

Accepted Manuscript

Lower Perinatal Exposure to Proteobacteria is an Independent Predictor of Early Childhood Wheezing

Benjamin A. Turturice, BS, Diane R. Gold, MD MPH, Augusto A. Litonjua, MD, Emily Oken, MD MPH, Sheryl Rifas-Shiman, MPH, David L. Perkins, MD PhD, Patricia W. Finn, MD

PII: S0091-6749(18)31277-6

DOI: [10.1016/j.jaci.2018.06.051](https://doi.org/10.1016/j.jaci.2018.06.051)

Reference: YMAI 13604

To appear in: *Journal of Allergy and Clinical Immunology*

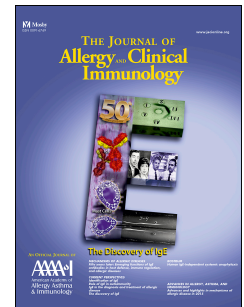
Received Date: 28 March 2018

Revised Date: 11 June 2018

Accepted Date: 25 June 2018

Please cite this article as: Turturice BA, Gold DR, Litonjua AA, Oken E, Rifas-Shiman S, Perkins DL, Finn PW, Lower Perinatal Exposure to Proteobacteria is an Independent Predictor of Early Childhood Wheezing, *Journal of Allergy and Clinical Immunology* (2018), doi: 10.1016/j.jaci.2018.06.051.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**Lower Perinatal Exposure to Proteobacteria is an Independent Predictor of Early
Childhood Wheezing**

Benjamin A. Turturice BS^{1, 2}, Diane R. Gold MD MPH^{3, 4, 5}, Augusto A. Litonjua MD⁶, Emily Oken
MD MPH⁷, Sheryl Rifas-Shiman MPH⁷, David L. Perkins MD PhD^{8, 9} §, Patricia W. Finn MD^{1, 2} §

1. Department of Microbiology and Immunology, University of Illinois at Chicago, Chicago, IL,
USA

2. Department of Medicine, Division of Pulmonary, Critical Care, Sleep, and Allergy, University
of Illinois at Chicago, Chicago, IL, USA

3. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's
Hospital, Harvard Medical School, Boston, MA, USA

4. Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and
Women's Hospital, Harvard Medical School, Boston, MA, USA

5. Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston,
MA, USA

6. Department of Pediatrics, University of Rochester, Rochester, NY, USA

7. Division of Chronic Disease Research Across the Lifecourse, Department of Population
Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

8. Department of Medicine, Division of Nephrology, University of Illinois at Chicago, Chicago, IL,
USA

9. Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, USA

§ These authors contributed equally to the work.

Corresponding Author

Patricia W. Finn, MD

840 S. Wood St.
CSB 1020N, M/C 787
Chicago, IL 60612 USA
Email: pwfinn@uic.edu
Phone: 312-996-7700
Fax: 312-413-0342

Word Count: 613

Disclosures: The authors report no conflicts of interest.

Acknowledgements: This work was funded by National Institutes of Health grants R01 AI 053878, F30 HL 136001, R01 HD 034568, and UG3 OD 023286.

Keywords

Asthma, Atopy, Gram Negative, Microbiota, Proteobacteria, Wheezing

43 Capsule Summary

44 Circulating free Proteobacterial DNA in umbilical cord serum is associated with features of
45 childhood asthma including maternal atopy, early childhood wheezing, and bronchodilator
46 response.

47

48

To the editor:

We postulated that perinatal risk factors for asthma would be associated with early exposure to bacteria or bacterial products. To evaluate this hypothesis, we assessed circulating-free bacterial DNA (cfbDNA) in umbilical cord blood from mother-child pairs (**Table S1**) from Project Viva, a longitudinal study from birth through childhood (1). To assess cfbDNA we utilized 16S rRNA sequencing of the V1-3 region from circulating DNA isolated from umbilical cord blood serum. Of the 27 maternal-child dyads in the cohort only four children reported a physician's diagnosis of asthma, therefore we chose to assess association the between cfbDNA composition, common risk factors for asthma, and asthma associated physiologies. For full technical and analytical methods see Online Repository at www.jacionline.org. We assessed associations between clinical features at birth and cfbDNA composition. Of the demographics tested, only maternal atopy was associated ($P = 0.070$, $R^2 = 0.17$, **Table S2**) with the composition of cfbDNA (**Fig. 1A**). The four top abundant phyla were compared between children of atopic and non-atopic mothers. Of these four phyla, only circulating Proteobacteria DNA was significantly different. Specifically, children of non-atopic mothers had increased Proteobacteria DNA compared to children of atopic mothers ($P = 0.031$) (**Fig. 1B**).

