Is house dust a reservoir for infant gut bacteria?

MPH Project Final Paper

Theodore Konya

Dalla Lana School of Public Health, University of Toronto

2/27/2012

Abstract

Numerous studies have identified an association between the infant gut microbiota and the development of asthma and atopic disease. Certain environmental exposures are known to shape the succession of infant gut microbiota; including birth delivery mode, antibiotic use, breast or formula feeding, and the mother's own microbiota. This study investigates the extent to which the indoor environment itself may serve as a reservoir for gut bacteria. This study compared the microbiota of household dust with that of infant fecal samples collected at 3 months of age. Samples of infant stool and household dust were obtained for twenty 3-month-old infants from Winnipeg, Canada; these subjects were early recruits in the Canadian Healthy Infant Longitudinal Development (CHILD) study. For each sample, community bacterial 16S rDNA was sequenced using a novel Serial Illumina Sequencing (SI-seq) method. Summary and statistical analyses were performed using MS Excel, PC-ORD, and R on the full data set and a conserved data set that was reduced by an abundance cut-off. Both stool and dust samples were dominated by members of the phyla Actinobacteria, Firmicutes, and Proteobacteria with dust samples also having a large content of Cyanobacteria. Analysis by nonmetric dimensional scaling (NMDS) revealed dust and stool communities as 2 distinct groups. A modified permutation test revealed that the number of shared OTUs between dust and stool of the same subject was significantly higher than expected ($p = 0.0328$). The findings demonstrate a clear difference in the microbiomes of infant gut and household dust. However, the co-occurrence of select taxa in paired dust and stool samples suggests an association between the bacterial populations of these communities. Using a more conserved data set had minimal affect on the overall outcome of analyses.

Introduction

The human gut contains approximately 10^{14} bacteria from at least 400-500 different species (Penders et al. 2007). These commensal bacteria create an ecological niche within the intestines and help protect the human from pathogenic bacteria. In addition to that, there is a belief that the first colonizers of the infant gut are pivotal to immune development. Numerous studies have found associations with the microbial composition of the infant gut and asthma and allergy (Sjögren et al. 2009; Johansson et al. 2011; Martin et al. 2010; van Nimwegen et al. 2011; Abrahamsson et al. 2012; Nakayama et al. 2011).

Bacterial colonization of the gut occurs in four different phases. Phase 1 occurs over the first 2 weeks of life and has the greatest fluctuation in species that are able to colonize. Phase 2 includes the period of breast-feeding and a difference between breast-fed and bottle-fed babies is seen. Phase 3 begins at the point of weaning and the introduction of solid foods; the colonization differences between types of feeding disappear. Lastly Phase 4 begins at the conclusion of weaning and is when the infant gut microbiota begins to resemble the adult gut microbiota (Penders et al. 2007).

The infant gut is considered to be sterile while inside the womb; the first exposure to bacteria comes during delivery where the infant encounters the mother's vaginal, fecal, and perineal microbial communities (Adlerberth et al. 2006; Orrhage and Nord 1999; Koening et al. 2011). Bacteria that is ingested by the infant from habitats other than the gastrointestinal tract will mainly pass through the

intestince unable to colonize, but some thrive under the gut conditions and become members of the resident microbial community (Apajalahti J 2005). Facultative bacteria are the first colonizers of the infant gut due to the adbundance of oxygen (Vael and Desager 2009; Orrhage and Nord 1999). As the oxygen begins to be consumed, the oxidation-reduction potential decreases, allowing for anaerobic bacteria to begin colonization (Orrhage and Nord 1999).

Phase 2 begins to show stability in the types of bacteria colonizing the gut. At the start of Phase 2 the facultative bacteria are in high numbers but gradually decrease as the reduced oxygen environment makes it more favourable for complex anaerobic bacteria to colonize (Adlerberth et al. 2006; Vael and Desager 2009). Phase 3 data shows the microbiota of bottle-fed and breast-fed infants becoming more similar as both sets are weaned from liquid food and introduced to solid food (Mackie et al. 1999) and the microbiota becomes more adult-like (Vael and Desager 2009). The microbiota becomes much more complex, dominated by obligate anaerobes, and provides a strong barrier against the establishment of new bacterial strains (Adlerberth and Wold 2009).

Infants encounter different factors during the first year of life that can affect bacterial colonization. Some intrinsic/host factors are anatomical development of the intestinal tract, peristalsis, bile acids, intestinal pH and immune responses, microbial interactions, mucosal receptors, and drug therapy (Fanaro et al. 2003; Penders et al. 2007; Apajalahti J 2005). Several environmental factors have been found to be associated with microbial succession patterns in the developing infant gut, such as gestational age, mode of delivery, breastfeeding, age of weaning, number of siblings, and exposure to antibiotics (Yap et al. 2011; Torrazza and Neu 2011).

More specifically, caesarean section delivered babies have been found to have delayed colonization by Bacteroidetes, *Bifidobacteria*, *Lactobacillus* and *E. coli* (O'Toole and Claesson 2010; Torrazza and Neu 2011). Prematurity has been linked to increased colonization by *C. difficile*, along with the administration of antibiotics to these babies that lead to reductions and extinctions of parts of the microbiota. Type of feeding has also been associated with changes in the gut microbiota; where formula enriched with oligosaccharides stimulates bifidobacterial numbers (O'Toole and Claesson 2010), and breast-fed babies have higher counts of *Staphylococci* and lower counts of *Clostridia*, *Bacteroides*, and *Enterobacteria* (Vael and Desager 2009; Orrhage and Nord 1999; Adlerberth and Wold 2009; Fanaro et al. 2003). Poor hygiene and sanitation has been linked to earlier colonization with Enterobacteriaceae (O'Toole and Claesson 2010).

A factor that has not been studied as having an effect on the developing gut microbiota is the indoor home environment. Vacuum collected dust is a reservoir of microbial communities that can act as a representative of the microbial exposure from the indoor environment. Only one study has been found in the literature that can link bacteria in indoor dust and the infant gut. Nevas et al. found *Clostridium botulinum* type B in both the intestinal contents of a deceased 11-week-old infant and vacuum cleaner dust from the deceased's household (Nevas et al. 2005). Beyond this study, there has not been an in depth look at a possible association between the microbial communities of household dust and the infant gut.

Studying microbes in dust and their effect on human health is not a new idea; Cooke's paper in 1922 was the first to link house dust as a cause of symptoms in rhinitis and asthma (Baldwin et al. 1957), followed by Cruickshank in 1935 and White in 1936 linking other illnesses to dust (Garrod 1944). Early studies of the bacterial communities in dust were conducted by Lidwell and Lowbury in 1950 and revealed a large variation between same samples of dust (Lidwell and Lowbury 1950). More recently studies have found links between bacteria in house dust and allergies and asthma (Pakarinen et al. 2008; Maier 2010; Gehring 2007; Hyvärinen 2006).

In this study we attempt to find an association between the microbial communities of household dust and the infant gut. As opposed to searching for specific species, we use a community based approach as "ecological measures indicate that the developing infant gut microbiota is composed of interacting bacterial consortia, not of randomly assembled suites of bacteria" (Koening et al. 2011). Understanding the succession of bacteria in the human gut can aid in developing strategies to guide the formation of health-promoting microbiotas (Koening et al. 2011), specifically in manipulating the indoor home environment.

