

# SCIENTIFIC REPORTS



OPEN

## Sex-specific associations of infants' gut microbiome with arsenic exposure in a US population

Anne G. Hoen<sup>1,2,3</sup>, Juliette C. Madan<sup>1,2,4</sup>, Zhigang Li<sup>3,8</sup>, Modupe Coker<sup>1</sup>, Sara N. Lundgren<sup>1</sup>, Hilary G. Morrison<sup>5</sup>, Thomas Palys<sup>2</sup>, Brian P. Jackson<sup>6</sup>, Mitchell L. Sogin<sup>5</sup>, Kathryn L. Cottingham<sup>2,7</sup> & Margaret R. Karagas<sup>1,2</sup>

Arsenic is a ubiquitous environmental toxicant with antimicrobial properties that can be found in food and drinking water. The influence of arsenic exposure on the composition of the human microbiome in US populations remains unknown, particularly during the vulnerable infant period. We investigated the relationship between arsenic exposure and gut microbiome composition in 204 infants prospectively followed as part of the New Hampshire Birth Cohort Study. Infant urine was analyzed for total arsenic concentration using inductively coupled plasma mass spectrometry. Stool microbiome composition was determined using sequencing of the bacterial 16S rRNA gene. Infant urinary arsenic related to gut microbiome composition at 6 weeks of life ( $p = 0.05$ , adjusted for infant feeding type and urine specific gravity). Eight genera, six within the phylum Firmicutes, were enriched with higher arsenic exposure. Fifteen genera were negatively associated with urinary arsenic concentration, including *Bacteroides* and *Bifidobacterium*. Upon stratification by both sex and feeding method, we found detectable associations among formula-fed males ( $p = 0.008$ ), but not other groups ( $p > 0.05$  for formula-fed females and for breastfed males and females). Our findings from a US population indicate that even moderate arsenic exposure may have meaningful, sex-specific effects on the gut microbiome during a critical window of infant development.

Arsenic is a ubiquitous environmental toxicant found in food and water worldwide<sup>1–3</sup>, and is present almost exclusively as inorganic arsenic in water supplies<sup>4</sup>, and as inorganic or organic species including metabolites of inorganic arsenic in dietary staples such as rice<sup>5</sup>. Environmental contaminants undergo biotransformation by gut bacteria. Biotransformation likely occurs with arsenic<sup>6</sup> and both absorption and toxicity of arsenic are highly dependent on its speciation<sup>4,7</sup>. A human *in vitro* gut microbiome model suggests that bacteria may methylate inorganic arsenic into highly toxic species, such as monomethylarsonous acid, which absorbs into the blood stream<sup>8</sup>. Thus, the relationship between arsenic and the human microbiome is bidirectional – direct by its antimicrobial effects on gut microbes, affecting gut permeability and immunity, and indirect by increasing the concentrations of microbial metabolites of arsenic capable of affecting human health<sup>1,4,6,8–10</sup>.

Fetuses, infants and young children are especially vulnerable to the effects of contaminants such as arsenic (reviewed in<sup>11</sup>). Early life also is a critical window for establishment of the gut microbiome, and, concomitantly, priming of the immune system<sup>12–21</sup>. While the infant gut microbiome may develop partially in fetal life, colonization chiefly occurs during birth, early feeding and beyond<sup>22–25</sup>. Immune maturation and programming involve interaction with gut microbes, and alteration in early life colonization patterns has been linked to lifelong disease risk, including allergy, asthma, and obesity<sup>24,26–33</sup>.

<sup>1</sup>Department of Epidemiology, The Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA.

<sup>2</sup>Children's Environmental Health & Disease Prevention Research Center at Dartmouth, Hanover, New Hampshire, USA. <sup>3</sup>Department of Biomedical Data Science, The Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA. <sup>4</sup>Division of Neonatology, Department of Pediatrics, Children's Hospital at Dartmouth, Lebanon, New Hampshire, USA. <sup>5</sup>Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, Massachusetts, USA.

<sup>6</sup>Department of Earth Sciences, Dartmouth College, Hanover, New Hampshire, USA. <sup>7</sup>Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, USA. <sup>8</sup>Present address: Department of Biostatistics, University of Florida, Gainesville, Florida, USA. Correspondence and requests for materials should be addressed to A.G.H. (email: [anne.g.hoen@dartmouth.edu](mailto:anne.g.hoen@dartmouth.edu))

Variable	All subjects (N = 204)	Males (N = 118)	Females (N = 81)
	Mean (Range) or N(%)	Mean (Range) or N(%)	Mean (Range) or N(%)
Gestational age, wk	39.3 (30.0–43.4)	39.3 (32.6–43.4)	39.6 (34.4–42.1)
Delivery mode			
Vaginal	140 (69)	82 (69)	58 (72)
Cesarean	55 (27)	34 (29)	17 (21)
Not reported	9 (4)	2 (2)	6 (7)
Infant birth weight, g**	3474 (1910–4689)	3496 (1910–4565)	3434 (2296–4689)
Feeding at 6 wk			
Exclusively breastfed	146 (72)	85 (72)	58 (72)
Combination fed	49 (24)	27 (23)	20 (25)
Exclusively formula-fed	9 (4)	6 (5)	3 (4)
Urinary arsenic concentration at 6 wk, $\mu\text{g/L}$	0.63 (<LOD – 4.8)	0.60 (<LOD – 3.9)	0.67 (<LOD – 4.8)
Home tap water arsenic concentration, $\mu\text{g/L}$	1.5 (<LOD – 57.0)	1.8 (<LOD – 57.0)	1.3 (<LOD – 23.1)

**Table 1.** Selected characteristics of participants overall and by sex\*. LOD: limit of detection (0.05  $\mu\text{g/L}$ ); \*sex not reported for N = 5; \*\*birth weight missing for n = 7 subjects.

Human arsenic exposure is a consequence of consumption of water and food containing arsenic. Infant formula powder and drinking water have been identified as important sources of arsenic exposure in infants<sup>34,35</sup>. US drinking water has varying levels of arsenic contamination and concentrations exceeding Environmental Protection Agency limits are particularly problematic in rural communities where private, unregulated wells commonly provide drinking water<sup>13</sup>.

Arsenic exposure is associated with effects on human health mediated through altered immunity, including infection, cancer, diabetes, cardiovascular disease, and in young children, impaired neurodevelopment<sup>12,36–45</sup>. Accumulating evidence indicate these effects occur as a consequence of arsenic exposure levels lower than previously appreciated. Investigations by our group of levels of arsenic exposure common in the US, beginning in fetal life through well water and maternal diet, identified potential health impacts that included dose-related trends in risk of infection and wheezing, as well as fetal growth restriction<sup>12,13,42,46</sup>. Immune profiling from this cohort has identified T cell decreases in cord blood<sup>18</sup>, increased placental gene expression of proinflammatory marker IL1B and developmental gene Gli3<sup>48</sup>, and DNA methylation changes in cord blood and placenta<sup>20,48</sup>.

