University of Queensland (2012000449) and Mater Mother's Hospital (2012_10LNR). The methods were conducted in accordance with the approved guidelines. Parents of the eligible infants gave informed consent for their infants to be recruited for this study. Breast-fed and formula-fed infants on antibiotics or infants whose mothers were on antibiotics before or at the time of collection, were excluded from the study. Swabs of the oral/cheek surfaces were collected from healthy, full-term, vaginally-delivered, Caucasian infants aged 4, 6 and 8 weeks who had not been fed for at least 30 min prior to specimen collection. Infants' oral cheek swab specimens were collected at the participants' homes from breast-fed (n = 26) and formula-fed (n = 12) infants. The inner cheek surface of each infant was swabbed gently for approximately 5 min using sterile swabs (Copan flocced swabs, Italy); sterile gloves were worn by the collector during the collection process. Three swabs were collected from each infant (at 4, 6 or 8 weeks), placed in sterile 1.5 mL Eppendorf tubes and snap-frozen on dry ice. Samples were immediately delivered to the laboratory and stored at -80°C for further analysis. Characteristics of infants included in this study are summarized in Table 1.

DNA extraction and purification. For total genomic DNA extraction, each cheek swab (1–2 swabs (of the total of 3 swabs) per neonate per time point) was mixed with 100 μL of 4 g/L lysozyme (20 mM Tris-HCl, pH 8.0; 2 mM EDTA; 1.2% Triton) (Sigma-Aldrich Pty Ltd, Castle Hill, NSW, Australia) and then incubated in a heating block for 35 mins at 37°C . After incubation, 100 μL of 100 $\mu\text{g}/\text{mL}$ Proteinase K (Qiagen Pty Ltd, Chadstone Centre, VIC, Australia) solution was added to the mixture and this was further incubated for 30 mins at 37°C . Finally, 400 μL of lysis buffer was added to the mixture, which then was incubated at 56°C for a further 10 mins. The total genomic DNA was subsequently purified using the QIAamp DNA Mini Kit (Qiagen Pty Ltd) following the manufacturer's instructions for Spin-Column Protocol for "DNA Purification from Buccal Swabs". The genomic DNA was eluted in 100 μL of PCR grade DNase/RNase-free distilled water (Gibco UltraPure™, Invitrogen, Mulgrave, VIC, Australia). Concentrations and the DNA quality was confirmed using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Scoresby, VIC, Australia) and by agarose gel electrophoresis in a 2% w/v tris borate EDTA (TBE) agarose gel containing 1 $\mu\text{g}/\text{mL}$ ethidium bromide (Sigma-Aldrich Pty Ltd) (results not shown). The genomic DNA concentration was standardised for all samples to 10 ng/ μL before PCR amplification. These DNA samples then were sent to the Australian Genome Research Facility (AGRF, University of Queensland, Brisbane, QLD, Australia) for bacterial 16S rRNA gene sequencing using the Roche 454 platform.

PCR amplification of the bacterial 16S rRNA gene. The PCR assay was performed at AGRF using a Forward 27 F primer (AGAGTTTGATCMTGGCTCAG) and a Reverse 519 R primer (GWATTACCGCGGCKGCTG)⁴¹, with fusion primer sequences for both the forward and reverse directions for each amplicon. Each primer sequence included the appropriate adapter sequence for Lib-L or Lib-A libraries followed by the key sequence of TCAG. This was followed by the unique "barcode" sequences for a GS FLX MID sequencing platform for forward and reverse directions; the barcode sequences were used subsequently to identify and distinguish each of the samples. Table 2 summaries the design structure of the fusion primers, and the 16S rRNA target sequence primers [Adapter]-[Key]-[MID barcode]-[Target primers].

PCR amplification was conducted in 384 well plates using 8 μL total volume per reaction. Each PCR reaction contained: 2 μL DNA template, 4 μL master mix containing Ampli Taq Gold (AmpliTAQ GOLD 360, Applied Biosystems, USA), and 2.25 μM of primers. The PCR cycling parameters were: 94°C for 3 mins; followed by 34 cycles of 94°C for 45 sec, 50°C for 60 sec, 72°C for 60 sec; a final extension step at 72°C for 7 mins; and then the reactions were held at 10°C . All amplicons were purified using a robotic liquid handler (Beckman Coulter 384 Biomek NX, CA, USA) and Agencourt AMPure (Beckman Coulter, CA, USA).

Sample QC post PCR purification. The PCR amplicons were quantified using PicoGreen® (Invitrogen, Mulgrave, VIC, Australia) to determine the DNA concentration and samples were normalised to the same molarity, then a qPCR reaction was performed to check each sample. The samples were re-normalised based on the qPCR copy number results. This double normalisation technique was effective at ensuring an even number of de-multiplexed reads post sequencing on the GSFLX.