Methods

Sample collection

Fecal samples were collected from infants as part of the 3 month home visit of the Canadian Healthy Infant Longitudinal Development (CHILD) study. Mothers were instructed to place a diaper liner in her child's diaper the day before the home visit and to change the liner with the diaper until stool has been deposited in the liner. Once deposited, she was instructed to place the diaper with diaper liner into a collection bag, record the time of collection, and place the sealed bag into a refrigerator until collected by the research team at the home visit. Trained research assistants collected the fecal sample from the diaper using a screw cap container with spoon (Globe Scientific, Paramus, NJ, USA) and placed the collection container in a cooler for transport until it could be placed in a -80°C freezer for long-term storage. An approximately 40mg subsample was aliquoted and stored at -80°C until processed.

Vacuum collected dust samples were also collected by the research technicians at the 3 month home visit. Technicians used a clean, depyrogenated CHILD study designed aluminum collection device attached to the end of a vacuum (Sanitaire Canister Vac: Model S3680 Serial#074002221) to collect floor dust. The collection device holds two nylon DUSTREAM™ filters (Indoor Biotechnologies Inc, Charlottesville, VA). Technicians were instructed to collect dust over 2 square meters of carpeted floors, or the entire floor in a room with only hard floors. Dust samples were sent to the dust processing centre in Toronto where the two thimbles from a single room were combined and shaken in 150um sieve (No 100 sieve, Legends Inc, Sparks, NV) for 15 minutes to separate the fine and coarse portions of dust. The coarse portion used for this study was stored at room temperature until DNA extraction could begin.

At the same visit, the technicians completed a detailed environmental home assessment and worked with mothers to complete a questionnaire outlining the family's activities during the child's first 3 months of life. Data from both the assessment and questionnaire was pulled to determine

environmental factors that could have an effect on the microbial composition of household dust or infant stool.

DNA Extraction and Amplification

Whole genome DNA was extracted from the stool aliquot and coarse dust using the FastPrep DNA for Soil Kit (MP Biomedicals Inc, Solon, OH, USA). Bacterial 16S DNA, hypervariable regions V5-V7, was amplified through PCR using forward primer V5+791 (5'-3' sequence: AATCAGGCGGGKAKCRAACVGGATTAGATACCCBGGTAGTCCWNRCHSTAAACGDTG) and reverse primer Uni-1104 (5'-3' sequence: AATCAGGCGGSCRTRMKGAYTTGACGTCRYCCCCDCCTTCCTCC). Each primer contained a phosphate added to the 5' end for linking to the Illumina plate. The primers were also barcoded so each sample could be uniquely identified post-sequencing. Each PCR mixture (50µl) contained 5 µl 10X Hotstart Buffer, 400mM dNTPs, 1.5mM $MgCl_2$, 2.5Uof Hotstart Taq polymerase (Fermentas, Glen Burnie, MD, USA), 1mg Ultrapure Bovine Serum Albumin (Ambion, Austin TX), molecular biology reagent grade water (Sigma-Aldrich, St. Louis, MO, USA) , 0.16µM primer, and 2µl bacterial template DNA (10ng/µl). The PCR program for stool consisted of an initial DNA denaturation step at 94°C (4 min), followed by 18 cycles of DNA denaturation at 94°C (45sec), an annealing step at 56°C (30 sec) and an elongation step at 72 °C (2 min 30 sec), and was performed on the PTC-200 Peltier Thermal Cycler (MJ Research, St. Bruno, QC, Canada).

qPCR of the dust samples revealed a large amount of inhibition and varying optimum cycle numbers, which resulted in a modified amplification procedure. All dust samples first followed the same PCR program as stool, stated above, except the number of cycles was based on the qPCR results and ranged from 16 to 25 cycles. The PCR product was diluted 1:10 and re-amplified following the same PCR program for only 12 cycles.

All final PCR product was cleaned with GENECLEAN® Turbo Kit (MP Biomedicals Inc, Solon, OH, USA) and gel purified using the E-gel® SizeSelect™ 2% agarose gel cutting system (Invitrogen, Carlsbad, CA, USA). Fifty nanograms of cleaned/extracted product from each sample was combined and concentrated, using an Amicon® Ultra-4 30K centrifugal filter (Millipore, Billerica, MA, USA), for sequencing.

Sequencing

Sequencing was performed by the Centre for the Analysis of Genome Evolution & Function (CAGEF) at the University of Toronto using a novel Serial Illumina Sequencing method developed by the Dr. David Guttman lab (publication in process). The method involved only sequencing the hypervariable V5, V6, and V7 regions of the 16S rDNA gene to create fragmented reads which are then concatenated and mapped to a Silva dataset to produce the best result for a "correct orientation". FASTA files obtained from CAGEF were mapped to a SILVA reference database to identify taxonomic classifications for each operational taxonomic unit (OTU) identified through sequencing.

A single species control was placed in the run to determine a minimum abundance cut-off that would exclude rare taxa that may have only been present due to contamination or sequencing errors.

The minimum abundance cut-off for our data was 0.18%. Data analysis was performed twice; once with the full sequence data set and a second time with the taxa below the abundance cut-off removed.

Data Analysis

Summary statistical analysis was performed using Microsoft Excel.

With PC-ORD version 6 (McCune & Mefford 2011), patterns of covariation were observed using the unconstrained ordination method Nonmetric Multidimensional Scaling (NMDS). NMDS was chosen as the best ordination method due to our dataset being highly hetergeneous and having non-linear relationships among responses (Peck 2010). An initial NMDS was performed using the autopilot mode to find the number of axes that best represents the variation in the dataset. The process was repeated 3 times to ensure qualitatively consistent results.

A modified permutation test was performed in R to find an association between dust and stool communities. The usual use of the permutation test is like a t-test, where one compares two groups. Accordingly, the permutation test permutes the labels of the two groups. In this study, we lack two groups but rather have "actual pairings" and "randomly assigned pairings", which are the agents that are permuted. Two different scoring rules were used, but our results were the same for both scoring approaches. The score reported here is the square root of the sum of squares:

$$\text{Score} = \sqrt{\sum_{i=1}^{n} d_i^2}$$

where *n* is the number of samples, *d* is the number of shared OTUs between a dust and stool sample, and *i* is the index of summation.

In our test, 100,000 permutations were performed in which 1 of 20 randomly selected stool samples was paired with 1 of 20 randomly selected dust samples. The resulting observed overlaps in OTUs were used to compute an overall score for the number of overlaps in a set of permuted pairings. The distribution of overlapped scores was plotted.

Results

Summary statistics

Sequence reads ranged from 487 to 451,927 reads per sample, averaging 43,271 reads per sample for dust, and 116,971 reads per sample for stool. Overall, 3,408 OTUs were identified with 838 unclassified OTUs. A total of 3,287 OTUs were found in the dust samples and 446 OTUs in stool. After applying the abundance cut-off the total number of OTUs was reduced to 734 with 37 unclassified. 652 of those OTUs were present in the dust samples while 108 were in the stool.

Both stool and dust samples were dominated by members of the phyla Actinobacteria, Firmicutes and Proteobacteria, with dust samples also showing a large content of Cyanobacteria (Figure 1). After applying the abundance cut-off the phyla proportions change slightly (Figure 2). Most notably,

the number of phyla present drops from 20 to 13. Also, the proportion of Actinobacteria becomes higher in dust (42%) than stool (39%), whereas it was higher in stool before the cutoff (57% versus 43%). The proportion of Firmicutes in stool becomes much higher compared to stool after the abundance cut-off (49% versus 28%).