In our comprehensive, molecular epidemiology study of over 1,500 pregnant women and subsequently, their offspring, we are investigating the effects of arsenic on pregnant women, their fetuses and infants. We are conducting the study in two regions of New Hampshire, including an area we previously determined has arsenic concentrations in well water that exceed 10  $\mu\text{g/L}$ , the Environmental Protection Agency maximum contaminant level for drinking water. Here, we hypothesized that infant arsenic exposure is associated with perturbations to the infant gut microbiome at 6 weeks of age and that these differences would be sex-dependent and most pronounced in formula-fed infants. Clarifying and quantifying human exposure and ultimate health impacts, including those that may be mediated by shifts in the composition of the microbiome, can lead to policy changes for protection of human health, especially in our most vulnerable populations.

## Results

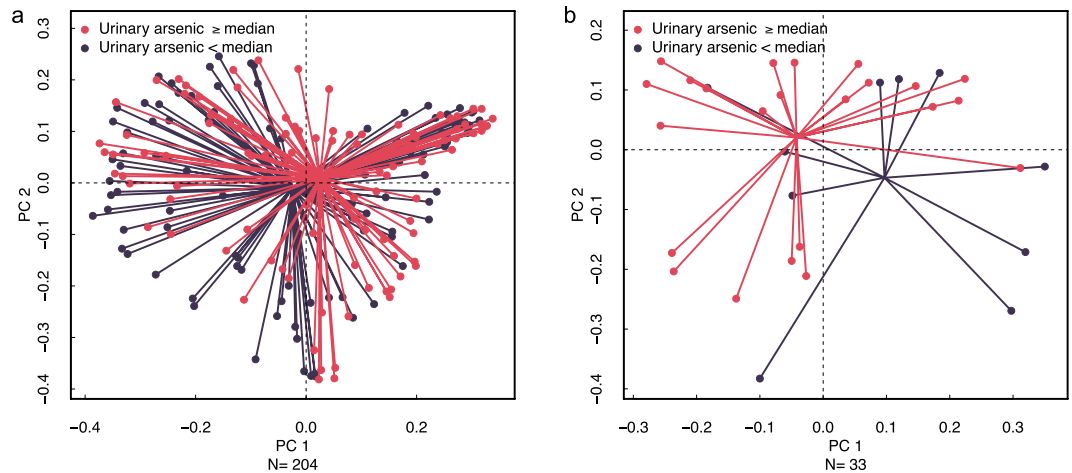
As of June 17, 2016, 1,572 pregnant women had been enrolled in the NHBCS, and 1,544 NHBCS infants had reached 6 weeks of age. Inclusion criteria for the current study required the availability of infant urine and stool samples collected at 6 weeks of age and information on infant diet during the first 6 weeks of life. We began collecting 6-week stool samples on August 15, 2012, and of the enrolled infants, 205 met inclusion criteria for the current study; of these, one sample was removed due to sequencing failure, resulting in a final sample size of 204.

Selected subject characteristics are shown in Table 1. The mean urinary arsenic concentration among 204 infants aged 6 weeks was 0.6  $\mu\text{g/L}$  with a median of 0.4  $\mu\text{g/L}$  and range of below the limit of detection (<0.05  $\mu\text{g/L}$ ) to 4.8  $\mu\text{g/L}$ . Three infants had urinary arsenic levels below the limit of detection. As we have demonstrated previously in this cohort, infant urinary arsenic levels were lowest among those who were exclusively breast fed<sup>34</sup>. Among the subjects included in the current study, infants who were exclusively breast fed had a mean urinary arsenic level of 0.5  $\mu\text{g/L}$  compared with a mean of 0.9  $\mu\text{g/L}$  for those fed both breast milk and formula and 1.2  $\mu\text{g/L}$  for those who were fed only formula.

Sequencing of the V4–V5 regions of the bacterial 16S rDNA present in DNA extracted from infant stool samples collected at the same time yielded a total of 19,099,004 (mean: 93,623, range: 10,818–630,397) bacterial DNA reads that passed quality filtering procedures (details in<sup>49</sup>).

### Relationship between urinary arsenic concentration and infant stool microbiome composition at 6 weeks of age.

Among the 204 infants in our study, natural log-transformed urinary arsenic concentration was associated with infant stool microbiome adjusted for infant diet and urine specific gravity: pseudo *F* statistic: 1.85, *p* = 0.05; adjusted for urine specific gravity only: pseudo *F* statistic: 1.88, *p* = 0.05; Fig. 1a). We identified several associations between natural log-transformed infant urinary arsenic concentration and the relative abundance of individual bacterial taxa (Supplementary Table S1 and Fig. 2). Adjusting models for feeding status and urine specific gravity, we identified 8 OTUs with positive associations with natural log-transformed



**Figure 1.** Principal coordinate plot of generalized UniFrac distance matrix comparing microbiome community profiles according to urinary arsenic concentration. (a) All  $N = 204$  infants in the study; (b)  $N = 33$  formula-fed males. Corresponding plots for other groups are presented in the supplementary information (Fig. S1). Statistical testing was performed using a continuous measure of arsenic exposure, the natural log of urinary arsenic concentration; however, for visualization purposes, here subjects were divided at the median urinary arsenic concentration of  $0.36 \mu\text{g/L}$ .

infant urinary arsenic concentrations at six weeks of age, including a strong positive association between natural log-transformed infant urinary arsenic and the relative abundance of bacteria in the genus *Ruminococcus*. We also identified 14 OTUs that were negatively associated with natural log-transformed infant urinary arsenic concentration, among them, OTUs assigned to the family Clostridiaceae and to the genera *Bacteroides* and *Bifidobacterium*. Overall, positively associated OTUs were largely from the phylum Firmicutes, while the majority of negatively associated OTUs were from the phyla Firmicutes or Bacteroidetes. Different OTUs representing bacteria of the genera *Lactobacillus* and *Dorea* were found to be both positively and negatively associated with natural log-transformed urinary arsenic concentration.