Two lines of evidence suggest that abundance of Proteobacterial cfbDNA is related to risk of asthma; lower abundance is associated with offspring who have increased likelihood of Th2 responses to common aeroallergens (2) and atopic mothers, shown in this study. Children who had at least one response to an annual questionnaire assessing wheezing were analyzed for Proteobacteria abundance. Proteobacteria was dichotomized by greater or less than the median and its association with wheezing within the first seven years after birth using Cox proportional hazards models (**Fig 2A, Table S3**). Lower exposure to Proteobacteria was associated with an increased risk of wheezing in a univariate model (HR (95%CI) = 3.60 (1.70-7.64), $P = 0.001$).

To assess if this association was independent of maternal atopic status, we adjusted for maternal atopy. Lower exposure to Proteobacteria was associated with an increased risk of wheezing in the multivariate model (aHR (95%CI) = 2.80 (1.46-5.36), $P = 0.002$). Children were evaluated by pulmonary function tests at mid-childhood follow up (median age 8.0 years, $n = 9$). Exposure to lower Proteobacteria trended towards an association with increased bronchodilator responses at mid-childhood ($P = 0.060$) (**Fig. 2B**). Only two children in the low Proteobacteria and no children in the high Proteobacteria group had clinically defined positive bronchodilator responses.

We previously reported that reduced Proteobacteria is associated with IL-13 responses to common allergens (2). In this study, we identified associations between Proteobacteria DNA in cord blood, maternal atopy (a strong risk factor for asthma), and common features of asthma (e.g., early-life wheezing and airflow reversibility). Proteobacteria are a diverse phyla of Gram negative bacteria and, as compared to Bacteroidetes (also Gram negative), lipopolysaccharide (LPS) derived from Proteobacteria may have additional immunomodulatory properties (3). LPS in mattress dust has previously been negatively correlated with peripheral blood mononuclear cell IL-13 responses to allergen challenge in childhood (4). In concordance, administration of an LPS mimetic in murine models of asthma skews immune response towards IFN- γ (Th1 cytokine) production and suppresses airway hyper-reactivity (5). Together, these data suggest that LPS exposure, both endogenous and exogenous, could suppress allergic responses implicated in the pathogenesis of childhood asthma. These findings support the hypothesis that early bacterial exposure patterns immune development which may drive later pulmonary disease.

In conclusion, we observed that lower circulating Proteobacteria DNA in umbilical cord blood serum is associated with maternal atopy, greater hazard of wheezing, and greater

bronchodilator response within the first seven years after birth. These findings are limited in their generalizability due to the small sample size, however, cfbDNA merits evaluation in a larger population as a novel biomarker for asthma risk independent of common risk factors.

Benjamin A. Turturice BS^{1, 2}

Diane R. Gold MD MPH^{3, 4, 5}

Augusto A. Litonjua MD⁶

Emily Oken MD MPH⁷

Sheryl Rifas-Shiman MPH⁷

David L. Perkins MD PhD^{8, 9§}

Patricia W. Finn MD^{1, 2 §}

1. Department of Microbiology and Immunology, University of Illinois at Chicago, Chicago, IL, USA

2. Department of Medicine, Division of Pulmonary, Critical Care, Sleep, and Allergy, University of Illinois at Chicago, Chicago, IL, USA

3. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

4. Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

5. Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

6. Department of Pediatrics, University of Rochester, Rochester, NY, USA

7. Division of Chronic Disease Research Across the Lifecourse, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

8. Department of Medicine, Division of Nephrology, University of Illinois at Chicago, Chicago, IL, USA

9. Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, USA

§ These authors contributed equally to the work.