Classifying to the genus level shows a clearer difference between the taxa present in dust and stool. There were 487 genera identified in the total dataset; 476 of those were present in dust while 71 were present in stool. Figure 3 shows the genus-level taxa present at greater than 1% abundance of the total when samples were pooled before the abundance cut-off; dust samples yielded 18 genera while stool had only 9. *Bifidobacterium* clearly dominates within the stool samples at 67% compared to the next highest genus in abundance *Streptococcus* at 5%. The genera abundance in dust is more spreadout, although *Streptococcus* and *Corynebactierum* are higher than the others at 16% each. After applying the abundance cut-off the total number of genera present was reduced to 199; 187 of those were present in dust while just 32 were present in stool. The abundance of *Bifidobacterium* in stool was greatly reduced to 34%, although it is still clearly more dominant than any other genus. The next highest abundant genus in stool changes to *Escherichia* at 9%. In dust, the distribution of genera abundance is similar after the cut-off with *Streptococcus* and *Corynebactierum* still being the highest at 10% and 12% respectively (Figure 4).

Nonmetric Multidimensional Scaling

OTU level analysis

Iterative NMDS analyses showed a three-dimensional solution was appropriate (randomization test p=0.004) with a final stress of 15, representing 45% of the variance of the full data set. When looking at the ordination plot of the first two axes (Figure 5), representing 34.9% of variation, there are two clear clusters present. The dust samples (red triangles) are clustered on the left side of the plot, while the stool samples (green triangles) are clustered on the right.

After the abundance cut-off, iterative NMDS analyses again showed a three-dimensional solution was appropriate (randomization test p=0.004) but with a final stress of 14, representing 44.1% of the variance. Looking at the ordination plot of the first two axes (Figure 6), representing 35.2% of variation, again shows two distinct clusters for the dust and stool samples. Interestingly one dust sample, subject ID# 40010, is situated to the far left of the plot away from the dust cluster and appears to potentially be an outlier. Performing outlier analysis, however, reveals that none of the samples fall outside 2 standard deviations and are thus not outliers.

Genus level analysis

Given the relatively low amount of variation explained by NMDS at the OTU level, we condensed the taxa to the genus level and performed the same NMDS analysis. Iterative NMDS analyses also showed a three-dimensional solution was appropriate (randomization test p=0.004) with a final stress of 10, representing 80% of the variance of the full data set. Stool and dust communities were also clearly separated in the ordination plot of the first two NMDS axes (Figure 7a), representing 67.6% of the variation. However, with the genus level data the stool samples were further subdivided into 2 groups

(blue shading). Overlays of genus-based abundance (indicated by symbol marker size) were assessed visually to determine which genera were most influential in shaping the gradients along the axes. One stool subgroup showed increased *Bifidobacterium* abundance (Figure 7b). In contrast to *Bifidobacterium*, the genera *Actinoplanes* (Figure 3c) and *Friedmaniella* (Figure 3d) showed increased abundance in dust communities.

After performing the abundance cut-off, iterative NMDS analyses revealed the same solution as before the cut-off was appropriate, however the percent of variation explained was reduced to 75%. Another change was with the ordination plot that represented the most variation; this time axis 1 versus axis 3 was better, representing 61.9% of the variation (Figure 8a). Visually, the ordination plot looks very similar to before the cut-off; and again the greatest explanation for the separation of stool into 2 subgroups can be visually seen by overlaying the abundance of *Bifidoacterium* (Figure 8b). Abundance of *Acitonplanes* was again responsible for the separation of dust from stool (Figure 8c), but this time abundance of *Modestobacter* was also explanatory (Figure 8d).

For all NMDS ordination plots, a second matrix consisting of environmental variables (obtained from the questionnaires and home assessment) that might also explain variations in bacterial communities was overlaid onto the plots. The variables used were antibiotic administered to child, vaginal or caesarean section delivery, breast milk or formula feeding, room of dust collection, total amount of dust collected, percent of fine dust, endotoxin levels in fine dust, B-D-glucan levels in fine dust, type of floor where dust was collected, and size of area where dust was collected. Visually, there was no clear variable that could reasonably explain the variation represented in any of the ordination plots.

Shared OTUs

Prior to the abundance cut-off, all subjects had at least 1 OTU shared between their stool and dust samples. The highest number of shared OTUs was 14 and the average for all 20 subjects was 5.2 OTUs. When looking at the taxonomic classification of the five specific OTUs that were shared most between the samples, there is a variety of types present (Table 1). Four of the five are from the phylum Firmicutes while the other one is a Proteobacterium. At the genus level, five different genera are represented: *Lactococcus*, *Escherichia*, *Streptococcus*, *Finegoldia*, and *Veillonella*.

After the cut-off the number of shared OTUs dropped considerably. Five subjects did not have any shared OTUs between their stool and dust sample, the highest number of shared OTUs was 4, and the average was 1.35. The taxonomic classifications of the specific OTUs shared most between the samples differs from the ones before the cut-off (Table 1). Differing from before the cut-off, two of the six are Actinobacteria, three are Firmicutes, and one is a Proteobacterium. Only two genera are the same as before the cut-off, *Escherichia* and *Streptococcus*. *Bifidobacterium* is represented in the two Actinobacteria, and two other Firmicutes could only be classified to the family level, Lachnospiriceae.

Permutation Test

The results of the permutation test before and after the abundance cut-off can be seen in Table 2. The score value assigned to the actual pairings of dust and stool samples for the same subject before

the cut-off was 28.67. The permutation tests revealed a mean of 23.67 (SD = 2.28) and a 95% confidence interval of 19.34, 28.16 for the randomly generated scores. When comparing the actual pairings score to the random score we see that it falls outside the 95% confidence interval and is thus statistically higher than expected by random (α=0.05, p (2-sided) = 0.0282). This can also be visually seen when plotting the distribution of the random pairings scores and placing the actual score as a vertical line over the distribution (Figure 9).

The score value for the actual pairings after the abundance cut-off dropped by 21 points to 7.81; reflecting the loss in number of shared OTUs between stool and dust samples of the same subject described above. The permutation tests revealed a mean of 5.82 (SD = 0.86) and a 95% confidence interval of 4.24, 7.62 for the randomly generated scores. When comparing the actual pairing score to the random score we again see that it falls outside the 95% confidence interval and is statistically higher than expected by random (α=0.05, p (2-sided) = 0.0327). Again, this can be visually seen when plotting the distribution of the random pairings scores and placing the actual score as a vertical line over the distribution (Figure 10).

Discussion

Comparison to other studies

The sequencing results of household dust revealed similarities and differences between other published studies of the same materials. First, the number of taxa found in our samples (3,287 before the cut-off or 652 after) was higher than previously reported. Using culture-based methods previously reported counts in dust were 55 species (Horak et al. 1996), 167 species (Górny and Dutkiewicz 2002), and 501 bacterial strains (Bouillard et al. 2005). Culture-independent studies have reported 190 bacterial species (Lee et al. 2007) and 248 OTUs (Täubel et al. 2009). At the higher taxonomic level, the dust results from this study are quite similar to other studies. For instance, other studies have found the bacteria present in household dust are predominately gram-positive (Górny and Dutkiewicz 2002; Täubel et al. 2009; Kärkkäinen 2009), of which this study agrees with Firmicutes and Actinobacteria present in the greatest abundance. A unique finding in our results that was not well reported before is the large proportion of Cyanobacteria found in the house dust samples. This group of bacteria is the only one that can perform photosynthesis and is well documented in outdoor environments (Dworkin et al. 2007). The presence of this group helps to support the notion that some bacteria in household dust are carried in from outside the home.

Further, the higher abundant genera present in our dust samples, *Streptococcus* and *Corynebacterium*, is similar to that found in two studies using culture-independent methods (Rintala et al. 2008; Täubel et al. 2009) but not studies using culture based methods (Górny and Dutkiewicz 2002; Horak et al. 1996). However, culture-based methods are limited in their ability to identify organisms that are not expected to be found. Since our study used culture-independent 16S rDNA sequencing, it is not surprising that our results resemble the culture-independent studies over the culture-based studies.

