Serial Analysis of the Gut and Respiratory Microbiome in Cystic Fibrosis in Infancy: Interaction between Intestinal and Respiratory Tracts and Impact of Nutritional Exposures

J. C. Madan,[a] D. C. Koestler,[b] B. A. Stanton,[c] L. Davidson,[d] L. A. Moulton,[e] M. L. Housman,[c] J. H. Moore,[b] M. F. Guill,[e] H. G. Morrison,[f] M. L. Sogin,[f] T. H. Hampton,[c] M. R. Karagas,[g] P. E. Palumbo,[h] J. A. Foster,[i] P. L. Hibberd,[j] and G. A. O'Toole[c]

Division of Neonatology, Department of Pediatrics, Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire, USA[a]; Computational Genetics, Translational Genomics Research Institute, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA[b]; Department of Microbiology & Immunology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA[c]; Center for Global Health Research, Division of Geographic Medicine and Infectious Diseases, Tufts University School of Medicine, Boston, Massachusetts, USA[d]; Division of Pediatric Pulmonology, Department of Pediatrics, Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire, USA[e]; Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, Massachusetts, USA[f]; Section of Biostatistics and Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA[g]; Section of Infectious Diseases and International Health, Department of Medicine, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA[h]; Department of Biological Sciences and Initiative for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho, USA[i]; and Division of Global Health, Department of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts, USA[j]

ABSTRACT **Pulmonary damage caused by chronic colonization of the cystic fibrosis (CF) lung by microbial communities is the proximal cause of respiratory failure. While there has been an effort to document the microbiome of the CF lung in pediatric and adult patients, little is known regarding the developing microflora in infants. We examined the respiratory and intestinal microbiota development in infants with CF from birth to 21 months. Distinct genera dominated in the gut compared to those in the respiratory tract, yet some bacteria overlapped, demonstrating a core microbiota dominated by *Veillonella* and *Streptococcus*. Bacterial diversity increased significantly over time, with evidence of more rapidly acquired diversity in the respiratory tract. There was a high degree of concordance between the bacteria that were increasing or decreasing over time in both compartments; in particular, a significant proportion (14/16 genera) increasing in the gut were also increasing in the respiratory tract. For 7 genera, gut colonization presages their appearance in the respiratory tract. Clustering analysis of respiratory samples indicated profiles of bacteria associated with breast-feeding, and for gut samples, introduction of solid foods even after adjustment for the time at which the sample was collected. Furthermore, changes in diet also result in altered respiratory microflora, suggesting a link between nutrition and development of microbial communities in the respiratory tract. Our findings suggest that nutritional factors and gut colonization patterns are determinants of the microbial development of respiratory tract microbiota in infants with CF and present opportunities for early intervention in CF with altered dietary or probiotic strategies.**

IMPORTANCE **While efforts have been focused on assessing the microbiome of pediatric and adult cystic fibrosis (CF) patients to understand how chronic colonization by these microbes contributes to pulmonary damage, little is known regarding the earliest development of respiratory and gut microflora in infants with CF. Our findings suggest that colonization of the respiratory tract by microbes is presaged by colonization of the gut and demonstrated a role of nutrition in development of the respiratory microflora. Thus, targeted dietary or probiotic strategies may be an effective means to change the course of the colonization of the CF lung and thereby improve patient outcomes.**

Received 23 July 2012 **Accepted** 26 July 2012 **Published** 21 August 2012

Citation Madan JC, et al. 2012. Serial analysis of the gut and respiratory microbiome in cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and impact of nutritional exposures. mBio 3(4):e00251-12. doi:10.1128/mBio.00251-12.

Editor Frederick Ausubel, Massachusetts General Hospital

Address correspondence to George A. O'Toole, georgeo@dartmouth.edu.

J.C.M., P.L.H., and G.A.O. contributed equally to this research as co-senior authors.

Cystic fibrosis (CF) is the most common life-limiting autosomal recessive disease in Caucasians and causes multisystem complications, with progressive lung disease, pancreatic insufficiency, and a profound impact on growth and nutrition (1, 2). Mutations in CFTR are known to alter the airway and intestinal microenvironment, and patterns of bacterial colonization are different in infants, children, and older adults with CF than in subjects without CF (3–11). The pattern of respiratory and enteric colonization in children with CF is thought to differ from that in children without CF beginning in infancy, and colonization with pathogens including *Staphylococcus aureus* and *Pseudomonas aeruginosa* is associated with worse long-term lung function, more frequent exacerbations, and earlier morbidity and mortality (1, 9, 12, 13). Recent CF porcine studies have identified that defective bacterial clearance occurs soon after birth, with early colonization and impaired eradication resulting in the inflammatory damage of

lung tissue (14). Chronic colonization of the CF lung by antibiotic-resistant microbial communities results in progressive respiratory failure and is the leading cause of mortality in these patients (1, 13, 15). Because delaying chronic pathogen colonization in CF improves clinical outcomes, it is important to understand early patterns of colonization to inform future strategies to prevent or modify the acquisition of pathogenic colonization (12, 16, 17).

Long-term morbidity and mortality in CF are closely related to suboptimal nutritional status and growth failure, which further increase the risk of morbidity and mortality (2, 18), and this nutritional status likely impacts, and is impacted by, the microbiota. Furthermore, twin studies of CF show that approximately 50% of disease progression is explained by genetic factors, and approximately 50% is explained by environmental and/or stochastic effects that have yet to be further elucidated (11). We hypothesize that a significant contribution to variable disease expression in patients with CF who are genetically identical is due to differences in microbial acquisition and that acquisition of the gut microbiome may impact the acquisition of the lung microbiome in infants with CF.

The purpose of this study was to investigate the developing respiratory and intestinal microbiome in infants with CF, beginning in the first month of life, using culture-independent 454 pyrosequencing of the bacterial 16S rRNA gene. We studied seven infants approximately every 3 months up to the first 21 months of life and compared the acquisition of the microbiome in the gut and respiratory tract as the infants were exposed to breast milk and infant formula and introduced to solid foods. The ultimate goal of this work is to understand how to promote and maintain a microbiome that prevents or resists colonization with pathogenic bacteria, maintain a more health-promoting microbiota in the gut and the respiratory tract for a longer period of time, and decrease morbidity and mortality in patients with CF.

RESULTS

Patient cohort. Seven subjects with CF (diagnosed by newborn screen, sweat chloride test, and confirmatory genetic testing) were enrolled in the study within a month of birth. Subjects were followed for a total of 9 to 21 months beginning at birth, with samples collected at regularly scheduled CF clinic visits at <1 month of age and at 3-month intervals serially. Respiratory samples were obtained using an oropharyngeal swab, a method that has been reported to accurately reflect the microbial populations in the respiratory tract (19), including assessing species in a culture-independent evaluation of respiratory microbes (20). In particular, this method is the preferred sampling method for infants who do not produce sputum and do not undergo bronchoalveolar lavage (BAL). Stool served as the source of gastrointestinal samples.

All infants were born in a hospital and hospitalized for 2 days to 7 weeks after birth. Of the seven subjects, one (subject 4) was delivered at 31 weeks' gestation and spent 7 weeks in neonatal intensive care. The remaining six subjects were delivered at full term. Two subjects (subjects 3 and 4) had meconium ileus immediately after birth and required surgery in the first week of life. Overall, 59 samples were collected and analyzed: 33 intestinal samples and 26 respiratory samples.

Eight genera dominate distinct microbial communities of the upper respiratory tract and gut. 16S rRNA gene pyrosequencing of the 59 study samples was conducted, yielding a total of 1,205,894 sequences, which comprised ~1,500 unique operational taxonomical units (OTUs) (see Table S1 in the supplemental material). Among these sequences, which collectively represented 79 unique genera, 74% and 72% of the total number of reads across samples were represented by 8 genera for the intestinal and respiratory microbiome samples, respectively (Fig. 1). *Streptococcus*, *Veillonella*, and *Prevotella* dominated the respiratory samples, comprising nearly 60% of the total number of reads across the respiratory samples (Fig. 1b). *Bacteroides*, *Bifidobacterium*, and *Veillonella* were the most abundant genera in the intestinal tract, collectively comprising 40% of the total number of reads (Fig. 1d).