References

1. Oken E, Baccarelli AA, Gold DR, Kleinman KP, Litonjua AA, De Meo D, et al. Cohort profile: project viva. *Int J Epidemiol*. 2015 February;44(1):37-48.
2. Benjamin A. Turturice, Ravi Ranjan, Brian Nguyen, Lauren M. Hughes, Kalista E. Andropolis, Diane R. Gold, et al. Perinatal Bacterial Exposure Contributes to IL-13 Aeroallergen Response. *Am J Res Cell Mol Bio*. 2017 Oct;57(4):419-427.
3. Vatanen T, Kostic A, d'Hennezel E, Siljander H, Franzosa E, Yassour M, et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell*. 2016 May 5;165(4):842-53.
4. Abraham JH, Finn PW, Milton DK, Ryan LM, Perkins DL, Gold DR. Infant home endotoxin is associated with reduced allergen-stimulated lymphocyte proliferation and IL-13 production in childhood. *Journal of Allergy and Clinical Immunology*. 2005 August;116(2):431-7.
5. Schaub B, Bellou A, Gibbons FK, Velasco G, Campo M, He H, et al. TLR2 and TLR4 stimulation differentially induce cytokine secretion in human neonatal, adult, and murine mononuclear cells. *J Interferon Cytokine Res*. 2004 September;24(9):543-52.

157

158

159

160

161

162

163

164

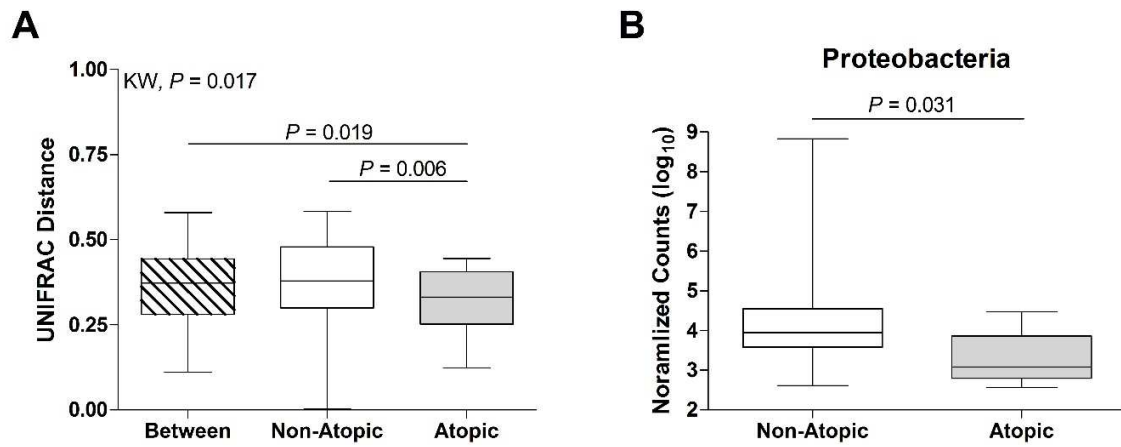
165

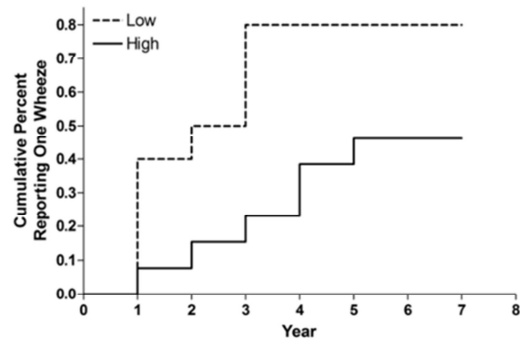
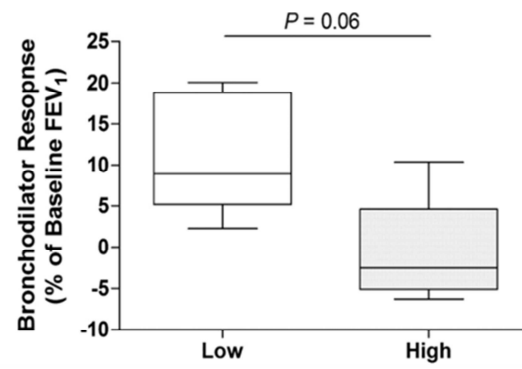
166