Comparing the stool sequencing results to other studies is more difficult since few studies have used the 3 month time point and the infant gut microbiome changes considerably each month of succession. However, similarities were found with *Bifidobacterium* where we found a similar abundance in our samples to Yap et al, which found 49.5% abundance in their 3 month samples (Yap et al. 2011). Other studies have also found *Bifidobacterium* to be the dominant species at the same time frame (Adlerberth et al. 2006; Orrhage and Nord 1999). *Escherichia* is suspected of being an early colonizer with reduced presence as more obligate anaerobes begin to colonize the infant gut (Adlerberth et al. 2006; Mackie et al. 1999). It was interesting to note that the *Escherichia* genus was second highest in abundance in our samples. It should be expected that this genus would be represented in smaller amounts since many more obligate anaerobes were present. Another difference was found with the genus *Ruminococcus*; two other culture-independent studies noted decreased abundance of *Bifidobacterium* and increased abundance of *Ruminococcus* (Adlerberth and Wold 2009; Favier et al. 2002). *Ruminococcus* did not even appear as a genus that was present greater than 1% abundance in our stool samples.

Differences between communities

The results reveal vastly different microbial communities between the dust and stool samples. The most obvious difference is displayed by the number of OTUs identified; with dust having much more than the stool samples. This isn't entirely surprising since the infant gut microbiota is still developing at 3 months of age while the bacterial communities in household dust can represent living or dead microbes that may have been present for months/years. Additionally the dust environment is exposed to more organisms through open windows and movement by residents and visitors than the infant gut which is only exposed to microbes ingested by the infant.

At the phylum level, the two communities do not appear to be vastly different since they share large proportions of the same phyla: Actinobacteria, Firmicutes, and Proteobacteria. However, when looking deeper to the genus level the differences become much more apparent. As suggested by the larger number of OTUs, the variety of genera in dust is greater than in dust. Also, stool samples have a very large proportion of one genus, *Bifidobacterium*, while dust has a greater spread of proportions across numerous genera.

NMDS also clearly demonstrates the differences between the communities. Whether before or after the abundance cut-off or using OTU or genus level data, the ordination plots produced clearly show 2 distinct clusters representing dust and stool samples. The surprising outcome from the NMDS was the splitting of stool into 2 subgroups by *Bifidobacterium* abundance with the genus reduced data. Certain environmental factors have been linked to the increased colonization of *Bifidobacterium* in infants, such as decreased colonization with *E. coli*, streptococci, *Bacteroides*, and *Clostridum* and type of diet (breast-fed versus formula-fed) (Dworkin et al. 2007). Unfortunately, our N is too small to determine an environmental factor that is responsible for this difference, but future research with larger samples sizes should look into these environmental factors and/or if these 2 groups have differing health outcomes.

The percent of variation explained by the NMDS ordination plots doubles when using the data reduced to the genus level compared to just OTUs. This implies that there is a greater variation between OTUs and by condensing down to the genus-level we may be losing that variation when interpreting the data. For instance, two OTUs classified as genus *Clostridium* may actually come from different sources and be quite variable, but when the data is reduced to the genus they are considered to be the same thus taking away the variability those OTU's possess.

Association between communities

While there are clear differences between the two communities, there are certain OTUs that co-occur frequently between dust and stool samples of the same subject implying an association may exist. There were more OTUs shared between the samples before the abundance cut-off, but the permutation test with either set of data show that our data have significantly more shared OTUs between samples than predicted by random assortment. This is important because it is the first time a significant association has been found between the microbiomes of household dust and stool from the inhabitants of the house.

While the permutation test shows an association between the communities, it is not able to provide a direction for the association. The taxonomic classification of the shared OTUs can help to determine where the direction of association might be going. For example, before the abundance cut-off, 3 OTUs identified as genus *Finegoldia*, *Lactococcus*, and *Veillonella* were shared frequently. These three genera are all found in the oral cavities of human and animals and not frequently found in soil. Further both *Finegoldia* and *Veillonella* are obligate anaerobes (Dworkin et al. 2007) suggesting that it is unlikely the originated from the soil and were transferred to the baby orally. It is more likely that the child's mouth was the source of these genera in the dust. However, those three OTUs were no longer some of the highest shared OTUs after the abundance cut-off was performed; the resulting OTUs are not as clear to determine a direction of association.

Escherichia is a facultative anaerobe that is commonly found in the gastrointestinal tract of humans and animals and has also been found in soil and water (Dworkin et al. 2007). Given that this organism can grow in the presence of oxygen and has been found in soil, this may suggest that the association could be potentially go from dust to baby. However, generally the presence of this genus in the environment is considered to reflect fecal contamination and not the ability to replicate feely outside the intestine (Dworkin et al. 2007). Two other OTU classifications are common in the gastrointestinal tract of humans, Lachnospiraceae family and the genus *Bifidobacterium*. Unlike *Escherichia* these OTUs are obligate anaerobes and not likely to grow in dust due to the presence of oxygen. As a result, it can be assumed that the direction of association for these is probably from baby to the dust. *Streptococcus* is another facultative anaerobe that is predominately found in the mouth and upper respiratory tract of humans, but not exclusively as it has been isolated from feces & soil (Dworkin et al. 2007). Therefore this OTU could have grown in the presence of oxygen in dust and then transferred to the baby orally.

It appears that the majority of the OTU classification information suggests the direction of association is from baby to dust, but it's not definitively clear. Not all are obligate anaerobes and some

can be found in soil environments. Again, this study has a small sample size, so a larger sample size will help to determine the OTUs that are truly shared frequently between stool and dust samples. Additionally, a different study altogether will be needed to identify the true direction of association, one that cannot be completed with the current data collected in the CHILD study.

Effect of using more conservative data

The abundance cut-off was applied to prevent sequence errors or contamination from being incorporated into the data set. However, by removing all OTUs that have abundance below a set value, we are not only removing ones created by error but also OTUs that are truly rare in our communities. The rare OTUs accounted for 75% of our entire data set, so we wanted to see how that changes analysis results. Thus we performed analyses with both sets of data to determine if the rare OTUs had a significant effect on our results.

Based on the results presented, the rare OTUs did not have a significant impact on the overall outcome. There were some changes in proportions of phyla and genera as seen in Figures 1-4 and a slight change in the NMDS ordination plots, however the differences between the dust and stool communities was so great that these slight changes did not affect the interpretable result. The greatest change came in the number of shared OTUs between samples. The conservative data resulted in 5 subjects not having any shared OTUs at all and the average shared OTU number to drop from 5.2 to 1.35. Even though there was a large drop in the number of shared OTUs, the permutation tests revealed they were still significantly higher than expected by random.

Conclusion

Although the sample size in this pilot project was small, we found substantial general differences between stool and dust microbial communities. Nevertheless, the co-occurrence of OTUs in paired dust and stool samples was significantly greater than would be predicted by random assortment, suggesting an association between these two communities. Further research should investigate the causality of dust borne gut microbes and determine if these repositories influence gut microbial succession in early life. Additionally, while an abundance cut-off removes a large amount of data, these rare OTUs appear to have minimal affect on the overall outcome of analyses.