We evaluated both the most abundant microbes in the intestinal and respiratory tracts and the inter- and intraindividual variation in the microbial composition and diversity within these two sampling locations. As depicted in Figure 1a and c, there is a pattern of lower interindividual variability in the normalized abundance of the eight most dominant genera within the respiratory tract than in normalized bacterial abundance within the intestinal tract.

We more formally examined this observation by computing and comparing the intraclass correlation coefficient (ICC) between the intestinal and respiratory tracts (Table 1). The ICC is a statistical measure that captures the intrasubject similarity in normalized microbial abundance. The values of ICC are bounded between 0 and 1, with values approaching 1 signifying that a greater proportion of the total variation in normalized microbial abundance is accounted for by between-subject variation (intersubject variation). In this way, the ICC can also be interpreted as the proportion of total variance in normalized microbial abundance that is "between subjects." The pairwise comparison of ICCs for all 77 bacterial genera between the intestinal and respiratory tracts revealed significantly greater ($P = 0.02$) ICCs in the intestinal tract than in the respiratory tract. These data indicate that the respiratory microbiome may be more consistent across the patients than the intestinal populations.

Identifying a core bacterial cluster. Hierarchical clustering of the 59 samples based on the $\log_{10}$-normalized abundance of the 79 genera identified here revealed that the intestinal and respiratory samples clustered exclusively into two discrete clusters (Fig. 2, left; see also Fig. S1 in the supplemental material). A common core of bacteria are present throughout all sample time points and both sample types (respiratory and intestinal), with a dominance of *Streptococcus* and *Veillonella* ($P = 0.05$) (Fig. 2, box 1). There are two significant clusters (Fig. 2, boxes 2 and 4) that are highly represented in the respiratory tract but not the gut. These data indicate that while a small group of organisms is present between the respiratory tract and the gut, these compartments are largely distinct with respect to their microbial communities.

Select genera in the gut presage microbial populations in the respiratory tract. Among the subset of microbes present in the respiratory tract and gut, we noticed a high degree of concordance between the bacterial genera that were increasing or decreasing over time in both systems. That is, 14 of 16 genera that were increasing in the gut were also increasing in the respiratory tract ($P < 0.001$) (Fig. 3A). These data suggest that while the microbiome of the respiratory tract and gut are distinct (Fig. 1 and 2), there are some microbes common to both sites whose populations change similarly over time. Importantly, the respiratory tract changes indicate a large increase in several bacterial genera, the majority of

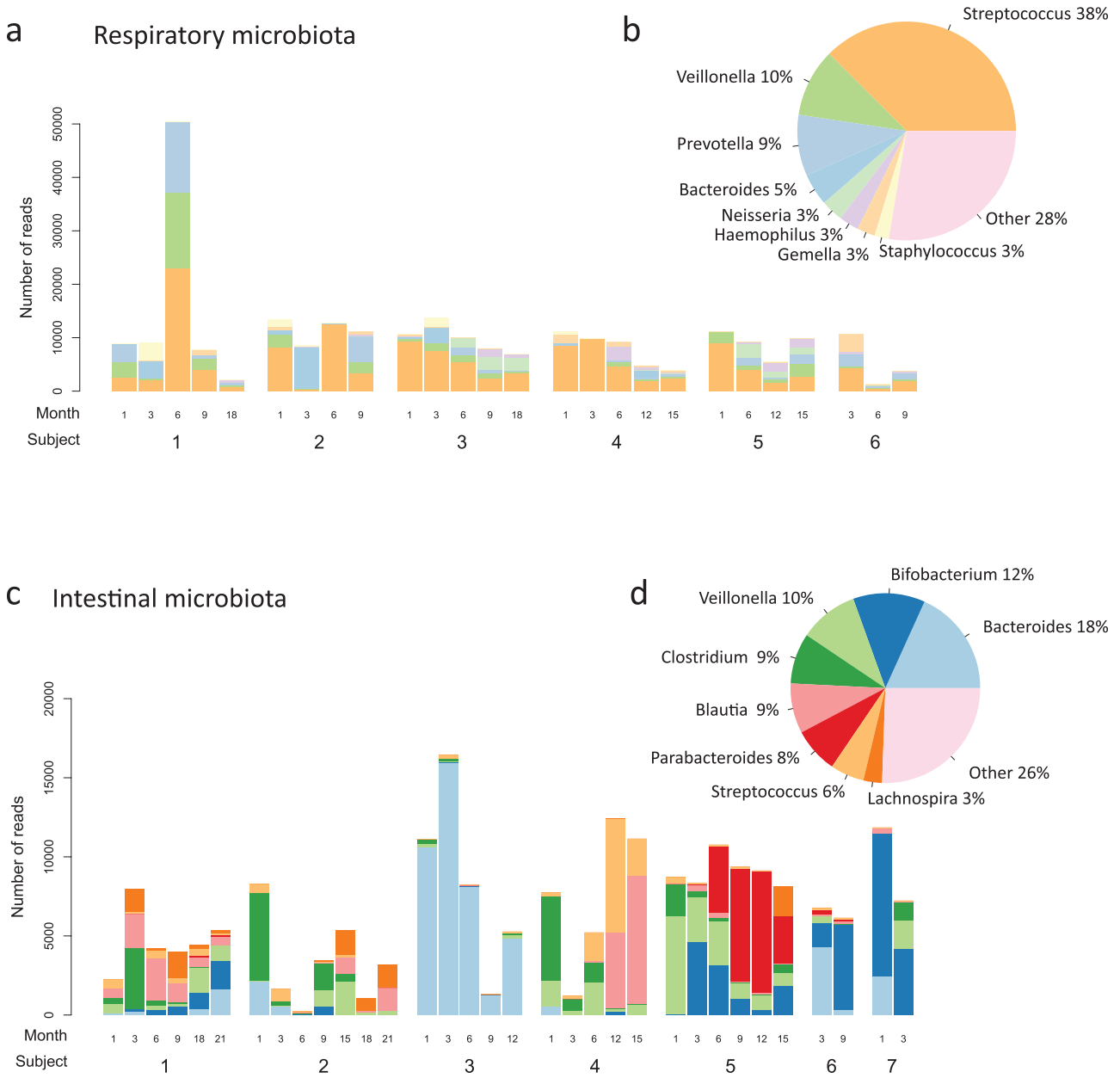


FIG 1 Bacterial normalized abundance of the respiratory and intestinal microbiomes. (a) Stacked bar plot representing bacterial genera from serial respiratory microbiome samples from the 7 subjects. The *y* axis represents the number of reads for particular bacterial genera; the *x* axis represents subject numbers (1 to 7) and serial samples at progressive months of life. (b) Pie chart of the eight most abundant genera within the respiratory microbiome. (c) Stacked bar plot representing bacterial genera from serial intestinal microbiome samples. The *y* axis represents the number of reads for particular bacterial genera; the *x* axis represents subject numbers (1 to 7) and serial samples at progressive months of life. (d) Pie chart of the eight most abundant genera within the intestinal microbiome.

which are known to populate the intestines (including *Escherichia*, *Parabacteroides*, and *Bifidobacterium*; see Fig. S2 in the supplemental material). Changes in the relative normalized abundance of bacteria in the intestines reveal a decrease in *Staphylococcus* and *Clostridium* and an increase in several of the bacteria that also increase in the respiratory tract, including *Escherichia* and *Parabacteroides* (see Fig. S2).