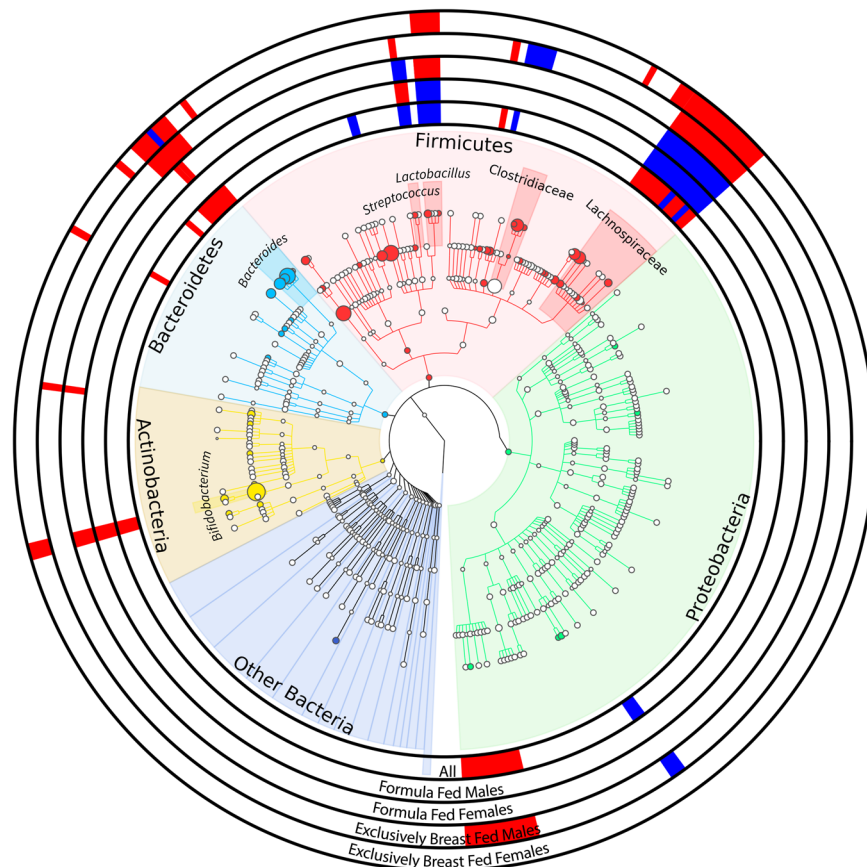
### Sex-specific relationships between urinary arsenic concentration and infant stool microbiome composition at 6 weeks of age in formula-fed infants.

We stratified study subjects by sex to identify sex-specific effects given that sex-specific perturbations to gut microbiome community structure by arsenic exposure through drinking water have been observed in mice<sup>50</sup>. In addition, our previous research in the NHBCS indicated clear associations between the infant gut microbiome at 6 weeks of age and infant feeding and we also previously reported in this cohort that breast milk fed infants have lower urinary arsenic than formula-fed infants and is not a likely source of dietary arsenic<sup>15,34</sup>. Therefore, to address the potential for confounding or modification of the relationship between arsenic exposure and the infant microbiome by feeding method, we also stratified based on infant feeding, analyzing infants who did and did not receive formula separately. Thus, we analyzed four groups: formula-fed males, formula-fed females, exclusively breastfed males and exclusively breastfed females.

We found strong associations between natural log-transformed urinary arsenic concentration and infant stool microbiome composition among exclusively or partially formula-fed male infants ( $n = 33$ , adjusted for formula exclusivity vs breast milk with formula supplementation and urine specific gravity: pseudo  $F$  statistic: 2.30,  $p = 0.008$ ; adjusted for urine specific gravity only: pseudo  $F$  statistic: 1.87,  $p = 0.03$ ; Fig. 1b). No such association was observed in formula-fed females ( $n = 23$ , adjusted for formula exclusivity vs breastfed with formula supplementation and urine specific gravity: pseudo  $F$  statistic: 0.54,  $p = 0.93$ ; adjusted for urine specific gravity only: pseudo  $F$  statistic: 0.56,  $p = 0.92$ ; Fig. S1A); exclusively breastfed males ( $n = 85$ , adjusted for urine specific gravity: pseudo  $F$  statistic: 0.58,  $p = 0.83$ ; Fig. S1B); or exclusively breastfed females ( $n = 58$ , adjusted for urine specific gravity: pseudo  $F$  statistic: 1.14,  $p = 0.39$ ; Fig. S1C).

To further investigate sex-specific differences among formula-fed infants, we compared mean UniFrac distances between pairs of samples in extreme quartiles of urinary arsenic among all formula-fed babies and, for male and female formula-fed babies only. In this analysis, we observed that males in extreme quartiles of urinary arsenic concentration had microbiome profiles that were more dissimilar than their female counterparts (Fig. 3).

Among males who were exclusively or partially formula-fed, natural log-transformed urinary arsenic concentration related to the relative abundance of 9 individual OTUs after adjusting for any breastfeeding and urine specific gravity—2 OTUs positively and 7 negatively (Fig. 2 and Supplementary Table S2). Microbes negatively associated with higher urinary arsenic in this subset included *Bifidobacterium*, *Bacteroides*, and *Lactobacillus*, whereas an OTU of the family Lachnospiraceae and another of the genus *Streptococcus* were positively associated. Associations between natural log-transformed urinary arsenic concentration and the relative abundance of individual OTUs for other stratification groups (formula-fed females, exclusively breastfed males, and exclusively breastfed females) are provided in the Supplementary Information (Supplementary Tables S3–S5).



**Figure 2.** Cladogram showing bacterial OTUs with differential relative abundances based on arsenic levels. Color-shaded areas of the cladogram delineate the most abundant phyla in the infant gut microbiota, with the size of corresponding nodes proportional to the abundance level of the clades. Several OTUs were significantly associated with arsenic levels; bars in the outer rings identified individual OTUs that were positively (blue bars) and negatively (red bars) correlated with urine arsenic concentrations. Rings correspond to all infants in the study (N = 204), male formula-fed infants only (N = 33), female formula-fed infants only (N = 23), male exclusively breastfed infants only (N = 85) and female exclusively breastfed infants only (N = 58).

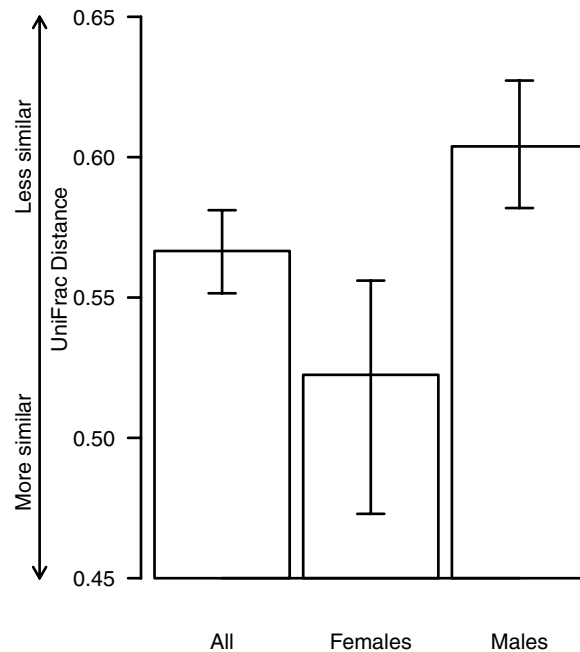
## Discussion

To our knowledge this is among the first studies of arsenic exposure and the human microbiome. In our analysis of 204 infants from our pregnancy cohort, we identified associations between urinary arsenic in 6-week-old infants and the early intestinal microbiome, both in terms of overall microbiome community composition and among bacterial taxa that are critical for immune training in infancy. We observed that male formula-fed infants may be more susceptible to the effects of arsenic on the microbiota than their female counterparts.