Figures/Tables

Figure 1. Phylum composition before the abundance cut-off

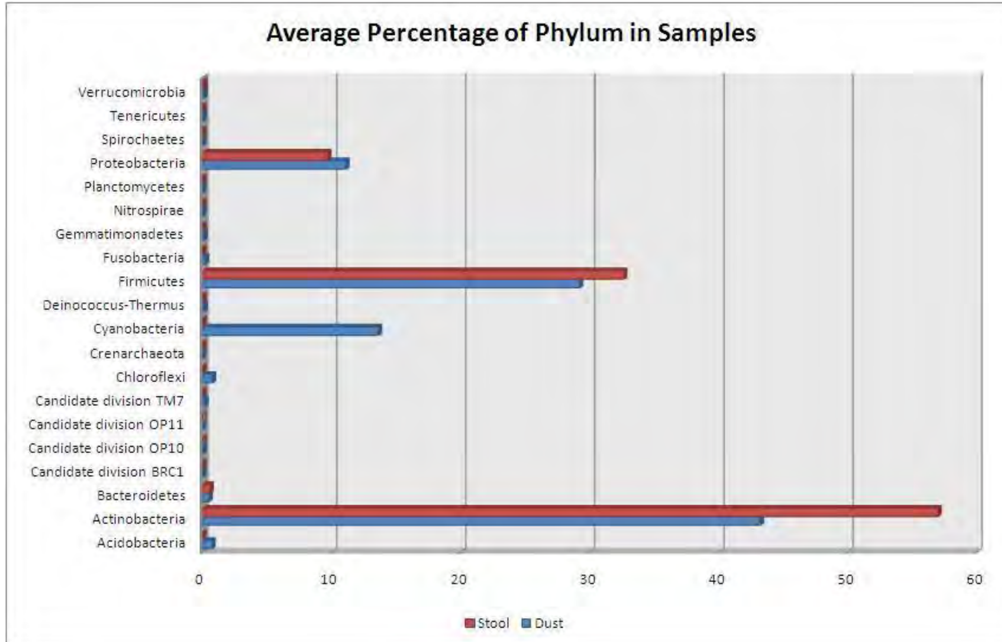

Figure 2. Phylum composition after the abundance cut-off

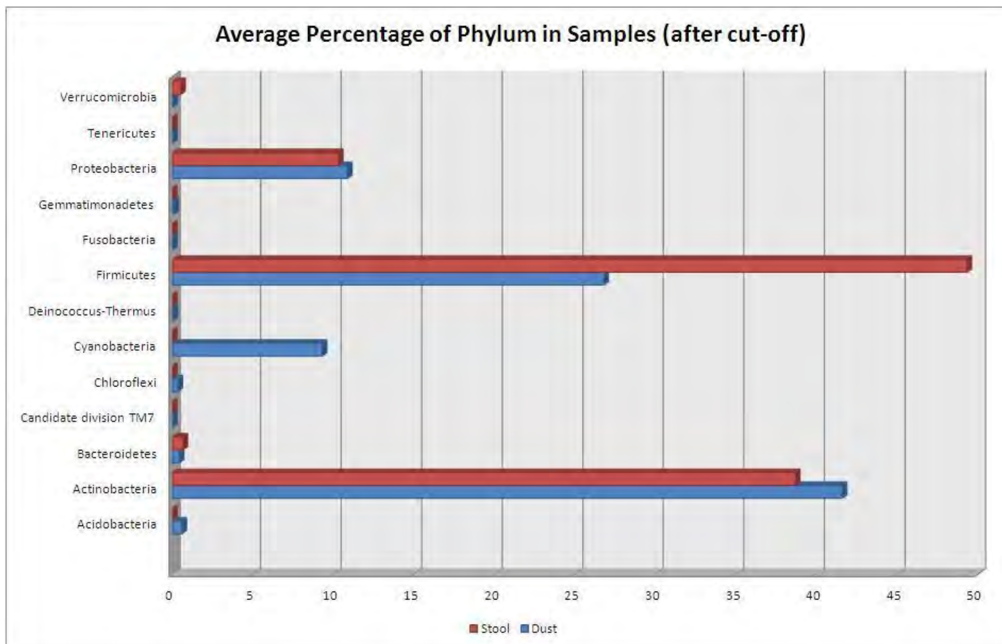

Figure 3. Genus composition before the abundance cut-off

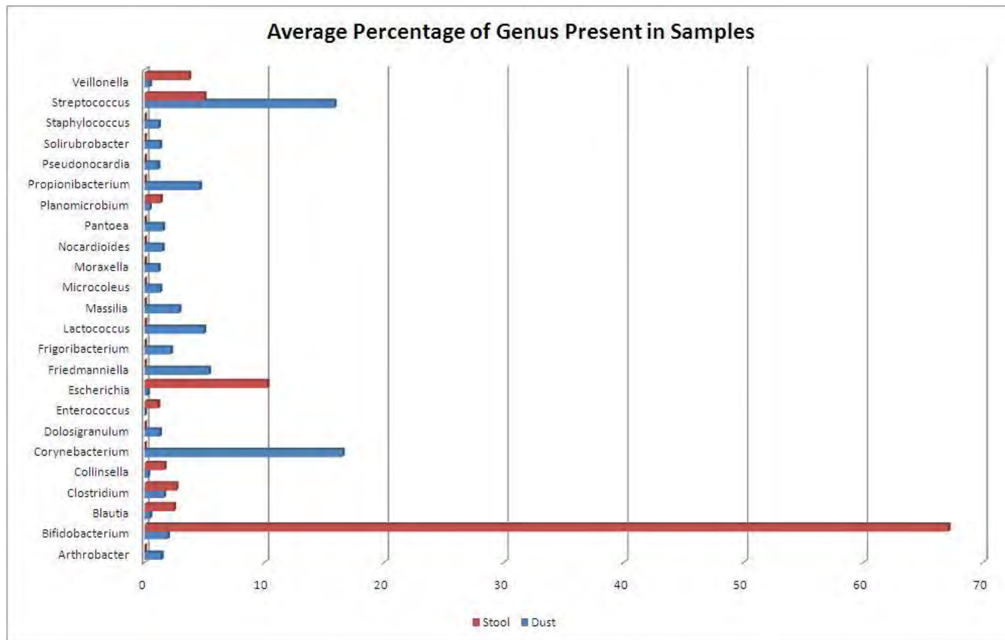

Figure 4. Genus composition after the abundance cut-off

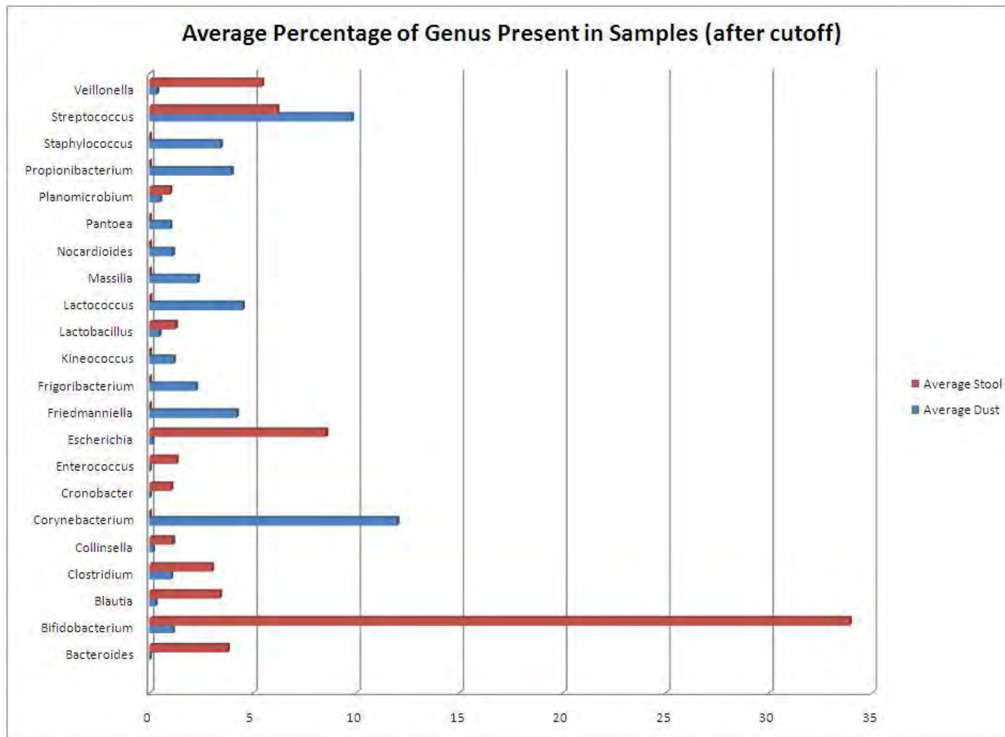

Figure 5. OTU level NMDS before the abundance cut-off

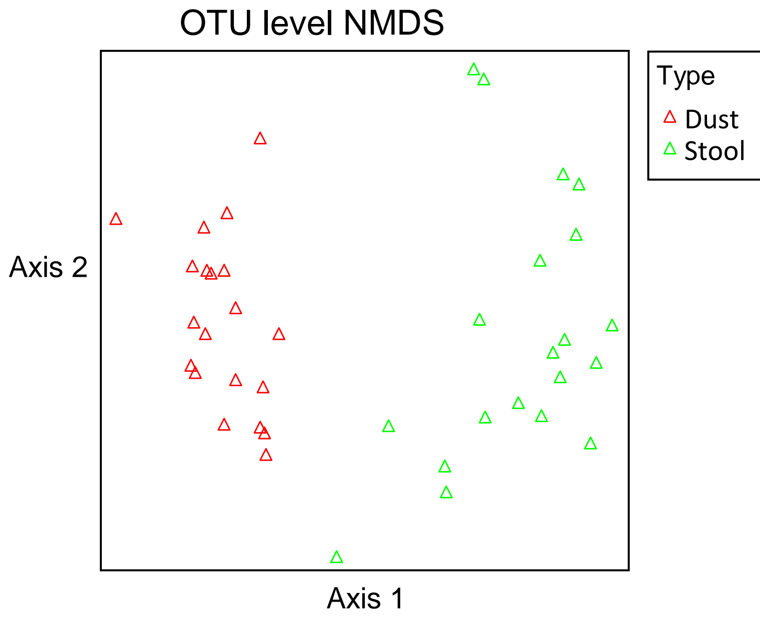

Figure 6. OTU level NMDS after the abundance cut-off

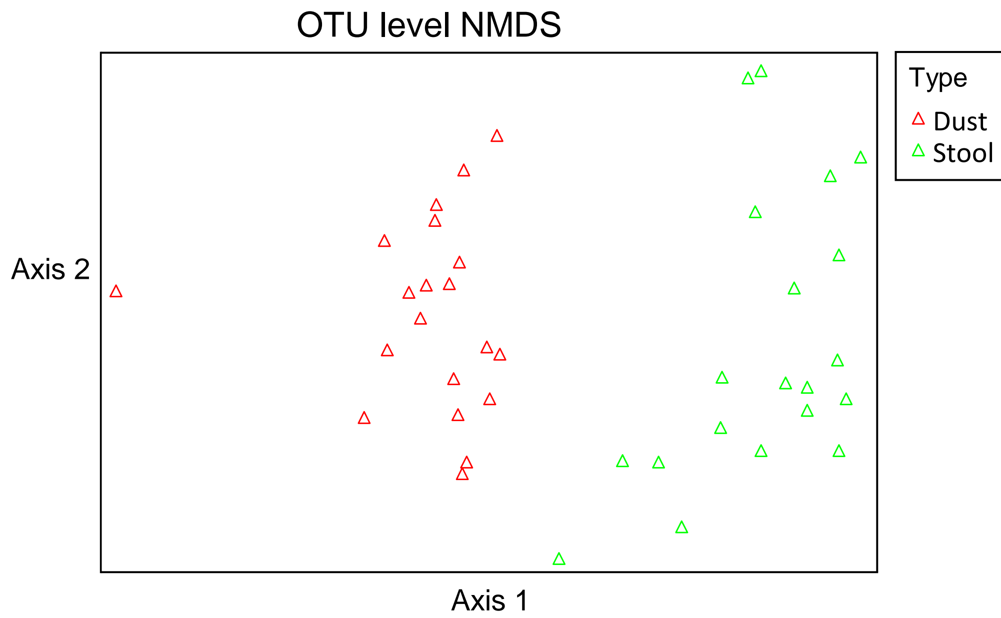

Figure 7. Genus level NMDS before the abundance cut-off

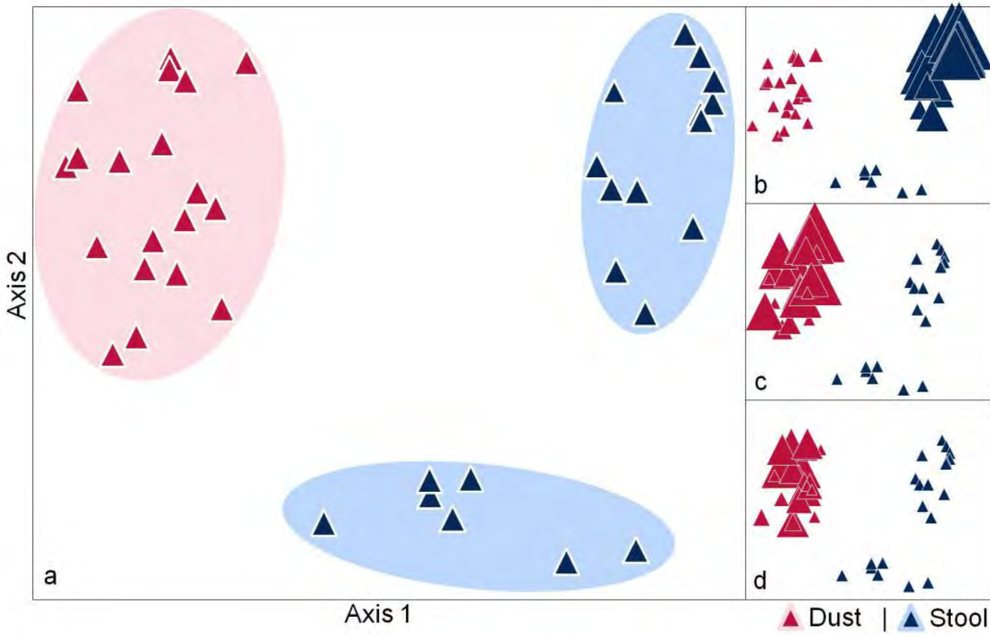

Figure 8. Genus level NMDS after the abundance cut-off

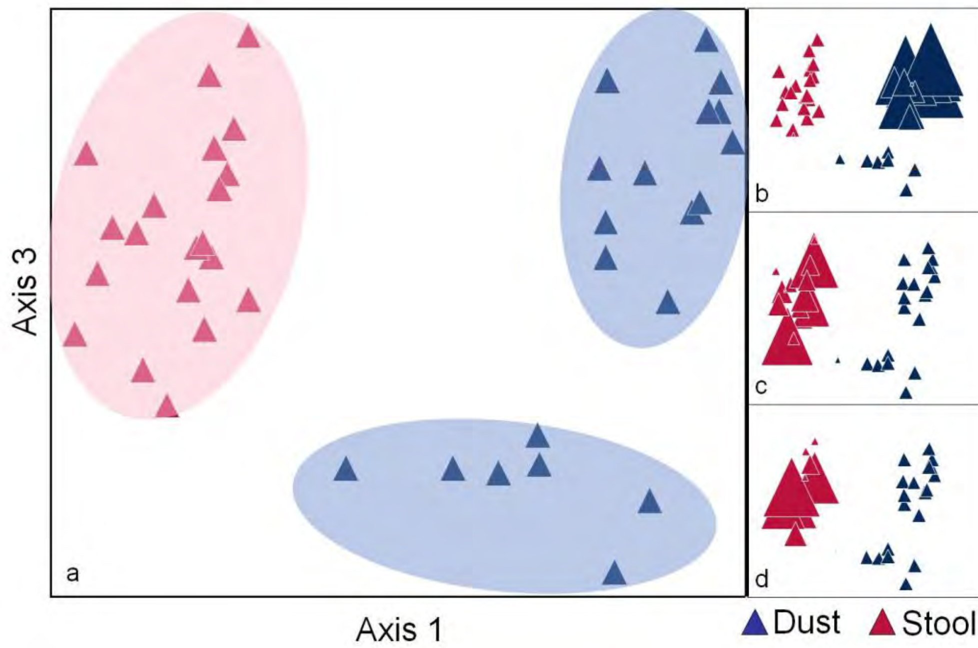