Further inspection of the hierarchical clustering analysis revealed a group of microbes present in samples from early life in the intestines that are also present later in life in the respiratory tract (Fig. 2, box 3, compare inset a to inset b). This group of bacteria includes *Enterococcus*, *Coprococcus*, *Escherichia*, and *Parabacteroides*. For 7 genera of bacteria, plotting the populations over time indicates that gut colonization occurs prior to their appearance in the respiratory tract (Fig. 3B). Thus, the microbial populations are dynamic over time, and a subset of genera is detected in the gut prior to their appearance in the respiratory tract.

Microbial diversity increases over time in both the respiratory and intestinal tracts. Microbial diversity, assessed using the Simpson's diversity index (SDI) (21), showed subtle but increas-

TABLE 1 Interclass correlation coefficients calculated for respiratory and intestinal tracts

Sample type	Genus	Proportion of total no. of reads (%)	ICC[a]
Respiratory	*Bacteroides*	18	0.54
	Bifidobacterium	12	0.61
	Veillonella	10	0.29
	Clostridium	9	<0.01
	Blautia	9	0.36
	Parabacteroides	8	0.60
Intestinal	*Streptococcus*	38	<0.01
	Veillonella	10	0.23
	Prevotella	9	0.05
	Bacteroides	5	<0.01
	Neisseria	3	0.30
	Haemophilus	3	<0.01

[a] ICC, intraclass correlation coefficient. The values of ICC are bounded between 0 and 1, with values approaching 1 signifying that a greater proportion of the total variation in microbial normalized abundance is accounted for by between-subject variation (intersubject variation). The ICC can also be interpreted as the proportion of total variance in microbial normalized abundance that is "between subjects."

ing trends across time among both the respiratory and intestinal tracts (Fig. 4a and b). We further note that the trends in increasing microbial diversity appeared to be more pronounced in the respiratory tract (Fig. 4a) than in the intestine (Fig. 4b). We used a linear mixed-effect model to examine the association between time and microbial diversity when comparing the respiratory and intestinal tracts. The results from this model revealed that while the microbial diversity of both the respiratory and intestinal tracts increased significantly over time ($P = 2e{-}4$ and $P = 0.04$, respectively), the diversity of the respiratory tract increased more acutely over time than that of the intestinal tract ($P = 0.07$) (Fig. 4c). Thus, while both sites are dynamic, the rate of change in the respiratory tract appears to be greater than that observed in the gut in this cohort.

Characterizing effect of exposures on microbiome development. We next assessed the impact of nutritional exposures on overall diversity and on particular subgroups of microbes. An analysis of the relationship between nutrition and overall diversity revealed that nonsignificant diversity trajectories occurred with

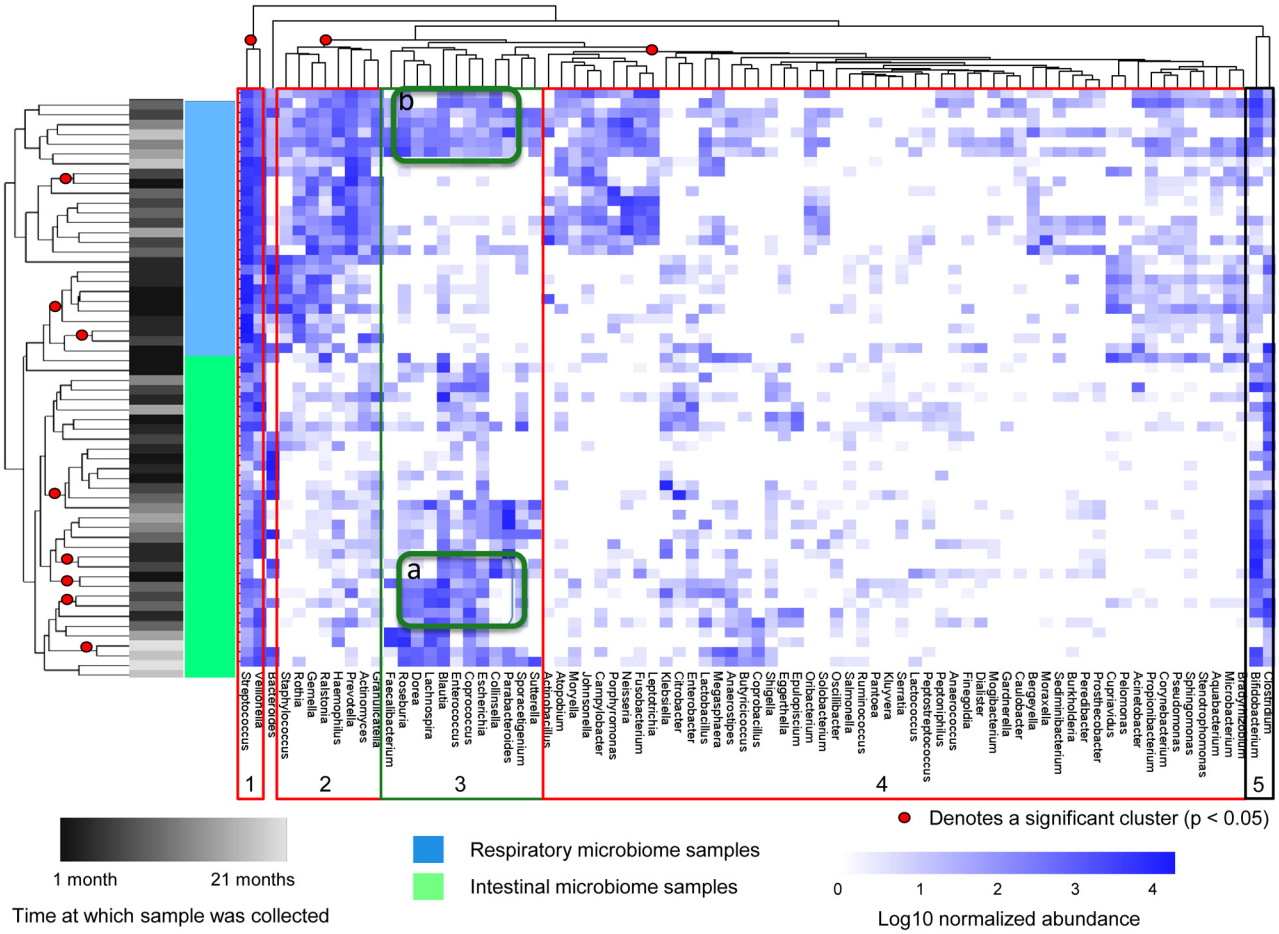


FIG 2 Heat map and SDI of the intestinal and respiratory microbiome samples. Heat map based on the hierarchical clustering solution (Euclidean distance metric and average linkage) of the 59 intestinal and respiratory microbiome samples. Rows represent samples (subjects and respective time points at which the samples were collected), columns represent genera, and the values in the heat map represent the $\log_{10}$-normalized numbers of sequencing reads, with increasing grades of purple representing greater relative abundance. Box 1, core microbiome crossing respiratory and intestinal samples across all time points ($P < 0.05$). Boxes 2 and 4, microbes more abundantly represented in the respiratory microbiome. Box 3, microbes highlighted as having increased normalized abundance in the intestines early in life (inset a) and in the respiratory tract later in life (inset b), highlighted by a dark-green outline. Box 5, genera more highly represented in the intestinal samples. Pvclust (http://bioinformatics.oxfordjournals.org/content/22/12/1540.full), a method for assessing the uncertainty in hierarchical cluster analysis via multiscale bootstrap resampling, was used to determine the significance of identified clusters.