The range of arsenic exposure in well water in our US cohort study in New Hampshire is low in comparison to studies of arsenic exposure in other cohorts such as in Bangladesh<sup>38</sup>. Well water used by the mothers in this cohort had total arsenic concentrations ranging from below the limit of detection to 57.0  $\mu\text{g/L}$  (mean: 1.5  $\mu\text{g/L}$ ; median: 0.1  $\mu\text{g/L}$ ), with 2% above the EPA maximum contaminant level of 10  $\mu\text{g/L}$ . The infants in our study had urinary arsenic concentrations of ranging from below the limit of detection to 4.8  $\mu\text{g/L}$  (mean: 0.6  $\mu\text{g/L}$ ; median: 0.4  $\mu\text{g/L}$ ). At these exposure levels, we identified an association with the developing intestinal microbiome, alterations of which associate with major health outcomes<sup>12</sup>.

We found an overall marginal association between infant urinary arsenic and microbiome composition among all infants in the study after adjustment for feeding method and urine specific gravity. Breast milk is known to be low in arsenic in our population and others<sup>15,34,35,51–55</sup> and exclusively breastfed babies had lower urinary arsenic concentrations than formula or mixed fed infants in our cohort<sup>34</sup>. Previous research by our group showed that this is to the contamination of both formula powder and drinking water used to mix formula<sup>34</sup>, with popular brands of formula containing from 2.2–12.6 ng/g of arsenic, most of it inorganic<sup>56</sup>.

In our study, we identified associations with overall microbiome composition among formula-fed males but not among breastfed infants or among formula-fed females. Chi and colleagues<sup>50</sup> reported recently on sex-specific effects of arsenic on the microbiome in mice. In contrast to our finding in humans, female mice exhibited greater susceptibility to microbiome perturbation in association with exposure to arsenic-contaminated drinking water than males<sup>50</sup>. The bacterial communities that colonize mice and humans are distinct, a factor which may explain the different patterns of sex-specific associations between arsenic exposure and microbiome



**Figure 3.** Mean UniFrac distances between pairs of samples that are in extreme quartiles of urinary arsenic for all babies who were formula-fed, for formula-fed males only and for formula-fed females only. Bars represent bootstrapped 95% confidence intervals around the means.

composition; however, taken together, this experimental work in mice and our observational human study suggests that arsenic is one of many factors shaping the gut microbiome of exposed infants and that its effect may be different in males and females. In arsenic-exposed populations in Bangladesh, Mexico, and central Europe, sex has been implicated as a factor influencing the efficiency of inorganic arsenic metabolism<sup>57–59</sup>. In experimental gut simulation models, gut microbes are capable of metabolizing, thiolating and methylating arsenic into more or less toxic forms prior to absorption into the blood stream<sup>4,6,10</sup>. The underlying dynamics of any interactions between arsenic exposure, the gut microbiome and sex are therefore likely complex, as the composition of the gut microbiome may be both a driver and consequence of arsenic exposure, with sex hormones potentially influencing both arsenic metabolism and gut microbiome composition<sup>1,4,6,8–10,50,60</sup>. Metabolomic profiling may offer mechanistic clues.

Arsenic exposure is a risk factor for a range of diseases that are also associated with gut microbial community composition. Numerous epidemiological studies support associations between arsenic exposure and diabetes risk (reviewed in<sup>61</sup>). We found that genera from the phylum Firmicutes and the family Lachnospiraceae were enriched with higher arsenic exposure. In obese mice, intestinal colonization with Lachnospiraceae has been implicated in the development of diabetes<sup>62</sup>. Arsenic has also been shown to have determinantal effects on immune function, even in low doses<sup>63–66</sup>. In our study, Clostridiaceae and the genera *Bacteroides* and *Bifidobacterium*, in particular, were inversely related to arsenic exposure. *Bacteroides* species play a critical role in the maturation of the immune system and help induce regulatory T-cells<sup>67,68</sup>. *Bifidobacterium* also is implicated in immune training in early infancy, and decreases are associated with occurrence of allergy and atopy<sup>69–71</sup>. Studies linking the neonatal gut microbiome composition to development of asthma, identified a depletion of *Bifidobacterium*, *Faecalibacterium*, and *Akkermansia* as being associated with greater risk of asthma at age four<sup>71,72</sup>. Among male infants, negative associations were observed between infant urinary arsenic and *Bifidobacterium*, *Bacteroides* and *Lactobacillus*. The presence of these particular microbes in early life has been implicated in association with complex immune training of the adaptive and innate immune system; absence is associated with decreased immune responsiveness<sup>73</sup>, early onset allergy<sup>27,74</sup> and asthma and atopy<sup>71</sup>.

The differential abundance of many bacterial taxa along an arsenic exposure gradient that we observed is consistent with the antimicrobial effects of arsenic. For instance, diamino dihydroxy arsenobenzol (arsphenamine) known as Salvarsan was used as an antibiotic to treat syphilis prior to the introduction of penicillin<sup>4,6,10</sup>. Future studies should evaluate a potential role for gut microbiome in a causal pathway linking arsenic with immune function and health outcomes such as respiratory and gastrointestinal disease.

Our study benefited from the careful characterization of arsenic exposure during infancy using advanced methods in arsenic detection and next generation sequencing of the gut microbiome. Nonetheless, findings are limited by the analysis of one time point in infancy at six weeks of life in a single US cohort in New Hampshire. Thus, our results may not be generalizable to other populations, groups with higher levels of arsenic exposure, or at later time points in development. In addition, our observational study design limits our ability to make causal inferences. However, our study draws on one of the largest birth cohorts in the US, with ongoing longitudinal sampling of infants and young children. With ongoing accrual and future collaborations efforts, we hope larger, multi-center studies and prospective analysis of additional time points in infancy can be achieved.