Figure 9. Permutation test distribution before the abundance cut-off

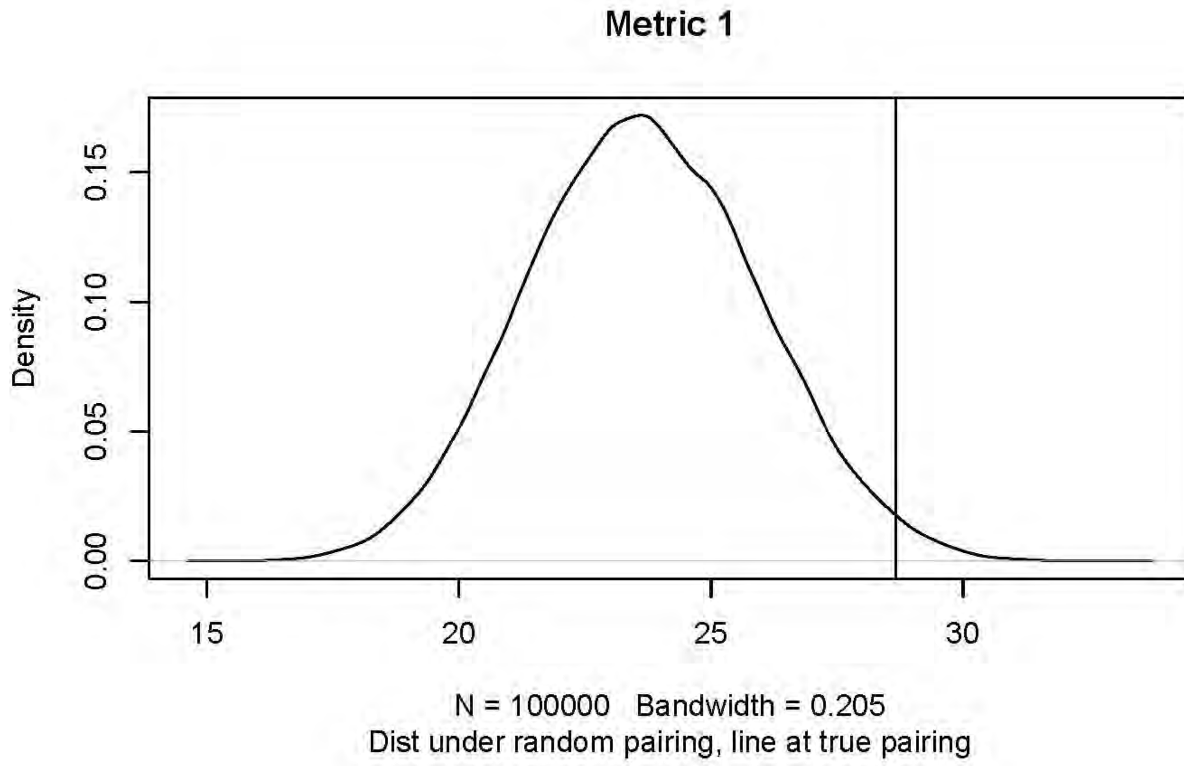

Figure 10. Permutation test distribution after the abundance cut-off

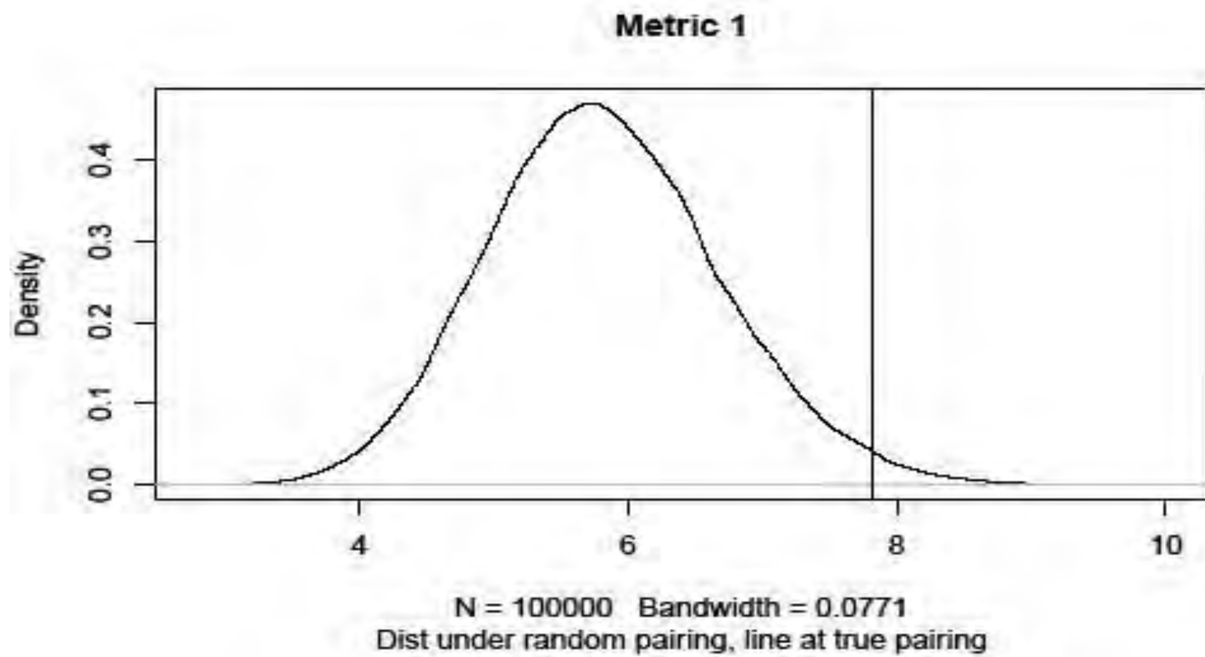

Table 1. Taxonomic Classification of Most Shared OTUs

	OTU	# Samples	Phlyum	Class	Order	Family	Genus
Before the cutoff							
	167480	10	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus
	159036	9	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia
	170861	7	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
	60526	5	Firmicutes	Clostridia	Clostridiales	Clostridiales Family XI Incertae Sedis	Finegoldia
	57112	4	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Veillonella
After the cutoff							
	159036	5	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia
	170861	5	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
	426	3	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	
	48	2	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	
	196	2	Actinobacteria	Actinobacteria	Actinobacteridae	Bifidobacteriaceae	Bifidobacterium
	32660	2	Actinobacteria	Actinobacteria	Actinobacteridae	Bifidobacteriaceae	Bifidobacterium

Table 2. Permutation Test Results

	Score	
	Before cut-off	After cut-off
Value under actual pairing	28.67	7.81
Estimated value under random pairing in permutations test		
Mean (SD):	23.67 (2.28)	5.82 (0.86)
95% interval:	(19.34, 28.16)	(4.24, 7.62)
Prob (true>random):	0.9859	0.9836
p-value (1-sided):	0.0141	0.0164
p-value (2-sided):	0.0282	0.0328

References

Abrahamsson TR et al. 2012. Low diversity of the gut microbiota in infants with atopic eczema. *Journal of Allergy & Clinical Immunology*. 129(2): 434-440.