Figure Legends

Figure 1. *Maternal atopy is associated with composition of cfbDNA and lower Proteobacteria cfbDNA.* **A)** UNIFRAC pairwise distances are shown for distances between children with atopic mothers and those without (Between), and the with-in group distances for children of atopic mothers (Atopic, $n = 9$) and those of non-atopic mothers (Non-Atopic, $n = 17$). $P < 0.05$ was considered statistically significant, Kruskal-Wallis pre-hoc test or Dunn's post-hoc Test. **B)** Abundance of Proteobacteria displayed as normalized counts for samples from non-atopic and atopic mothers. $P < 0.05$ was considered statistically significant, Mann-Whitney U -test.

Figure 2. *Lower Proteobacteria exposure is associated with increased risk of wheezing.* **A)** Kaplan-Meier curves for median of imputed cumulative percent parental reported wheezing at each year post-birth for children with high (solid, $n = 13$) and low (dashed, $n = 10$) exposure to Proteobacteria cfbDNA. **B)** Children who had pulmonary function tests at mid-childhood follow up (median age 8.0 years) were dichotomized for abundance of Proteobacteria, high (grey, $n = 4$) and low (white, $n = 5$). The groups were assessed for associations with changes in pulmonary function tests at mid-childhood. Bronchodilator response displayed as box and whisker plots for children with high or low (greater or less than the median, respectively) Proteobacteria exposure. $P < 0.05$ was considered statistically significant, Mann-Whitney U -test.



A**B**

Online Data Supplement

All methods involving sample preparation and sequencing have previously been described (1) and are reiterated here, in addition to methods for subsequent reanalysis with outcome and demographic data.

Study population

The current study using was approved by the University of Illinois at Chicago IRB (#2016-0326). Volunteers were recruited from the eastern Massachusetts general population. Exclusion criteria were multiple gestation, inability to answer questions in English, gestational age ≥ 22 weeks at recruitment and plans to move away before delivery. The Project Viva cohort profile has been described in Oken et al. 2014 (2). Mother-child dyad demographic data from the 27 individuals assessed in this report is displayed in **Table S1**.

DNA isolation

DNA was extracted from 1 mL of cord blood serum using QIAmp DNA Blood Midi Kit (Qiagen). All extractions were carried out under sterile conditions. Each extraction was recorded for lot number on the kit and time of extraction and tested for any effect in our analysis. No effect was observed for lot number or date of extraction. To control for any possibility of bacterial DNA contamination introduced in our procedure, we used 1 mL of molecular grade water through our extraction protocol as a negative control. Our negative control showed both a minimum 5 cycle time difference assessing by qPCR from any sample and did not yield a 16S rRNA amplicon library. DNA concentration and base pair size was quantified from all samples using Qubit dsDNA High Sensitivity Dye (Invitrogen) and on 2100 Bioanalyzer using High Sensitivity DNA kit (Agilent).

16s rDNA amplicon sequencing

V1-V3 regions of the 16s rRNA gene were amplified using the NEXTflex™ 16s V1-V3 Amplicon-Seq Kit (BIOO Scientific) with an initial 10 cycle step of 16s rRNA V1-V3 amplification followed by a 20 cycle step for amplification and barcoding. Libraries were pooled at equal molar concentrations and size selected for amplicons between 550-700 base pairs (bp) using Pippin Prep™ 1.5% Agarose Dye-Free Pippin Gel Cassette (Sage Science) to remove any non-specific amplicons. The pooled libraries were further diluted to 15 pM and 5% PhiX was added to increase the diversity, and sequenced for a paired-end 301 read length. The DNA libraries were sequenced on Illumina's MiSeq platform using MiSeq Reagent v3-600 cycle kit, following manufacturer's protocol.