16S rRNA gene sequencing ('deep sequencing'). Samples for bead-based sequencing were set up according to Roche's standard Lib-L emPCR method for XLR70 chemistry on the GSFLX + instrument, emulsion

PCR (emPCR) Method Manual – Lib-L MV GX FLX Titanium Series October 2009 (Rev. Jan 2010), Germany. Emulsions comprised 2 x medium volume, both at 0.35 copies per bead (cpb) to achieve recoveries of 18% and 20% (the recommended range is 5% to 20%).

Two million beads were sequenced in a half Pico titre plate in Roche-454 GS FLX amplicon sequencing platform (Roche, Castle Hill, NSW, Australia) according to Roche's standard XLR70 protocol for the GSFLX + instrument. Subsequently, the quantified libraries were amplified in micro-reactors through emPCR followed by streptavidin bead enrichment and emulsion breaking. The beads attached to amplified DNA fragments were denatured with 1 N NaOH solution and annealed to a specific sequencing primer. All these steps and subsequent sequencing steps on the Roche-454 GS FLX amplicon sequencing platform were performed according to the Roche sequencing manual protocol. All sequence files were taken in fastq format then analysed using FASTQC. The FASTQC quality analysis tool assisted in determining the length of amplicons and its quality score distribution⁴². Standard bioinformatics quality control (QC) involved reviewing internal QC metrics and de-multiplexing the results as per the barcodes included in the pool.

454 bacterial 16S rRNA sequence analysis and statistics. The composition of species diversity in high-throughput amplicon sequencing data samples was analyzed using the Quantitative Insights Into Microbial Ecology (QIIME) software package version 1.7.0 (<http://qiime.org/>) for profiling bacterial 16S rRNA⁴³. In order to analyse the sequence dataset using the QIIME software/protocol, the sequence dataset was first depleted of sequences with lengths below 250 bp but not higher than 550 bp⁴⁴. The FASTQC length distribution is shown in the Supplementary Info Figure 1.

Sequences with more than six homopolymer runs were discarded and no barcode or primer mismatches were allowed. Low quality sequences with a phred score <25 (<Q25) and the reads from the forward primer were then selected and sample identities (breast-fed, formula-fed; 4, 6, or 8 weeks of age) were assigned based on barcode sequences. All barcodes and primer sequences were trimmed. The remaining high quality reads were then used to determine the diversity and richness of bacterial communities and the relative abundance based on operational taxonomic unit (OTU) using QIIME version 1.7.0.

Mean values of infant characteristics (gestational age, birth weight and age at the time of sampling) in breast-fed and formula-fed infants were analysed using an unpaired, two-tailed t-test. One-way analysis of variance (ANOVA) was performed to compare the relative abundance of phyla and genera in breast-fed and formula-fed babies. Regression analyses were employed for analysis of feeding mode over time (i.e. at 4, 6 and or 8 weeks of age). All p-values were calculated at 95% confidence intervals and statistical significance was accepted as $p < 0.05$.

Total bacterial community composition. Bacterial communities were grouped into clusters (6101, in total) based on 3% (or 97% identity cut-off) sequence variation using UCLUST clustering algorithm⁴⁵. Each cluster considered as an OTU and a representative OTU was selected to represent each cluster. Representative OTU sequence reads were aligned using QIIME based Nearest Alignment Space Termination (NAST) alignment. For taxonomy-based analysis, the Ribosomal database project (RDP) Naïve Bayesian classifier⁴⁴ was used at a confidence threshold of 85% against the Greengenes (version 13_8) database. QIIME based composition analysis was used to quantify the total number of reads per each OTU and their relative abundance at phylum and genus levels. Bacterial OTUs assigned with one sequence read were removed prior to composition analysis. To account for different number of reads across individuals, raw abundance counts were normalised using total sum scaling method, which present the contribution of each bacterial taxa as a relative proportion to individual reads count⁴⁶.

The Shannon diversity index was calculated as a measure of species richness and evenness of bacterial taxa within individual samples of breast-fed and formula-fed babies. The diversity between the samples (Beta diversity) was estimated using unweighted UniFrac distances⁴⁷. Box plot graphs, cluster and heatmap diagrams were generated to illustrate bacterial community diversity and compositions using R project (ggplot2 package) for statistical computing (<http://www.r-project.org>) and Calypso web Online tool (<http://bioinfo.qimr.edu.au/calypso>). Phylogenetic trees were constructed using Randomised accelerated Maximum Likelihood (RaxML) HPC version 7.2.8⁴⁸. Sequence reads were stored and managed using Filemaker Pro Advanced relational database management software (10.0 Ver.1) and diagrams were edited using Pages software on Mac OSX.