In conclusion, we have identified potential impacts of infant arsenic exposure and differences in overall intestinal microbiome composition and in key taxa that are critical to proper immune training in early life. Our data suggest sex specific differences among formula-fed, arsenic exposed young infants, with male infants experiencing differences in the microbiome with increasing arsenic exposure. Clarifying the impact of low level arsenic exposure on the developing infant microbiome and determining the impact on immune function and health outcomes affords opportunity to alter exposures and impact lifelong health.

## Methods

**The New Hampshire Birth Cohort Study.** The New Hampshire Birth Cohort Study (NHBCS) recruited pregnant women ages 18–45 from New Hampshire prenatal clinics beginning at approximately 24 to 28 weeks gestation as previously described<sup>75,76</sup>. Women were eligible if they used a private, unregulated well at the residence occupied since their last menstrual period and were not planning to move. Institutional review board approval was obtained from the Center for the Protection of Human Subjects at Dartmouth, and participants provided written informed consent. All methods were performed in accordance with the relevant guidelines and regulations.

**Infant urine and stool samples.** A subset of mothers enrolled in the NHBCS provided infant stool samples collected at the time of regularly scheduled maternal postnatal follow-up visits conducted at approximately 6 weeks post-partum. Infant urine samples were collected using provided diapers (Pampers) containing cotton pads (Shiseido) according to a protocol adapted from Fångström *et al.*<sup>35</sup>. Urine-saturated pads were placed in an acid washed specimen cup and sealed in a polyethylene bag. Stool was collected in provided diapers, sealed in a separate polyethylene bag and frozen in home freezer until transport. Urine and stool samples were transported in a cooler with ice packs and brought to the post-partum visit within 24 hours of collection. Urine samples were maintained at 4 °C and processed within 24 hours. During processing urine was removed from pads, aliquoted into acid washed tubes and frozen at –80 °C. Stool was maintained frozen and thawed at 4 °C prior to processing. Using sterile applicators, 0.5–1 g of stool was aliquoted into cryovial tubes containing 3 ml of RNAlater stabilizer, homogenized by vortexing to create a slurry, maintained at 4 °C for 24 hours and then frozen at –80 °C.

**Urine sample analysis for infant arsenic exposure.** Urine samples were analyzed for total arsenic concentration using inductively coupled plasma mass spectrometry at the Dartmouth Trace Metal Analysis Laboratory as described previously<sup>34</sup>. The limit of detection for total urinary arsenic was 0.05 µg/L, and samples with measured arsenic levels below this limit were assigned values of half of the limit of detection, or 0.025 µg/L. Though total urinary arsenic includes arsenobetaine, an organoarsenic compound considered relatively non-toxic, human exposure to arsenobetaine is primarily through the consumption of fish and seafood. In young infants who have not yet been introduced to complementary foods, total arsenic is therefore a reasonable surrogate for exposure to toxic arsenic species.

**Stool sample analysis for infant gut microbiome composition.** We performed microbiome characterizations of stool samples as described previously<sup>49</sup>. Briefly, DNA was extracted using the Zymo Research ZR Fecal DNA Mini-prep extraction kit with 0.5 mm bashing beads and sequenced with Illumina tag sequencing of the 16S rRNA gene V4-V5 hypervariable region at the Marine Biological Laboratories in Woods Hole, Massachusetts using established methods<sup>77,78</sup>. Details of sequencing methods, quality control and filtering are described elsewhere<sup>79</sup>.

**Infant feeding.** We have previously reported associations between infant feeding method (breast milk vs. formula vs. a combination of breast milk and formula) and infant gut microbiome composition<sup>49</sup> in the NHBCS. In the present analysis, therefore, we took care to ensure that observed associations between infant arsenic exposure and microbiome composition were not confounded by infant feeding by both adjusting models for infant feeding and by stratification by feeding group. We ascertained infant diet from birth until the time of stool collection by telephone questionnaires that included questions regarding the duration of breastfeeding and the timing of formula introduction, if any. Infants who up until the time of urine and stool collection were fed breastmilk and who had never been given formula were given the dietary status of exclusive breastmilk feeding; infants who had not been breastfed and who had been fed formula only were assigned the status exclusively formula-fed; and infants who had received both breastmilk and formula were identified as having a diet of a combination of breastmilk and formula. Adjusted models included these three dietary categories. Stratified analyses evaluated formula-fed (i.e. exclusively formula-fed and combination fed) subjects as a group and exclusively breastfed subjects as a group. Sample size considerations precluded us from analyzing exclusively formula-fed subjects separately from combination fed subjects; however, our previous research on this cohort indicated that breastfed infants supplemented with formula have microbiome compositions more similar to those fed formula exclusively than to those fed breast milk exclusively<sup>49</sup>.

**Data analysis.** We implemented the UCLUST algorithm<sup>80</sup> in QIIME version 1.9.1<sup>81</sup> to group sequences of ≥97 percent similarity into open reference operational taxonomic units (OTUs). Bacterial taxonomy was assigned using PyNAST alignment<sup>82</sup> to the Greengenes reference database<sup>83–85</sup>.

We evaluated the associations between infant urinary arsenic concentration and stool microbiome community composition using generalized UniFrac analysis<sup>86</sup>. To avoid sequencing depth bias, the OTU table was rarefied to the minimum sequencing depth of 10,818 reads in the data set prior to analysis. The phylogenetic tree required for UniFrac analysis was computed using FastTree<sup>87</sup> and was midpoint rooted. Phylogenetic distances between the microbiome community composition of pairs of infant stool samples were characterized using generalized UniFrac analysis, and relationships between pairwise generalized UniFrac distance matrices and natural

log-transformed infant urinary arsenic concentration were evaluated with a permutational multivariate analysis of variance (PERMANOVA) performed with 10,000 permutations. All PERMANOVA models were adjusted for urine specific gravity to control for the effects of urine concentration. While statistical analyses were performed on continuous measures of urinary arsenic concentrations, generalized UniFrac distances were visualized in two-dimensional space using principal coordinates analysis with points colored according to group membership defined by separating subjects at the median urinary arsenic concentration for all subjects of 0.36  $\mu\text{g/L}$ . To further visualize relationships, samples were divided according to infant urinary arsenic concentration quartile, and mean pairwise generalized UniFrac distances were computed for pairs in extreme (first and fourth) quartiles. 95 percent confidence intervals around mean pairwise generalized UniFrac distances were computed using the bias correction with acceleration method.