OTU Annotation

For operation taxonomic unit (OTU) annotation, we used the QIIME pipeline (3). Sequences were aligned to Greengenes Core reference alignment (4-5). Chimeric sequences were removed by UCHIME and OTUs were picked by USEARCH (7, 8). Taxonomy were assigned using the Greengenes database and a phylogenetic tree was constructed using FastTree 2.1.3 for the determination of UNIFRAC distances (9, 10).

Statistical Analysis

All statistical analysis was performed in R. Mann-Whitney U-test was used in all two group comparisons unless otherwise stated. For testing associations between demographic data and cfbDNA composition, samples were normalized using median sum-scaling (11), pair-wise sample distances were calculated using the weighted UNIFRAC method (9), ordinated using principle component analysis, and tested for associations (**Table S2**) using the *vegan envfit* function (12). Maternal atopy was assessed by interview response to "Has a health professional

ever told you that you have asthma, hay fever, or eczema?”. Phylum level abundances were calculated by combining normalized reads of each operational taxonomic unit (OTU) assigned to each phylum. Proteobacteria abundance was dichotomized in low and high based on normalized counts below or above the median, respectively. To determine associations between Proteobacteria abundance and early wheezing (response to annual questionnaire “Has your child had wheezing within the last 12 months?”) we assessed the time to first reported wheeze in univariate Cox-proportional hazards model and Cox-proportional hazards model adjusting for maternal atopy (**Table S3**). To account for missing questionnaire responses, missing data was imputed using the following method: subjects not reporting at any time point were not assessed, missing data for wheezing in annual questionnaires years one through seven was imputed 500 times with larger subset of subjects from Project Viva ($n = 293$) using *Amelia* (13), models were estimated for each of the imputed data sets, and multiple imputation model parameters were inferred using the method described by Rubin et al. 1987 (14). Missing data per each time point is reported in **Table S4**. There were no significant differences in missing data overall or at any time point. Pulmonary function test were compared at mid-childhood visit (median age 8.0 years) (15). There were no significant differences in FEV₁ or FVC between high and low Proteobacteria cfbDNA exposure groups (Data not shown). $P < 0.05$ was considered significant and $P < 0.10$ was considered a trend for all analyses.

Data Accessibility

Raw metagenomic sequence data for the samples have been uploaded onto NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>) and are available under the Accension:

PRJNA473573

References

1. Benjamin A. Turturice, Ravi Ranjan, Brian Nguyen, Lauren M. Hughes, Kalista E. Andropolis, Diane R. Gold, et al. Perinatal Bacterial Exposure Contributes to IL-13 Aeroallergen Response. 2017;57.
2. Oken E, Baccarelli AA, Gold DR, Kleinman KP, Litonjua AA, De Meo D, et al. Cohort profile: project viva. *Int J Epidemiol*. 2015 February;44(1):37-48.
3. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Meth*. 2010 May;7(5):335-6.
4. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol*. 2006 July 1;72(7):5069-72.
5. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 2012 March 1;6(3):610-8.
6. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*. 2010 January 15;26(2):266-7.
7. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010 October 1;26(19):2460-1.
8. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011 August 15;27(16):2194-200.
9. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl Environ Microbiol*. 2005 12/01;71(12):8228-35.
10. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5(3):e9490.

11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12).
12. Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin P, O'Hara RB, et al. *vegan: Community Ecology Package*. R package. 2016 "Feb ";2.3-3.
13. James Honaker, Gary King, Matthew Blackwell. *Amelia II: A Program for Missing Data*. *Journal of Statistical Software*. 2011 Dec 1;45(7).
14. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York [u. a.]: Wiley; 1987.
15. Oken E, Kleinman KP, Rich-Edwards J, Gillman MW. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC pediatrics*. 2003 Jul 8;3(1):6.