Adlerberth I and Wold A. 2009. Establishment of the gut microbiota in Western infants. *Acta Paediatrica*. 98: 229-238.

Adlerberth I et al. 2006. Reduced enterobacterial and increased Staphylococcal colonization of the infantile bowel: An effect of hygienic lifestyle?. *Pediatric Research*. 59(1): 96-101.

Apajalahti J. 2005. Comparative gut microflora, metabolic challenges, and potential opportunities. *Journal of Applied Poultry Research*. 14: 444-453.

Baldwin HS, Smart KM, Kahn MC, and Dworetzky M. 1957. Aerobic bacteria in house dust. *Journal of Allergy*. 28(2): 179-181.

Bisgaard H et al. 2011. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *Journal of Allergy & Clinical Immunology*. 128(3): 646-652.

Bouillard L, Michel O, Dramaiz M, and Devleeschouwer M. 2005. Bacterial contamination of indoor air, surfaces, and settled dust, and related dust endotoxin concentrations in healthy office buildings. *Annals of Agricultural & Environmental Medicine*. 12: 187-192.

Dworkin MM et al (eds). 2007. Prokaryotes: A handbook on the biology of bacteria. Vol 3. New York: Springer.

Fanaro S, Chierici R, and Vigi V. 2003. Intestinal microflora in early infancy: composition and development. *Acta Paediatrica Supplement*. 441: 48-55.

Favier C, Vaughan E, Willem M, and Akkermans A. 2002. Molecular monitoring of succession of bacterial communities in human neonates. *Applied and Environmental Microbiology*. 68(1): 219-226.

Garrod LP. 1944. Some observations on hospital dust. *British Medical Journal*. 1(4337): 245-247.

Gehring Y et al. 2007. Bacteria and mould components in house dust and children's allergic sensitisation. *European Respiratory Journal*. 29: 1144-1153.

Górny RL and Dutkiewicz J. 2002. Bacterial and fungal aerosols in indoor environment in central and eastern european countries. *Annals of Agricultural & Environmental Medicine*. 9: 17-23.

Horak B, Dutkiewicz J, and Solarz K. 1996. Microflora and acarofauna of bed dust from homes in Upper Silesia, Poland. *Annals of Allergy, Asthma, & Immunology*. 76: 41-50.

Hyvärinen, A et al. 2006. Characterizing microbial exposure with ergosterol, 3-hydroxy fatty acids, and viable microbes in house dust: Determinants and association with childhood asthma. *Archives of Environmental & Occupational Health*. 61(4): 149-157.

Ismail IH et al. 2011. Prenatal administration of Lactobacillus rhamnosus has no effect on the diversity of the early infant gut microbiota. *Pediatric Allergy and Immunology*. Dec 4 online.

Johansson MA, Sjögren YM, Persson J-O, and Nilsson C. 2011. Early colonization with a group of Lactobacilli decreases the risk for allergy at five years of age despite allergic heredity. *PLoS One*. 6(8): e23031.

Kärkkäinen PV, Valkonen M, Hyvärinen A, Nevalainen A, and Rintala H. 2009. Determination of bacterial load in house dust using qPCR, chemical markers and culture. *Journal of Environmental Monitoring*. 12: 759-768.

Koenig JE et al. 2011. Succession of microbial consortia in the developing nifant gut microbiome. *PNAS*. 108(Suppl 1): 4578-4585.
Lee L, Tin S, and Kelley ST. 2007. Culture-independent analysis of bacterial diversity in a child-care facility. *BMC Microbiology*. 7: 27.
Lidwell OM and Lowbury EJ. 1950. The survival of bacteria in dust. *The Journal of Hygiene*. 48(1): 6-20.
Mackie R, Sghir A, and Gaskins H. 1999. Developmental microbial ecology of the neonatal gastrointestinal tract. *American Journal of Clinical Nutrition*. 69(Suppl): 1035S-1045S.
Maier RM et al. 2010. Environmental determinants of an impact on childhood asthma by the bacterial community in household dust. *Applied and Environmental Microbiology*. Apr: 2663-2667.
Martin R et al. 2010. Early life: Gut microbiota and immune development in infancy. *Beneficial Microbes*. 1(4): 367-382.
McCune B and Mefford MJ. 2011. PC-ORD: Multivariate analysis of ecological data. Version 6.0. MjM Software: Gleneden Beach, OR.
Nakayama J et al. 2011. Aberrant structures of fecal bacterial community in allergic infants profiled by 16S rRNA gene pyrosequencing. *FEMS Immunology & Medical Microbiology*. 63: 397-406.
Nevas M et al. 2005. Infant Botulism acquired from household dust presenting as sudden infant death syndrome. *Journal of Clinical Microbiology*. 43(1): 511-513.
Orrhage K and Nord C. 1999. Factors controlling the bacterial colonization of the intestine in breastfed infants. *Acta Paediatrica Supplement*. 430: 47-57.
O'Toole PW and Claesson MJ. 2010. Gut microbiota: Changes throughout the lifespan from infancy to elderly. *International Daily Journal*. 20: 281-291.
Pakarinen J et al. 2008. Predominance of gram-positive bacteria in house dust in the low-allergy risk Russian Karelia. *Environmental Microbiology*. 10(12): 3317-3325.
Peck J. 2010. Multivariate analysis for community ecologists: Step-by-step using PC-ORD. MjM Software: Gleneden Beach, OR.
Penders J et al. 2007. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA birth cohort. *Gut*. 56: 661-667.
Rintala H, Pitkäranta M, Toivola M, Paulin L, and Nevalainen A. 2008. Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiology*. 8: 56.
Sjögren YM, Jenmalm MC, Böttcher MF, Björkstén B, and Sverremark-Ekström E. 2009. Altered early infant gut microbiota in children developing allergy up to 5 years of age. *Clinical & Experimental Allergy*. 29: 518-526.
Täubel M et al. 2009. The occupant as a source of house dust bacteria. *Journal of Allergy & Clinical Immunology*. 124(4): 834-840.
Torrazza RM and Neu J. 2011. The developing intestinal microbiome and its relationship to health and disease in the neonate. *Journal of Perinatology*. 31: S29-S34.
Vael C and Desager K. 2009. The importance of the development of the intestinal microbiota in infancy. *Current Opinion in Pediatrics*. 21: 794-800.
van Nimwegen FA et al. 2011. Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. *Journal of Allergy & Clinical Immunology*. 128(5): 948-955.
Yap GC et al. 2011. Evaluation of stool microbiota signatures in two cohorts of Asian (Singapore and Indonesia) newbonrs at risk of atopy. *BMC Microbiology*. 11: 193.