Associations between infant urinary arsenic concentration and specific OTU relative abundance were modeled by a zero-inflated logistic normal (ZILN) model<sup>88</sup>. We regressed the log-ratio transformations of the OTUs relative abundance on the arsenic exposure adjusting for specific gravity and feeding method. Zero relative abundance values were handled by the zero part of ZILN distribution in the regression model. Estimating-equations approach was employed for the estimation procedure to address the complicated inter-taxa correlations induced by the hierarchical phylogenetic tree structure. The minimax concave penalty<sup>89</sup> was used to select OTUs and minimize bias in parameter estimation in the regression model. Beta estimates from this model were aggregated based on the lowest taxonomic rank, annotated into a phylogenetic tree and visualized using GraPhlAn v0.9.7 (<http://huttenhower.sph.harvard.edu/graphlan>)<sup>90</sup>.

**Data availability.** The datasets generated during and/or analyzed during the current study are available in GENBANK <http://www.ncbi.nlm.nih.gov/genbank/> under accession number PRJNA296814 and/or from the corresponding author on reasonable request.

## References

- Choiniere, J. & Wang, L. Exposure to inorganic arsenic can lead to gut microbe perturbations and hepatocellular carcinoma. *Acta Pharm Sin B* **6**, 426–429, <https://doi.org/10.1016/j.apsb.2016.07.011> (2016).
- Wilson, D. Arsenic Consumption in the United States. *J Environ Health* **78**, 8–14 (2015). quiz 44.
- Gundert-Remy, U. *et al.* High exposure to inorganic arsenic by food: the need for risk reduction. *Arch Toxicol* **89**, 2219–2227, <https://doi.org/10.1007/s00204-015-1627-1> (2015).
- Rubin, S. S. C. D. C., Alava, P., Zekker, I., Laing, G. D. & Van de Wiele, T. Arsenic thiolation and the role of sulfate-reducing bacteria from the human intestinal tract. *Environ Health Perspect* **122**, 817–822, <https://doi.org/10.1289/ehp.1307759> (2014).
- Schoof, R. A. *et al.* A market basket survey of inorganic arsenic in food. *Food Chem Toxicol* **37**, 839–846 (1999).
- Alava, P., Tack, F., Laing, G. D. & Van de Wiele, T. Arsenic undergoes significant speciation changes upon incubation of contaminated rice with human colon micro biota. *J Hazard Mater* **262**, 1237–1244, <https://doi.org/10.1016/j.jhazmat.2012.05.042> (2013).
- Sun, G. X., Van de Wiele, T., Alava, P., Tack, F. & Du Laing, G. Arsenic in cooked rice: effect of chemical, enzymatic and microbial processes on bioaccessibility and speciation in the human gastrointestinal tract. *Environ Pollut* **162**, 241–246, <https://doi.org/10.1016/j.envpol.2011.11.021> (2012).
- Alava, P., Du Laing, G., Tack, F., De Ryck, T. & Van De Wiele, T. Westernized diets lower arsenic gastrointestinal bioaccessibility but increase microbial arsenic speciation changes in the colon. *Chemosphere* **119**, 757–762, <https://doi.org/10.1016/j.chemosphere.2014.08.010> (2015).
- Trinder, M., Bisanz, J. E., Burton, J. P. & Reid, G. Probiotic lactobacilli: a potential prophylactic treatment for reducing pesticide absorption in humans and wildlife. *Benef Microbes* **6**, 841–847, <https://doi.org/10.3920/bm2015.0022> (2015).
- Laird, B. D. *et al.* An investigation of the effect of gastrointestinal microbial activity on oral arsenic bioavailability. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **48**, 612–619, <https://doi.org/10.1080/10934529.2013.731357> (2013).
- Vahter, M. Health effects of early life exposure to arsenic. *Basic Clin Pharmacol Toxicol* **102**, 204–211, <https://doi.org/10.1111/j.1742-7843.2007.00168.x> (2008).
- Farzan, S. F., Karagas, M. R. & Chen, Y. In utero and early life arsenic exposure in relation to long-term health and disease. *Toxicol Appl Pharmacol* **272**, 384–390, <https://doi.org/10.1016/j.taap.2013.06.030> (2013).
- Gilbert-Diamond, D., Emond, J. A., Baker, E. R., Korricks, S. A. & Karagas, M. R. Relation between in Utero Arsenic Exposure and Birth Outcomes in a Cohort of Mothers and Their Newborns from New Hampshire. *Environ Health Perspect* **124**, 1299–1307, <https://doi.org/10.1289/ehp.1510065> (2016).
- Ferrario, D., Gribaldo, L. & Hartung, T. Arsenic Exposure and Immunotoxicity: a Review Including the Possible Influence of Age and Sex. *Curr Environ Health Rep* **3**, 1–12, <https://doi.org/10.1007/s40572-016-0082-3> (2016).
- Carignan, C. C., Karagas, M. R., Punshon, T., Gilbert-Diamond, D. & Cottingham, K. L. Contribution of breast milk and formula to arsenic exposure during the first year of life in a US prospective cohort. *J Expo Sci Environ Epidemiol* **26**, 452–457, <https://doi.org/10.1038/jes.2015.69> (2016).
- Quansah, R. *et al.* Association of arsenic with adverse pregnancy outcomes/infant mortality: a systematic review and meta-analysis. *Environ Health Perspect* **123**, 412–421, <https://doi.org/10.1289/ehp.1307894> (2015).
- Garry, M. R., Santamaria, A. B., Williams, A. L. & DeSesso, J. M. In utero arsenic exposure in mice and early life susceptibility to cancer. *Regul Toxicol Pharmacol* **73**, 378–390, <https://doi.org/10.1016/j.yrtph.2015.07.023> (2015).
- Nadeau, K. C. *et al.* In utero arsenic exposure and fetal immune repertoire in a US pregnancy cohort. *Clin Immunol* **155**, 188–197, <https://doi.org/10.1016/j.jclim.2014.09.004> (2014).
- Bloom, M. S., Surdu, S., Neamtiu, I. A. & Gurzau, E. S. Maternal arsenic exposure and birth outcomes: a comprehensive review of the epidemiologic literature focused on drinking water. *Int J Hyg Environ Health* **217**, 709–719, <https://doi.org/10.1016/j.ijheh.2014.03.004> (2014).
- Fei, D. L. *et al.* Association between In Utero arsenic exposure, placental gene expression, and infant birth weight: a US birth cohort study. *Environ Health* **12**, 58, <https://doi.org/10.1186/1476-069x-12-58> (2013).
- Vahter, M. Effects of arsenic on maternal and fetal health. *Annu Rev Nutr* **29**, 381–399, <https://doi.org/10.1146/annurev-nutr-080508-141102> (2009).
- Meropol, S. B. & Edwards, A. Development of the infant intestinal microbiome: A bird's eye view of a complex process. *Birth Defects Res C Embryo Today* **105**, 228–239, <https://doi.org/10.1002/bdrc.21114> (2015).
- Lu, C. Y. & Ni, Y. H. Gut microbiota and the development of pediatric diseases. *J Gastroenterol* **50**, 720–726 (2015). 10.1007/s00535-015-1082-z.
- Kerr, C. A. *et al.* Early life events influence whole-of-life metabolic health via gut microflora and gut permeability. *Crit Rev Microbiol* **41**, 326–340, <https://doi.org/10.3109/1040841x.2013.837863> (2015).