Determination of 'Bacteria_others'. Sequences from 'Phylum-Unclassified' bacteria identified by deep sequencing were selected from OTU clusters. In this study we focussed only on the OTUs with high abundance and occurrence. We excluded the OTUs with single read (singletons) and ones that appeared in less than 50% of samples. In total, 106 OTUs selected from 653 OTUs identified as 'Bacteria_others'. The taxonomic identities of these 'Bacteria_others' OTUs were further explored using RDP naïve Bayesian classifier (16S bacteria training set 9) and the relative abundance of 'Bacteria_others' OTUs were shown together with predicted class levels taxonomic assignments. The bootstrap sequence similarity percentage of OTUs and their composition was described using a heat map diagram constructed using R project for statistical computing. In addition, we traced the source of origin of identified bacterial taxa. A phylogenetic was then constructed a using Randomised accelerated Maximum Likelihood (RaxML) HPC version 7.2.8 with 1000 bootstrap replications and General Time Reversal (GTR GAMMA) model.

References

1. Aagaard, K. *et al.* The placenta harbors a unique microbiome. *Sci. Transl. Med.* **6**, 237ra265 (2014).
2. Gueimonde, M. *et al.* Effect of maternal consumption of lactobacillus GG on transfer and establishment of fecal bifidobacterial microbiota in neonates. *J. Pediatr. Gastroenterol. Nutr.* **42**, 166–170 (2006).
3. Vaishampayan, P. A. *et al.* Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biol. Evol.* **2**, 53–66 (2010).

4. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. USA* **108** Suppl 1, 4578–4585 (2011).
5. Al-Shehri, S. *et al.* Collection and determination of nucleotide metabolites in neonatal and adult saliva by high performance liquid chromatography with tandem mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **931C**, 140–147 (2013).
6. Al-Shehri, S. S. *et al.* Breastmilk-Saliva Interactions Boost Innate Immunity by Regulating the Oral Microbiome in Early Infancy. *PLoS ONE* **10**, e0135047 (2015).
7. Bertino, E. *et al.* Benefits of donor human milk for preterm infants: current evidence. *Early Hum. Dev.* **85**, S9–S10 (2009).
8. Kelley, L. Increasing the consumption of breast milk in low-birth-weight infants: can it have an impact on necrotizing enterocolitis? *Adv. Neonatal Care* **12**, 267–272 (2012).
9. Gephart, S. M., McGrath, J. M., Effken, J. A. & Halpern, M. D. Necrotizing enterocolitis risk: state of the science. *Adv. Neonatal Care* **12**, 77–87; quiz 88–79 (2012).
10. Schurr, P. & Perkins, E. M. The relationship between feeding and necrotizing enterocolitis in very low birth weight infants. *Neonatal Netw.* **27**, 397–407 (2008).
11. Meinzen-Derr, J. *et al.* Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. *J. Perinatol.* **29**, 57–62 (2009).
12. Levy, O. Innate immunity of the newborn: basic mechanisms and clinical correlates. *Nat. Rev. Immunol.* **7**, 379–390 (2007).
13. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. USA* **107**, 11971–11975 (2010).
14. Cephas, K. D. *et al.* Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. *PLoS ONE* **6**, e23503 (2011).
15. Holgerson, P. L. *et al.* Oral microbial profile discriminates breast-fed from formula-fed infants. *J. Pediatr. Gastroenterol. Nutr.* **56**, 127–136 (2013).
16. Costello, E. K., Carlisle, E. M., Bik, E. M., Morowitz, M. J. & Relman, D. A. Microbiome assembly across multiple body sites in low-birthweight infants. *mBio* **4**, e00782–00713 (2013).
17. Wade, W. G. The oral microbiome in health and disease. *Pharmacol. Res.* **69**, 137–143 (2013).
18. Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. *Science* **352**, 539–544 (2016).
19. Wang, W. L. *et al.* Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* **21**, 803–814 (2015).
20. Isolauri, E., Rautava, S. & Salminen, S. Probiotics in the development and treatment of allergic disease. *Gastroenterol. Clin. North Am.* **41**, 747–762 (2012).
21. Jonkers, D., Penders, J., Masclee, A. & Pierik, M. Probiotics in the management of inflammatory bowel disease: a systematic review of intervention studies in adult patients. *Drugs* **72**, 803–823 (2012).
22. Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J. Clin. Microbiol.* **43**, 5721–5732 (2005).
23. Lif Holgerson, P., Harnevik, L., Hernell, O., Tanner, A. C. & Johansson, I. Mode of birth delivery affects oral microbiota in infants. *J. Dent. Res.* **90**, 1183–1188 (2011).
24. Ward, T. L., Hosid, S., Ioshikhes, I. & Altosaar, I. Human milk metagenome: a functional capacity analysis. *BMC Microbiol.* **13**, 116 (2013).
25. Jost, T., Lacroix, C., Braegger, C. P. & Chassard, C. New insights in gut microbiota establishment in healthy breast fed neonates. *PLoS ONE* **7**, e44595 (2012).
26. Rodriguez, J. M. The origin of human milk bacteria: is there a bacterial entero-mammary pathway during late pregnancy and lactation? *Advances in nutrition (Bethesda, Md.)* **5**, 779–784 (2014).
27. Kononen, E., Jousimies-Somer, H. & Asikainen, S. Relationship between oral gram-negative anaerobic bacteria in saliva of the mother and the colonization of her edentulous infant. *Oral Microbiol. Immunol.* **7**, 273–276 (1992).
28. Kononen, E. *et al.* Transmission of oral *Prevotella melaninogenica* between a mother and her young child. *Oral Microbiol. Immunol.* **9**, 310–314 (1994).
29. Urbaniak, C., Angelini, M., Gloor, G. B. & Reid, G. Human milk microbiota profiles in relation to birthing method, gestation and infant gender. *Microbiome* **4**, 1 (2016).
30. Mutschlechner, W. *et al.* Mammary candidiasis: molecular-based detection of *Candida* species in human milk samples. *Eur. J. Clin. Microbiol. Infect. Dis.* (2016).
31. Chen, P. W., Tseng, S. Y. & Huang, M. S. Antibiotic Susceptibility of Commensal Bacteria from Human Milk. *Curr. Microbiol.* **72**, 113–119 (2016).
32. Boix-Amoros, A., Collado, M. C. & Mira, A. Relationship between Milk Microbiota, Bacterial Load, Macronutrients, and Human Cells during Lactation. *Frontiers in microbiology* **7**, 492 (2016).
33. Martin, V. *et al.* Sharing of bacterial strains between breast milk and infant feces. *J. Hum. Lact.* **28**, 36–44 (2012).
34. Luo, A. H., Yang, D. Q., Xin, B. C., Paster, B. J. & Qin, J. Microbial profiles in saliva from children with and without caries in mixed dentition. *Oral Dis.* **18**, 595–601 (2012).
35. Kononen, E. Development of oral bacterial flora in young children. *Ann. Med.* **32**, 107–112 (2000).
36. Uehara, Y. *et al.* H(2)O(2) produced by viridans group streptococci may contribute to inhibition of methicillin-resistant *Staphylococcus aureus* colonization of oral cavities in newborns. *Clin. Infect. Dis.* **32**, 1408–1413 (2001).
37. Reis, N. A. *et al.* Probiotic properties of lactic acid bacteria isolated from human milk. *J. Appl. Microbiol.* (2016).
38. Lal, D., Verma, M. & Lal, R. Exploring internal features of 16S rRNA gene for identification of clinically relevant species of the genus *Streptococcus*. *Ann. Clin. Microbiol. Antimicrob.* **10**, 28 (2011).
39. Rossi-Tamisier, M., Benamar, S., Raoult, D. & Fournier, P. E. Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int. J. Syst. Evol. Microbiol.* **65**, 1929–1934 (2015).
40. Roger, L. C., Costabile, A., Holland, D. T., Hoyles, L. & McCartney, A. L. Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* **156**, 3329–3341 (2010).
41. Turner, S., Pryer, K. M., Miao, V. P. & Palmer, J. D. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* **46**, 327–338 (1999).
42. Andrews, S. FastQC: a quality control tool for high throughput sequence data. From: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).
43. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
44. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
45. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
46. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5**, e1000352 (2009).
47. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
48. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