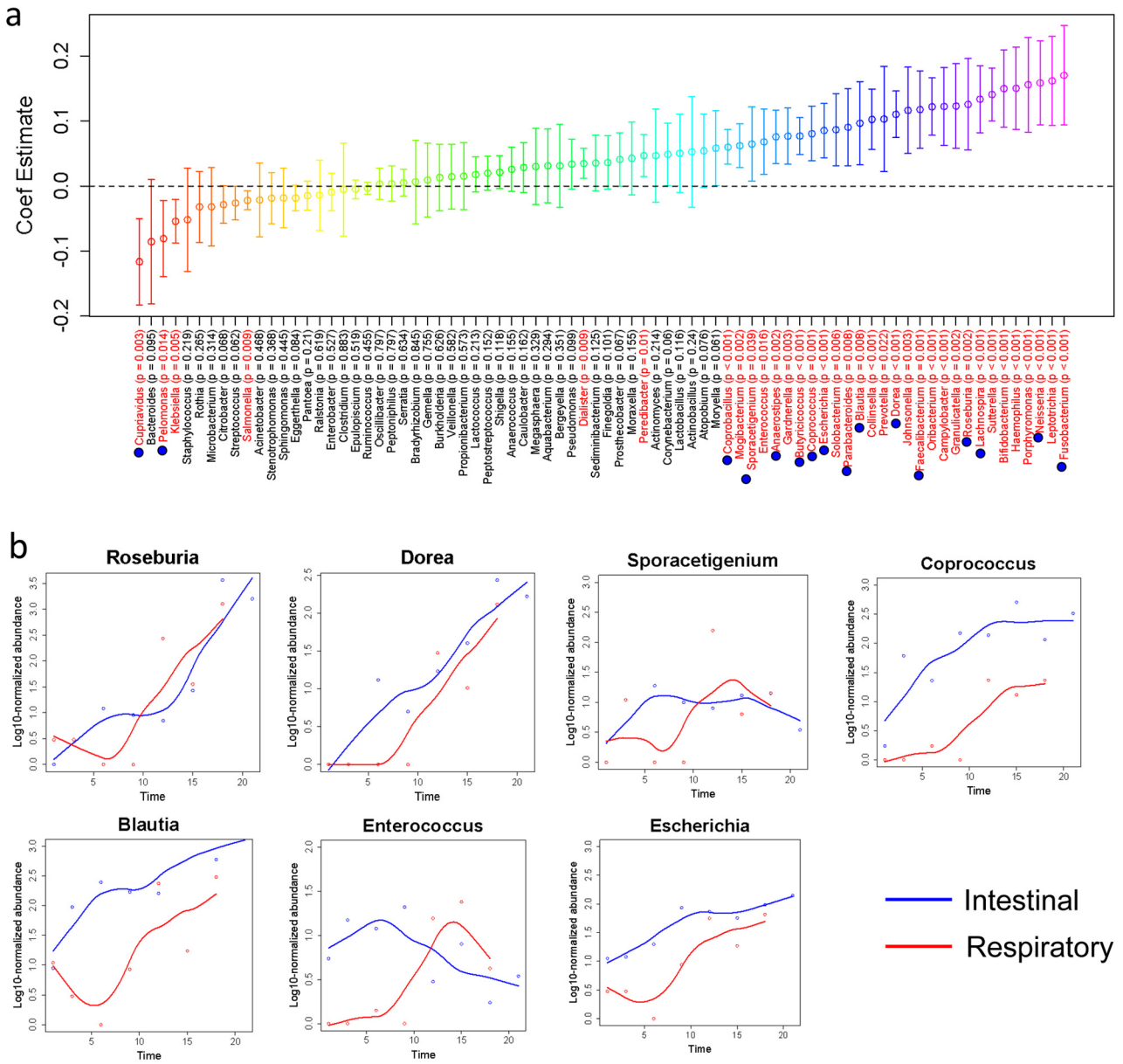
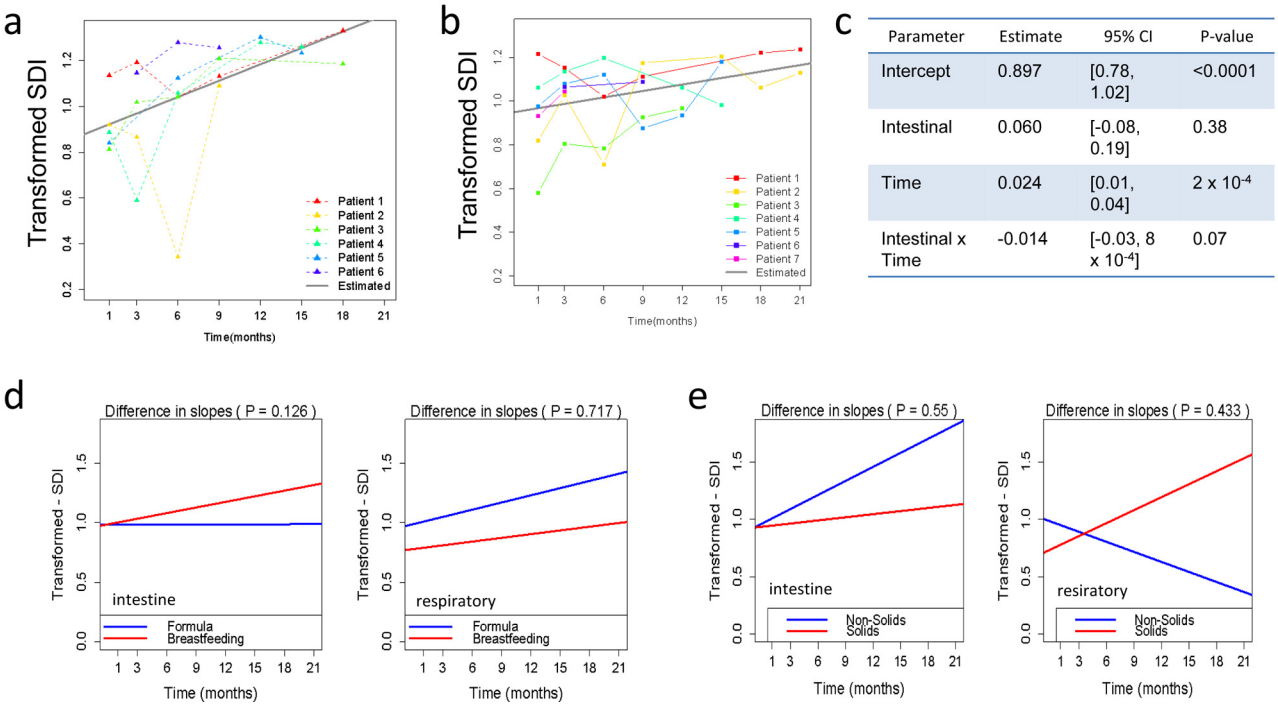


FIG 3 Changes in bacterial relative normalized abundance over time in the CF infant respiratory tract. (a) *y* axis is the coefficient (Coef) estimate representing the change in normalized $\log_{10}$ abundance over time. Values below the line represent a decrease in normalized abundance of the indicated bacterial genera over time (highlighted with red font if a statistically significant change with $P < 0.05$); points above the line represent bacteria that increase in normalized abundance over time. Blue dots highlight the bacteria at the genus level that are also present in the intestinal tract ($P < 0.05$). (b) *y* axis is normalized $\log_{10}$ abundance plotted versus time in months (*x* axis) for the 7 bacterial genera indicated. The intestinal relative normalized abundance is shown in blue, and the data for the respiratory samples are red.

breast milk compared with results with infant formula, as well as the introduction of solid foods (Fig. 4d and e). As discussed below, however, there are significant impacts on specific microbial populations in response to these nutritional parameters.

The recursive partitioned mixed-models (RPMM) (22) method was implemented to cluster samples on the basis of their microbiome profile and to examine the association between the resultant clusters and nutritional parameters. Briefly, RPMM is a hierarchical, model-based methodology for clustering samples on the basis of their observed normalized microbial abundances across a collection of genera. RPMM is similar to the well-known hierarchical clustering; however, given its model-based framework, this analysis requires the specification of a model for the underlying distribution of normalized microbial abundance. In our application of RPMM, a Gaussian (i.e., normal) distribution was assumed and fit to the $\log_{10}$-transformed, normalized microbial abundance data. The resulting RPMM solution provides an estimate of the number of underlying clusters, as well as the class membership probabilities for each of the samples within each of the estimated clusters.