Table S1. Subject demographics

Description of variable	% (N)
Maternal Characteristics	
Atopy	
No	63.0 (17)
Yes	33.3 (9)
No Response	3.7 (1)
BMI, kg/m ² (mean (SD))	24.9 (5.8)
Race/ Ethnicity	
White	63.0 (17)
Black	7.4 (2)
Hispanic	7.4 (2)
Other	18.5 (5)
No Response	3.7 (1)
Smoking	
Never	85.2 (23)
Former	7.4 (2)
During Pregnancy	7.4 (2)
No Response	0.0 (0)
Antibiotics during Pregnancy	
No	63.0 (17)
Yes	33.3 (9)
No Response	3.7 (1)
C-Section	
No	74.1 (20)
Yes	25.9 (7)
No Response	0.0 (0)
Child's Characteristics	
Sex	
Male	55.6 (15)
Female	44.4 (12)
Race/ Ethnicity	
White	59.3 (16)
Black	7.4 (2)
Hispanic	7.4 (2)
Other	22.2 (6)
No Response	0.0 (0)
Gestational Age, weeks (mean (SD))	39.7 (1.2)
Birth Weight for gestational age z-score (mean (SD))	0.41 (0.87) [†]

[†] Sex specific birth weight for gestational age z-scores were estimated from Oken et al. 2003 (15).

Table S2. Demographic associations with cfbDNA composition

Description of variable	R^2	P
Maternal Characteristics		
Atopy	0.17	0.07
BMI, kg/m ²	0.01	0.95
Race/ Ethnicity	0.19	0.28
Smoking	0.05	0.70
Antibiotics during Pregnancy	0.00	0.90
C-Section	0.07	0.15
Child's Characteristics		
Sex	0.03	0.41
Race/ Ethnicity	0.13	0.54
Gestational Age, weeks	0.05	0.53
Birth Weight for gestational age z-score	0.15	0.15

R^2 and P are estimated by vector and factor fitting, envfit function vegan, and permutation testing ($B=999$) of environmental variables onto the first two axis of PCoA using weighted UNIFRAC distance. 47.4% and 15.3% of variation was attributed to principle component 1 and 2, respectively.

Table S3. Associated risk for low Proteobacteria cfbDNA exposure and time to first parental reported wheezing event

	β	SE	P	95%CI	$R^{\dagger}$
Imputed*					
Univariate	1.283	0.383	0.001	(0.533, 2.032)	0.433
Adjusted [‡]	1.030	0.331	0.002	(0.381, 1.679)	0.511
No Missing Data**					
Univariate	1.466	0.956	0.125	(-0.407, 3.339)	
Adjusted [‡]	1.382	1.058	0.191	(-0.692, 3.496)	
Imputed Wheeze[¶]					
Univariate	0.910	0.496	0.066	(-0.061, 1.881)	
Adjusted [‡]	0.793	0.523	0.129	(-0.230, 1.815)	
Imputed Non-Wheeze[#]					
Univariate	1.619	0.625	0.010	(0.395, 2.843)	
Adjusted [‡]	1.309	0.662	0.048	(0.011, 2.607)	

[†] R is estimated relative increase in variance due to nonresponse

[‡] Adjusted for maternal atopy

* Only subjects with complete questionnaire responses from year 1 to 7. $N = 6$ and $N = 5$ for Proteobacteria > and Proteobacteria < the median, respectively.

[¶] All missing data was reported as wheezing event

[#] All missing data reported as non-wheezing event

Table S4. Missing Data in Each Group per Time-Point

	% (N) < Median	% (N) > Median	P[#]
First Year	20.0 (2)	18.2 (2)	0.999
Second Year	10.0 (1)	18.2 (2)	0.999
Third Year	10.0 (1)	18.2 (2)	0.999
Fourth Year	10.0 (1)	0.0 (0)	0.476
Fifth Year	30.0 (3)	45.5 (5)	0.659
Sixth Year	30.0 (3)	36.4 (4)	0.999
Seventh Year	10.0 (1)	18.2 (2)	0.999
Total	17.1 (12)	22.1 (17)	0.536

Significance determined by Fisher's Exact Test