25. Dogra, S. *et al.* Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *mBio* **6**, <https://doi.org/10.1128/mBio.02419-14> (2015).
26. Bjorksten, B., Sepp, E., Julge, K., Voor, T. & Mikelsaar, M. Allergy development and the intestinal microflora during the first year of life. *J Allergy Clin Immunol* **108**, 516–520, <https://doi.org/10.1067/mai.2001.118130> (2001).
27. Wang, M. *et al.* Reduced diversity in the early fecal microbiota of infants with atopic eczema. *J Allergy Clin Immunol* **121**, 129–134, <https://doi.org/10.1016/j.jaci.2007.09.011> (2008).
28. Adlerberth, I. & Wold, A. E. Establishment of the gut microbiota in Western infants. *Acta Paediatr* **98**, 229–238, <https://doi.org/10.1111/j.1651-2227.2008.01060.x> (2009).
29. Lee, Y. K. & Mazmanian, S. K. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* **330**, 1768–1773, <https://doi.org/10.1126/science.1195568> (2010).
30. Panzer, A. R. & Lynch, S. V. Influence and effect of the human microbiome in allergy and asthma. *Curr Opin Rheumatol* **27**, 373–380, <https://doi.org/10.1097/bor.0000000000000191> (2015).
31. Houghteling, P. D. & Walker, W. A. From Birth to “Immunohealth,” Allergies and Enterocolitis. *J Clin Gastroenterol* **49**(Suppl 1), S7–s12, <https://doi.org/10.1097/mcg.0000000000000355> (2015).
32. Johnson, C. C. & Ownby, D. R. Allergies and Asthma: Do Atopic Disorders Result from Inadequate Immune Homeostasis arising from Infant Gut Dysbiosis? *Expert Rev Clin Immunol*, 1–10, <https://doi.org/10.1586/1744666x.2016.1139452> (2016).
33. Lynch, S. V. & Boushey, H. A. The microbiome and development of allergic disease. *Curr Opin Allergy Clin Immunol* **16**, 165–171, <https://doi.org/10.1097/aci.0000000000000255> (2016).
34. Carignan, C. C. *et al.* Estimated exposure to arsenic in breastfed and formula-fed infants in a United States cohort. *Environ Health Perspect* **123**, 500–506, <https://doi.org/10.1289/ehp.1408789> (2015).
35. Fangstrom, B. *et al.* Breast-feeding protects against arsenic exposure in Bangladeshi infants. *Environ Health Perspect* **116**, 963–969, <https://doi.org/10.1289/ehp.11094> (2008).
36. Argos, M., Ahsan, H. & Graziano, J. H. Arsenic and human health: epidemiologic progress and public health implications. *Rev Environ Health* **27**, 191–195, <https://doi.org/10.1515/reveh-2012-0021> (2012).
37. Argos, M. *et al.* Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. *Lancet* **376**, 252–258, [https://doi.org/10.1016/s0140-6736\(10\)60481-3](https://doi.org/10.1016/s0140-6736(10)60481-3) (2010).
38. Yunus, F. M. *et al.* A Review of Groundwater Arsenic Contamination in Bangladesh: The Millennium Development Goal Era and Beyond. *Int J Environ Res Public Health* **13**, 215, <https://doi.org/10.3390/ijerph13020215> (2016).
39. Sanchez, T. R., Perzanowski, M. & Graziano, J. H. Inorganic arsenic and respiratory health, from early life exposure to sex-specific effects: A systematic review. *Environ Res* **147**, 537–555, <https://doi.org/10.1016/j.envres.2016.02.009> (2016).
40. Tsuji, J. S., Garry, M. R., Perez, V. & Chang, E. T. Low-level arsenic exposure and developmental neurotoxicity in children: A systematic review and risk assessment. *Toxicology* **337**, 91–107, <https://doi.org/10.1016/j.tox.2015.09.002> (2015).
41. Sung, T. C., Huang, J. W. & Guo, H. R. Association between Arsenic Exposure and Diabetes: A Meta-Analysis. *Biomed Res Int* **2015**, 368087, <https://doi.org/10.1155/2015/368087> (2015).
42. Farzan, S. F. *et al.* Gene-arsenic interaction in longitudinal changes of blood pressure: Findings from the Health Effects of Arsenic Longitudinal Study (HEALS) in Bangladesh. *Toxicol Appl Pharmacol* **288**, 95–105, <https://doi.org/10.1016/j.taap.2015.07.017> (2015).
43. Abdul, K. S., Jayasinghe, S. S., Chandana, E. P., Jayasumana, C. & De Silva, P. M. Arsenic and human health effects: A review. *Environ Toxicol Pharmacol* **40**, 828–846, <https://doi.org/10.1016/j.etap.2015.09.016> (2015).
44. Tolins, M., Ruchirawat, M. & Landrigan, P. The developmental neurotoxicity of arsenic: cognitive and behavioral consequences of early life exposure. *Ann Glob Health* **80**, 303–314, <https://doi.org/10.1016/j.aogh.2014.09.005> (2014).
45. Yunus, M., Soheli, N., Hore, S. K. & Rahman, M. Arsenic exposure and adverse health effects: a review of recent findings from arsenic and health studies in Matlab, Bangladesh. *Kaohsiung J Med Sci* **27**, 371–376, <https://doi.org/10.1016/j.kjms.2011.05.012> (2011).
46. Farzan, S. F. *et al.* Infant Infections and Respiratory Symptoms in Relation to in Utero Arsenic Exposure in a U.S. Cohort. *Environ Health Perspect* **124**, 840–847, <https://doi.org/10.1289/ehp.1409282> (2016).
47. Winterbottom, E. F. *et al.* GLI3 Links Environmental Arsenic Exposure and Human Fetal Growth. *EBioMedicine* **2**, 536–543, <https://doi.org/10.1016/j.ebiom.2015.04.019> (2015).
48. Green, B. B. *et al.* Epigenome-Wide Assessment of DNA Methylation in the Placenta and Arsenic Exposure in the New Hampshire Birth Cohort Study (USA). *Environ Health Perspect* **124**, 1253–1260, <https://doi.org/10.1289/ehp.1510437> (2016).
49. Madan, J. C. *et al.* Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants. *JAMA Pediatr* **170**, 212–219, <https://doi.org/10.1001/jamapediatrics.2015.3732> (2016).
50. Chi, L. *et al.* Sex-Specific Effects of Arsenic Exposure on the Trajectory and Function of the Gut Microbiome. *Chem Res Toxicol* **29**, 949–951, <https://doi.org/10.1021/acs.chemrestox.6b00066> (2016).
51. Bjorklund, K. L. *et al.* Metals and trace element concentrations in breast milk of first time healthy mothers: a biological monitoring study. *Environ Health* **11**, 92, <https://doi.org/10.1186/1476-069X-11-92> (2012).
52. Samanta, G. *et al.* Arsenic in the breast milk of lactating women in arsenic-affected areas of West Bengal, India and its effect on infants. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **42**, 1815–1825, <https://doi.org/10.1080/10934520701566785> (2007).
53. Concha, G., Vogler, G., Nermell, B. & Vahter, M. Low-level arsenic excretion in breast milk of native Andean women exposed to high levels of arsenic in the drinking water. *Int Arch Occup Environ Health* **71**, 42–46 (1998).
54. Sternowsky, H. J., Moser, B. & Szadkowsky, D. Arsenic in breast milk during the first 3 months of lactation. *Int J Hyg Environ Health* **205**, 405–409, <https://doi.org/10.1078/1438-4639-00161> (2002).
55. Islam, M. R. *et al.* Availability of arsenic in human milk in women and its correlation with arsenic in urine of breastfed children living in arsenic contaminated areas in Bangladesh. *Environ Health* **13**, 101, <https://doi.org/10.1186/1476-069X-13-101> (2014).
56. Jackson, B. P., Taylor, V. F., Punshon, T. & Cottingham, K. L. Arsenic concentration and speciation in infant formulas and first foods. *Pure Appl Chem* **84**, 215–223, <https://doi.org/10.1351/PAC-CON-11-09-17> (2012).
57. Lindberg, A. L. *et al.* Gender and age differences in the metabolism of inorganic arsenic in a highly exposed population in Bangladesh. *Environ Res* **106**, 110–120, <https://doi.org/10.1016/j.envres.2007.08.011> (2008).
58. Lindberg, A. L. *et al.* Metabolism of low-dose inorganic arsenic in a central European population: influence of sex and genetic polymorphisms. *Environ Health Perspect* **115**, 1081–1086, <https://doi.org/10.1289/ehp.10026> (2007).
59. Torres-Sanchez, L. *et al.* Sex differences in the reduction of arsenic methylation capacity as a function of urinary total and inorganic arsenic in Mexican children. *Environ Res* **151**, 38–43, <https://doi.org/10.1016/j.envres.2016.07.020> (2016).
60. Cong, X. *et al.* Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender. *PLoS One* **11**, e0152751, <https://doi.org/10.1371/journal.pone.0152751> (2016).
61. Maull, E. A. *et al.* Evaluation of the association between arsenic and diabetes: a National Toxicology Program workshop review. *Environ Health Perspect* **120**, 1658–1670, <https://doi.org/10.1289/ehp.1104579> (2012).
62. Kameyama, K. & Itoh, K. Intestinal colonization by a Lachnospiraceae bacterium contributes to the development of diabetes in obese mice. *Microbes Environ* **29**, 427–430, <https://doi.org/10.1264/jisme2.ME14054> (2014).
63. Dangleben, N. L., Skibola, C. F. & Smith, M. T. Arsenic immunotoxicity: a review. *Environ Health* **12**, 73, <https://doi.org/10.1186/1476-069X-12-73> (2013).
64. Kozul, C. D., Ely, K. H., Enelow, R. I. & Hamilton, J. W. Low-dose arsenic compromises the immune response to influenza A infection in vivo. *Environ Health Perspect* **117**, 1441–1447, <https://doi.org/10.1289/ehp.0900911> (2009).
65. Selgrade, M. K. Immunotoxicity: the risk is real. *Toxicol Sci* **100**, 328–332, <https://doi.org/10.1093/toxsci/kfm244> (2007).