Parameter	Estimate	95% CI	P-value
Intercept	0.897	[0.78, 1.02]	<0.0001
Intestinal	0.060	[-0.08, 0.19]	0.38
Time	0.024	[0.01, 0.04]	2 x 10^{-4}
Intestinal x Time	-0.014	[-0.03, 8 x 10^{-4}]	0.07

FIG 4 Trajectories of microbial diversity for the intestinal and respiratory microbiomes. SDIs were arc sine square root transformed (55) to satisfy normality conditions. (a) Subject-specific and estimated trajectories of the microbial diversity of the respiratory tract as a function of time. (b) Subject-specific and estimated trajectories of the microbial diversity of the intestinal tract as a function of time. (c) Summarized results of the linear mixed-effects model. (d, e) Trajectory of microbial diversity for breast milk versus formula (d) and for pre- and post-introduction of solid foods (e).

Four clusters were estimated from the RPMM solution applied to the intestinal microbiome samples only (Fig. 5). Subsequent statistical analysis revealed a significant association between the resultant clusters and the initiation of solid foods (permutation chi-square testing) (Fig. 5). We further examined this association using a permutation-based logistic regression model that was controlled for time and found that solid food remained statistically significant ($P = 0.01$).

RPMM was also applied to respiratory samples. Similar to the intestinal samples, four clusters were estimated from RPMM (see Fig. S3 in the supplemental material). Statistical analysis using a permutation chi-square test revealed a significant association between the resultant clusters and the start of solid foods ($P = 0.02$) and breast milk exposure ($P = 0.05$) (Table 2). A permutation-based logistic regression model controlled for time was fit, and only one variable remained marginally significant, breast-feeding ($P = 0.08$) (Table 2; see also Fig. S3).

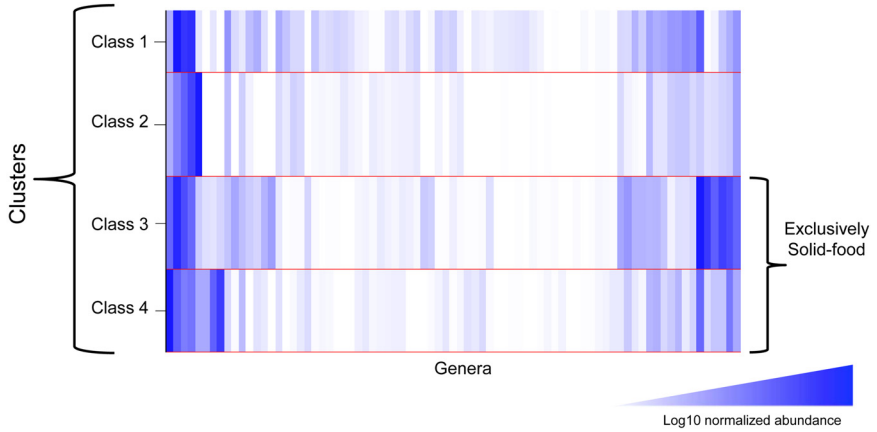


FIG 5 Heat map based on RPMM solution. The resulting clusters from model-based clustering of intestinal microbiome samples are shown. Rows represent samples, and columns represent microbes at the genus level, depicting the four clusters that were estimated from the model-based clustering solution. The actual values represent the mean (within cluster) $\log_{10}$-normalized abundances. Clusters, significantly associated with solid food ($P = 0.006$), adjusting for the time at which the sample was collected.

TABLE 2 Examining the association between RPMM clusters, derived from respiratory samples, and nutritional parameters

Variable	*P* value	
	Permutation chi square	Permutation logistic regression (controlled for time)
Histamine blockers	0.004	0.25
Solid food	0.021	0.89
CF exacerbation	0.83	0.52
Antibiotics	0.009	0.55
Breast-feeding	0.05	0.08
Pseudomonas colonization	0.05	0.87

DISCUSSION

Our in-depth study of a limited cohort of infants with CF, using massively parallel pyrosequencing technology, outlines the developing microbiome of the respiratory and intestinal tracts in infants with CF. Infants and young children with CF are highly susceptible to chronic colonization with pathogens and infection, and we were able to begin to outline their developing microbiome in detail, laying the groundwork for future studies to better characterize the microbiome in these patients to improve short- and long-term outcomes. The developing microbiota of the respiratory and intestinal tracts in infants with CF are distinct with an overlapping core and differentially increase in diversity over the first years of life. There is some concordance between the genera in both the gut and the respiratory tract and a subset of genera that are present in the gut prior to colonization of the respiratory tract.

We demonstrate that the majority of the bacteria present in the respiratory and intestinal tracts were represented by 8 distinct genera, with overlap between the two organ systems that was reflected in a core microbiota dominated by *Veillonella* and *Streptococcus*. Our group identified a separate cohort of adult subjects with CF showing a *Streptococcus* predominance in sputum samples, particularly during periods of clinical stability (23). Rogers et al., using non-culture-based studies, have highlighted two distinct phyla in the CF lung: the *Proteobacteria* (*Haemophilus influenzae*, *P. aeruginosa*, *Stenotrophomonas maltophilia*, and *Burkholderia cepacia* among others) and the *Firmicutes* (*S. aureus*), in addition to other members of these phyla, including *Veillonella*, and *Streptococcus* (24), findings which are mirrored in our study. It is interesting that our cohort of infants and young children with CF mirrors studies of adults with CF, where many of the bacteria in the CF respiratory tract are species that require anaerobic conditions for growth, such as *Veillonella* and *Prevotella* spp. (24–28). For example, Sibley et al. found that nearly half of the culture CF airway community in adult sputum was composed of obligate anaerobes (29).

Sample collection for respiratory microbiota evaluation involved oropharyngeal sampling which mirrors the clinical samples collected for culture at proscribed time points in the CF clinic. The oropharyngeal swab method has been reported to accurately reflect the microbial populations in the lung in CF (19), including the use to assess species in a culture-independent evaluation of respiratory microbes (20). In particular, this method is the preferred sampling method for infants, a population that does not produce sputum and for which bronchoalveolar lavage is not indicated clinically. Rosenfeld and colleagues analyzed data from three independent, prospective studies involving simultaneous collection of 286 oropharyngeal and BAL cultures from 141 children with CF under age 5 years (19). The concordance between the oropharyngeal cultures and simultaneous BAL cultures was high, with a 95% specificity, although the specificity of oropharyngeal cultures was higher for the presence of *P. aeruginosa* than for *S. aureus* (19). Although there are a limited number of studies that have investigated infant oral or salivary microbiota, one study of 5 edentulous infants' salivary bacterial microbiome identified that prior to tooth eruption, *Streptococcus*, *Veillonella*, and *Neisseria* were the predominant bacterial genera present (30). This is similar to the respiratory microbiota we identified with the three most common genera, including *Streptococcus*, *Veillonella*, and *Prevotella*. Of particular interest will be the comparison of clinical cultures with the paired deep-sequencing samples, which is the focus of the next analysis of this cohort.

We also discovered a surprising pattern of clusters of bacteria, including potential pathogens such as *Enterococcus*, that were present early in life in the gut and later in young life in the respiratory tract, highlighting the potential interrelatedness of these two organ systems and their microbiota. These findings suggest time-dependent development of the microflora in each organ system with an interaction that might prove to highlight clinically relevant information for intervention, for example with oral probiotic therapy. These findings are consistent with previous reports, with identification of bacteria in the respiratory tract in CF that are typically associated with the intestinal cavity (31) and are theorized to contribute to the continuum of interactions between the host and microbial community in CF that relates to both the lung and gut microbiota (24).