Acknowledgements

We thank Mater Research Institute (MRI) for recruiting study participants and organising sample collections. The authors acknowledge financial support from The Mater Children's Hospital Golden Casket Research Fund and the National Health and Medical Research Council (Grant 1043282).

Author Contributions

S.A. designed the experiments, collected samples, carried out the molecular genetic studies, participated in the sequence alignment, bioinformatics, data analysis, and drafted the manuscript. E.S. participated in the design of the study, helped in performing of the experiments and in the drafting of the manuscript. D.C. participated in the design of the study, supervised the project, participated in the drafting of the manuscript. H.L. supervised the project, participated in the drafting of the manuscript, coordinated samples collection. P.R. performed the sequence alignment, bioinformatics, and statistical analysis. B.C. supervised the project, participated in the drafting of the manuscript. N.S. supervised the project, participated in the drafting of the manuscript. D.V. participated in the statistical analysis. J.D. participated in the design of the experiment and statistical analysis, facilitated samples collection, main supervisor of the project, and drafting of the manuscript. C.K. participated in the design of the experiment, bioinformatics and statistical analysis, supervised the project, and drafting of the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Al-Shehri, S. S. *et al.* Deep sequencing of the 16S ribosomal RNA of the neonatal oral microbiome: a comparison of breast-fed and formula-fed infants. *Sci. Rep.* **6**, 38309; doi: 10.1038/srep38309 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016