Our investigation of the diversity of bacteria over time revealed the unexpected finding that diversity in the respiratory tract increased more rapidly than that in the intestines in this population. It has been demonstrated that over time, the intestinal microbiome increases in diversity in humans, beginning at birth (32, 33), with transition to the height of diversity by the age of 1 to 3 years, mirroring an adult pattern (33, 34). There have been several authors who have identified that lack of microbial diversity, or an abrupt decrease in microbial diversity, is associated with infection or disease (3, 35, 36). Conversely, it has been postulated that the respiratory tract microbiota is not particularly diverse in its healthy state, and that perhaps increased diversity would be associated with infection and morbidity (24). Specific to adults with CF, investigations into the relationship between CF lung microbiota diversity and lung function have overall pointed to decreased lung diversity as being inversely proportional to lung function (3, 36, 37), thus overall supporting the idea that several organ systems have been identified as correlating increased diversity with health and decreased diversity with disease (3, 35, 36). One study of 8 to 9 years in duration in six adults subjects with CF identified that diversity decreased over time in parallel with progressive disease and remained stable in patients with milder lung disease; however, they identified antibiotic use rather than lung function as the most significant driver of decreased microbial diversity in sputum samples (38). Additionally, based on sputum samples, Stressmann et al. corroborated similar findings for 14 patients with CF that antibiotic use was most associated with decreased microbial diversity and that overall decreased diversity was correlated to more-severe lung disease, as well as abundance of *Pseudomonas aeruginosa* (28). Further evaluation of predominant pathogens, microbial di-

versity patterns, and clinical outcomes in infants and children with CF is warranted, since Huang et al. have postulated that the airway microbiome composition in CF is related to the severity of disease and that microbial drivers of disease exacerbation are likely to be identified with new culture-independent techniques (39).

We characterized the developing microbiome in infants and children with CF in relation to the common dietary exposures experienced by this patient population. Interestingly, the effect of exposures on diversity trajectories over time revealed more significant effects of breast milk exposure on respiratory tract diversity than on the gut. The introduction of solid food, as might be expected, preferentially impacted the intestinal microbiota. These data highlight the potential mechanisms for understanding pilot studies of probiotic administration in subjects with CF, linking alteration of the gut microflora with decreased pulmonary exacerbations (40). Breast milk is the gold standard nutrition in the newborn and infant period, providing significant amounts of *Bifidobacterium* and *Lactobacillus*, with significant health-promoting associations (41–43). Understanding the potential interaction between the developing intestinal microbiota, microbe-host interactions, and respiratory colonization highlights potential time points for intervention. However, it is important to note that at this stage and based on our study, no conclusions can be drawn regarding any benefits of breast-feeding in this population.

Limitations of the study include the small number of subjects involved and overall small number of sampling time points. Another potentially significant limitation, which is particularly relevant in infant populations, is the recently reported difficulty with culture-independent investigations in detecting bifidobacterial species as reported by Turroni et al. (44), which is being attributed to technical biases related to DNA extraction protocols or primers utilized.

The developing microbiota of the respiratory and intestinal tracts in infants with CF are distinct with an overlapping core and differentially increase in diversity over the first years of life. There are concordances between the genera in both the gut and the respiratory tract and a large number of genera that are present in the gut prior to colonization of the respiratory tract. Furthermore, changes in diet also result in altered respiratory microflora, suggesting a link between nutrition and development of microbial communities in the respiratory tract. Our findings suggest that nutritional factors and gut colonization patterns are determinants of microbial development in the respiratory tract in infancy and present opportunities for early intervention in CF.

MATERIALS AND METHODS

Human subjects, inclusion criteria, and clinical data. Institutional review board (IRB) approval was obtained in April 2010 (Center for the Protection of Human Subjects at Dartmouth, number 21761), with yearly renewal of approval in April 2011 and 2012, and parents of subjects provided written informed consent.

Eligible participants included neonates diagnosed prenatally or immediately postnatally with CF whose care for CF would take place at the Dartmouth Hitchcock Medical Center in Lebanon, NH, or Manchester, NH. Subjects were enrolled prior to 1 month of life, with follow-up prospective data and sample collection on an every 1- to 3-month basis until the age of 21 months. Neonates with clinical syndromes or other chromosomal abnormalities were not eligible for the study.

Clinical information collected prospectively included the following: demographics, prenatal/birth history, and interim medical history (specifically CF exacerbations using the Akron Children's Hospital Exacerbation scoring system, growth failure, and hospitalizations for CF-related disease), as well as height/weight. Clinical culture results were collected prospectively, along with medical interventions, including medications, hospitalizations, and detailed dietary and environmental histories. Maternal antibiotic exposure and complete neonatal and infant antibiotic exposure data were collected and analyzed relative to changes in the microbiome over time.

Sample collection and storage. Oropharyngeal and stool samples were collected at regularly scheduled time periods (every 3 months) beginning in the first month of life for up to 2 years during follow-up visits in the CF clinic. Oropharyngeal swabs were collected at the same time as the routine surveillance oropharyngeal cultures to minimize the burden for the participating children in CF clinic. Surveillance throat cultures in CF patients are screened for the presence of *S. aureus* and other common pathogens, such as *P. aeruginosa*, *B. cepacia*, and *H. influenzae*. Each research swab was placed in a 2-ml microcentrifuge tube, and the sample tubes were capped, stored at −20°C, and transported to the lab on dry ice. Sterile swab samples were washed in RNALater reagent (Qiagen) and vortexed, and then samples were divided into two 1-ml samples and frozen at −80. Stool samples were collected by parents at home using a sterile wooden spatula, placed in a sterile collection cup, and placed on ice or in the freezer until transport to the lab on ice. In the lab, stool samples were aliquoted into 250-mg samples in sterile tubes with 1 ml RNAlater and frozen at −80.

Sample processing. (i) Respiratory samples. Samples were thawed in batches of 10, and bacterial DNA was extracted immediately upon thawing. DNA extraction was performed according to published methods (45) using the Mobio PowerSoil DNA extraction kit (Mobio), utilizing a combination of mechanical and chemical lysis. Sample was loaded into the PowerBead tube, and a homogenization and lysis procedure was completed. The PowerBead tube contains a buffer that disperses the sample and protects nucleic acids from degradation. The mix is vortexed, which mixes the components of the PowerBead tube and disperses sample in the PowerBead solution.

(ii) Stool samples. DNA was extracted using the MoBio Powersoil bacterial DNA isolation kit. Two hundred fifty milligrams of stool sample was dissolved in RNALater reagent, and then DNA was extracted according to the manufacturer's instructions with details outlined above. The DNA was used to construct PCR amplicon libraries for sequencing.

Massively parallel sequencing. 454 pyrosequencing was performed at the Josephine Bay Paul Center, Marine Biological Laboratories. Pyrosequencing targeted the 16S rRNA gene using Titanium amplicon sequencing. A mix of five fusion primers at the 5′ end and four primers at the 3′ end of the V6 region were designed to capture the full diversity of rRNA sequences represented in molecular databases (46, 47). Similarly, fusion primers with multiple mixed sites were used to amplify the V4 through V6 regions (a single 3′ primer with degenerate sites replaced the pool of four primers used for V6). The V6 data sets are all GS20 and 101.7 nucleotides (nt), average; the rest are GSFLX titanium and 543.5 nt, average. Taxonomic identifiers were assigned to pyrotags by using the rRNA indexing algorithm Global Assignment of Sequence Taxonomy (GAST) (48), which compares pyrotags to known rRNA genes that have already been placed in a phylogenetic framework of more than 1,000,000 nearly full-length rRNA reference sequences (RefSSU) on the SILVA database (49). The tag mapping methodology GAST for taxonomic assignments of environmental V6 pyrotags and the V4V6 and V6 reference databases are freely available through the VAMPS (Visualization and Analysis of Microbial Population Structure) website (http://vamps.mbl.edu/resources/faq.php#gasting). There were no low-biomass concerns from the samples with sequencing. 16S rRNA gene pyrosequencing of the 59 study samples was conducted using first the V6 region and ultimately using the V4V6 hypervariable region in the samples. These two analyses were assessed with regard to their similarities and were not statistically significantly

different (see Table S2 in the supplemental material), and therefore the data were pooled for the analyses presented here. A small proportion of sequences were assigned a terminal rank at the species level (only 41% of OTUs), and therefore our analysis was focused toward comparisons at the genus level, for which there was considerably better sequence specificity (77% of OTUs were classified at the genus level).

Data analysis. Statistical analysis was aimed at examining the richness and diversity of the intestinal and respiratory microbiome among the seven neonates in the study from the 59 total samples analyzed. Given the longitudinal nature of these data, we sought to investigate both how the richness and diversity of the intestinal and respiratory microbiome change over time in developing infants with CF and if these changes associate with dietary and nutritional factors. Richness was defined as the total number of taxonomic units (abundance, expressed as the number of reads) within a sample, and diversity was evaluated using the SDI (50). For the *i*th sample, the SDI was recorded as $1 - D_i$, where

$$Di = \sum_{j-1}^{N_i} \frac{n_{ij}(n_{ij} - 1)}{N_i(N_i - 1)}$$

and where n_{ij} represents the number of the *j*th species/organism for the *i*th sample and N_i represents the total number of species/organisms for the *i*th sample. Thus, the SDI is approximately continuously distributed between 0 and 1, with values approaching 1 signifying greater microbial diversity. OTUs defined at a 3% sequence difference cutoff were treated as the terminal taxonomical rank. Statistical analyses were applied comprehensively at all phylogenetic ranks (phylum, class, order, family, genus, species, and OTU).

Hierarchical clustering using the Euclidean distance metric and average linkage was used to cluster all samples based on the $\log_{10}$-normalized abundance for each bacterial taxon. In addition to hierarchical clustering, samples were also clustered using RPMM (22), a model-based hierarchical clustering methodology that has been demonstrated to perform effectively and efficiently for clustering high-dimensional array-based genomic data (51–53). The model-based formulation of RPMM provides a robust statistical framework for estimating the number of clusters *K*, while also providing the posterior probability of class membership for each sample across each of the *K* terminal clusters. RPMM was applied to all samples, to respiratory microbiome samples only, and to intestinal microbiome samples only. Specifically, a Gaussian-distributed RPMM was used to cluster samples based on the $\log_{10}$-normalized abundances for each bacterial taxon using the CRAN R software package RPMM. To avoid resulting in a sparse data-clustering solution (i.e., a small number of samples within the resulting clusters), the maximum number of clusters was restricted to four. The phenotypic relevance of the resulting clusters was then analyzed using permutation chi-square tests and a permutation-based logistic regression procedure (54) for robust small-sample inference. To ensure that the phenotypic relevance of the resulting clusters was not driven by time, the permutation-based logistic regression models were adjusted for the time at which the microbiome sample was collected. The permutation-logistic regression models were fit using the CRAN R software package logregperm.

Linear mixed-effects models were used to investigate changes in microbial diversity and normalized bacterial abundance over time and how these changes related to dietary and nutritional factors. The use of linear mixed-effects models facilitates the examination of the linear relationship between two variables when one or more of the subjects being examined contain repeated measurements. For examining changes in microbial diversity over time, the SDI was arc sine square root (55) transformed to satisfy the assumption of normality.

Transformed SDI was modeled as the response, with fixed-effects terms including main effects for time (time at which the sample was collected) and sample type (respiratory or intestinal) and the interaction term and a random effect for subject. A similar strategy was employed for investigating the relationship between dietary and nutritional factors in relation to changes in respiratory and intestinal microbial diversity over time. For examining changes in normalized bacterial abundance over time, we modeled $\log_{10}$ normalized abundance as the response, with a fixed term for time and a random effect for patient. All analyses were implemented in the R software environment (http://cran.r-project.org/), version 2.13.

Nucleotide sequence accession number. All data generated from sequencing have been deposited in the NCBI (National Center for Biotechnology Information of the National Institutes of Health and the U.S. National Library of Medicine) GenBank (SRA) with accession number SRP014429.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00251-12/-/DCSupplemental.

Figure S1, PDF file, 0.5 MB.
Figure S2, PDF file, 0.8 MB.
Figure S3, PDF file, 0.7 MB.
Table S1, PDF file, 0.2 MB.
Table S2, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We are deeply grateful to the families and patients involved in the study. We also acknowledge the terrific efforts of the nurses and staff in the Dartmouth Hitchcock pulmonology clinics, as well as the pulmonology research nurses, including Lynn Feenan and Dana B. Dorman. We acknowledge Tom Caldwell for database creation and management.

This work was supported by the Hearst Foundation, the Synergy Grant (Dartmouth), the Joshua Burnett Career Development Award through the Hitchcock Foundation (Dartmouth), and the Department of Pediatrics, Dartmouth, to J.C.M., with additional support from NIH/NIEHS 1P20ES018175-02 and EPA RD-83459901-0. This work was also supported by a pilot grant to G.A.O. from the Cystic Fibrosis Foundation Research Development Program (STANTO07R0). G.A.O. and J.H.M. received joint support from the Neukom Institute. J.H.M. is funded by NIH grant R01 AI59694. J.A.F. was funded in part by NIH grants P20RR16448 and P20RR016454 and by NSF DBI0939454. M.L.S. and H.G.M. are funded by NIH 4UH3DK083993-02. B.A.S. and T.H.H. were funded by the Cystic Fibrosis Foundation Research Development Program (STANTO07R0), NIEHS P42 ES7373, GMS/NCRR P20-RR01878, and NIH R01 HL074175. P.L.H. is funded by NIH grant 2K24AT003683. The translational research core (TRC) for collecting and processing clinical samples is supported by grants from the National Center for Research Resources (5P20RR018787) and the National Institute of General Medical Sciences (8 P20 GM103413) (NIH). M.R.K. is funded by P20 ES018175 from NIEHS and RD-83459901 from the EPA.

REFERENCES

1. **Emerson J, Rosenfeld M, McNamara S, Ramsey B, Gibson RL.** 2002. *Pseudomonas aeruginosa* and other predictors of mortality and morbidity in young children with cystic fibrosis. Pediatr. Pulmonol. **34**:91–100.
2. **Bronstein MN, et al.** 1992. Pancreatic insufficiency, growth, and nutrition in infants identified by newborn screening as having cystic fibrosis. J. Pediatr. **120**:533–540.
3. **Cox MJ, et al.** 2010. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. PLoS One **5**:e11044. http://dx.doi.org/10.1371/journal.pone.0011044.
4. **Rosenecker J.** 2000. Relations between the frequency of the deltaF508 mutation and the course of pulmonary disease in cystic fibrosis patients infected with *Pseudomonas aeruginosa*. Eur. J. Med. Res. **5**:356–359.
5. **McKone EF, Emerson SS, Edwards KL, Aitken ML.** 2003. Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study. Lancet **361**:1671–1676.
6. **Polizzi A, et al.** 2005. Phenotypic expression of genotype-phenotype correlation in cystic fibrosis patients carrying the 852del22 mutation. Am. J. Med. Genet. A 132:434–440.
7. **Rosenfeld M, et al.** 2010. Baseline characteristics and factors associated with nutritional and pulmonary status at enrollment in the cystic fibrosis EPIC observational cohort. Pediatr. Pulmonol. **45**:934–944.

8. **Schaedel C, et al.** 2002. Predictors of deterioration of lung function in cystic fibrosis. Pediatr. Pulmonol. **33**:483–491.
9. **Sagel SD, et al.** 2009. Impact of *Pseudomonas* and *Staphylococcus* infection on inflammation and clinical status in young children with cystic fibrosis. J. Pediatr. **154**:183–188.
10. **Green DM, et al.** 2010. Mutations that permit residual CFTR function delay acquisition of multiple respiratory pathogens in CF patients. Respir. Res. **11**:140.
11. **Collaco JM, Blackman SM, McGready J, Naughton KM, Cutting GR.** 2010. Quantification of the relative contribution of environmental and genetic factors to variation in cystic fibrosis lung function. J. Pediatr. **157**: 802–807.
12. **Hudson VL, Wielinski CL, Regelmann WE.** 1993. Prognostic implications of initial oropharyngeal bacterial flora in patients with cystic fibrosis diagnosed before the age of two years. J. Pediatr. **122**:854–860.
13. **Rosenfeld M, et al.** 2001. Early pulmonary infection, inflammation, and clinical outcomes in infants with cystic fibrosis. Pediatr. Pulmonol. **32**: 356–366.
14. **Stoltz DA, et al.** 2010. Cystic fibrosis pigs develop lung disease and exhibit defective bacterial eradication at birth. Sci. Transl. Med. **2**:29ra31.
15. **Sordé R, Pahissa A, Rello J.** 2011. Management of refractory *Pseudomonas aeruginosa* infection in cystic fibrosis. Infect. Drug Resist. **4**:31–41.
16. **Stutman HR, Lieberman JM, Nussbaum E, Marks MI.** 2002. Antibiotic prophylaxis in infants and young children with cystic fibrosis: a randomized controlled trial. J. Pediatr. **140**:299–305.
17. **Frederiksen B, Koch C, Høiby N.** 1997. Antibiotic treatment of initial colonization with *Pseudomonas aeruginosa* postpones chronic infection and prevents deterioration of pulmonary function in cystic fibrosis. Pediatr. Pulmonol. **23**:330–335.
18. **Norkina O, Burnett TG, De Lisle RC.** 2004. Bacterial overgrowth in the cystic fibrosis transmembrane conductance regulator null mouse small intestine. Infect. Immun. **72**:6040–6049.
19. **Rosenfeld M, et al.** 1999. Diagnostic accuracy of oropharyngeal cultures in infants and young children with cystic fibrosis. Pediatr. Pulmonol. **28**: 321–328.
20. **Klepac-Ceraj V, et al.** 2010. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. Environ. Microbiol. **12**:1293–1303.
21. **Simpson E.** 1949. Measurement of diversity. Nature **163**:688.
22. **Houseman EA, et al.** 2008. Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. BMC Bioinformatics **9**:365.
23. **Filkins LM, et al 29 June** 2012. The prevalence of streptococci and increased polymicrobial diversity associated with cystic fibrosis patient stability. J. Bacteriol. http://dx.doi.org/10.1128/JB.00566-12.
24. **Rogers GB, et al.** 2010. Comparing the microbiota of the cystic fibrosis lung and human gut. Gut Microbes **1**:85–93.
25. **Rogers GB, et al.** 2009. Studying bacteria in respiratory specimens by using conventional and molecular microbiological approaches. BMC Pulm. Med. **9**:14.
26. **Tunney MM, et al.** 2008. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am. J. Respir. Crit. Care Med. **177**:995–1001.
27. **Rogers GB, Skelton S, Serisier DJ, van der Gast CJ, Bruce KD.** 2010. Determining cystic fibrosis-affected lung microbiology: comparison of spontaneous and serially induced sputum samples by use of terminal restriction fragment length polymorphism profiling. J. Clin. Microbiol. **48**: 78–86.
28. **Stressmann FA, et al** Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. Thorax, in press.
29. **Sibley CD, et al.** 2011. Culture enriched molecular profiling of the cystic fibrosis airway microbiome. PLoS One **6**:e22702. http://dx.doi.org/10.1371/journal.pone.0022702.
30. **Cephas KD, et al.** 2011. Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. PLoS One **6**:e23503. http://dx.doi.org/10.1371/journal.pone.0023503.
31. **Rogers GB, et al.** 2003. Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. J. Clin. Microbiol. **41**:3548–3558.
32. **Favier CF, Vaughan EE, De Vos WM, Akkermans AD.** 2002. Molecular monitoring of succession of bacterial communities in human neonates. Appl. Environ. Microbiol. **68**:219–226.
33. **Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO.** 2007. Development of the human infant intestinal microbiota. PLoS Biol. **5**:e177. http://dx.doi.org/10.1371/journal.pbio.0050177.
34. **Yatsunenko T, et al.** 2012. Human gut microbiome viewed across age and geography. Nature **486**:222–227.
35. **Costello EK, et al.** 2009. Bacterial community variation in human body habitats across space and time. Science **326**:1694–1697.
36. **Zemanick ET, Sagel SD, Harris JK.** 2011. The airway microbiome in cystic fibrosis and implications for treatment. Curr. Opin. Pediatr. **23**: 319–324.
37. **Delhaes L, et al.** 2012. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community—implications for therapeutic management. PLoS One **7**:e36313. http://dx.doi.org/10.1371/journal.pone.0036313.
38. **Zhao J, et al.** 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc. Natl. Acad. Sci. U. S. A. **109**:5809–5814.
39. **Huang YJ, Lynch SV.** 2011. The emerging relationship between the airway microbiota and chronic respiratory disease: clinical implications. Expert. Rev. Respir. Med. **5**:809–821.
40. **Bruzzese E, et al.** 2007. Effect of *Lactobacillus* GG supplementation on pulmonary exacerbations in patients with cystic fibrosis: a pilot study. Clin. Nutr. **26**:322–328.
41. **Youn HN, et al.** 2012. Intranasal administration of live *Lactobacillus* species facilitates protection against influenza virus infection in mice. Antiviral Res. **93**:138–143.
42. **Tanaka A, et al.** 2011. *Lactobacillus pentosus* strain b240 suppresses pneumonia induced by *Streptococcus pneumoniae* in mice. Lett. Appl. Microbiol. **53**:35–43.
43. **Thomas DJ, et al.** 2011. *Lactobacillus rhamnosus* HN001 attenuates allergy development in a pig model. PLoS One **6**:e16577. http://dx.doi.org/10.1371/journal.pone.0016577.
44. **Turroni F, et al.** 2012. Diversity of bifidobacteria within the infant gut microbiota. PLoS One **7**:e36957. http://dx.doi.org/10.1371/journal.pone.0036957.
45. **Biesbroek G, et al.** 2012. Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. PLoS One **7**:e32942. http://dx.doi.org/10.1371/journal.pone.0032942.
46. **Sogin ML.** 2009. Characterizing microbial population structures through massively parallel sequencing, p 19–34. *In* Epstein SS (ed), Uncultivated Microorganisms. Springer, New York, NY.
47. **Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. **8**:R143. http://dx.doi.org/10.1186/gb-2007-8-7-r143.
48. **Huse SM, et al.** 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet. **4**:e1000255. http://dx.doi.org/10.1371/journal.pgen.1000255.
49. **Pruesse E, et al.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**:7188–7196.
50. **Hill M.** 1973. Diversity and evenness: a unifying notation and its consequences. Ecology **54**:427–432.
51. **Christensen BC, et al.** 2011. DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. J. Natl. Cancer Inst. **103**:143–153.
52. **Marsit CJ, et al.** 2011. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. J. Clin. Oncol. **29**:1133–1139.
53. **Christensen BC, et al.** 2009. Epigenetic profiles distinguish pleural mesothelioma from normal pleura and predict lung asbestos burden and clinical outcome. Cancer Res. **69**:227–234.
54. **Potter DM.** 2005. A permutation test for inference in logistic regression with small- and moderate-sized data sets. Stat. Med. **24**:693–708.
55. **Rocke D.** 1993. On the beta transformation family. Technometrics **35**: 72–81.