66. Nayak, A. S., Lage, C. R. & Kim, C. H. Effects of low concentrations of arsenic on the innate immune system of the zebrafish (*Danio rerio*). *Toxicol Sci* **98**, 118–124, <https://doi.org/10.1093/toxsci/kfm072> (2007).
67. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118, <https://doi.org/10.1016/j.cell.2005.05.007> (2005).
68. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313–323, <https://doi.org/10.1038/nri2515> (2009).
69. Di Gioia, D., Aloisio, I., Mazzola, G. & Biavati, B. Bifidobacteria: their impact on gut microbiota composition and their applications as probiotics in infants. *Appl Microbiol Biotechnol* **98**, 563–577, <https://doi.org/10.1007/s00253-013-5405-9> (2014).
70. Turrone, F., Ribbera, A., Foroni, E., van Sinderen, D. & Ventura, M. Human gut microbiota and bifidobacteria: from composition to functionality. *Antonie Van Leeuwenhoek* **94**, 35–50, <https://doi.org/10.1007/s10482-008-9232-4> (2008).
71. Fujimura, K. E. & Lynch, S. V. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. *Cell Host Microbe* **17**, 592–602, <https://doi.org/10.1016/j.chom.2015.04.007> (2015).
72. Fujimura, K. E. *et al.* Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med* **22**, 1187–1191, <https://doi.org/10.1038/nm.4176> (2016).
73. Sjogren, Y. M. *et al.* Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses. *Clin Exp Allergy* **39**, 1842–1851, <https://doi.org/10.1111/j.1365-2222.2009.03326.x> (2009).
74. Bisgaard, H. *et al.* Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *J Allergy Clin Immunol* **128**(646–652) **645**, e641, <https://doi.org/10.1016/j.jaci.2011.04.060> (2011).
75. Farzan, S. F. *et al.* In utero arsenic exposure and infant infection in a United States cohort: a prospective study. *Environ Res* **126**, 24–30, <https://doi.org/10.1016/j.envres.2013.05.001> (2013).
76. Gilbert-Diamond, D. *et al.* Rice consumption contributes to arsenic exposure in US women. *Proc Natl Acad Sci USA* **108**, 20656–20660, <https://doi.org/10.1073/pnas.1109127108> (2011).
77. Degnan, P. H. & Ochman, H. Illumina-based analysis of microbial community diversity. *ISME J* **6**, 183–194, <https://doi.org/10.1038/ismej.2011.74> (2012).
78. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621–1624, <https://doi.org/10.1038/ismej.2012.8> (2012).
79. Newton, R. J. *et al.* Sewage reflects the microbiomes of human populations. *mBio* **6**, e02574, <https://doi.org/10.1128/mBio.02574-14> (2015).
80. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461, <https://doi.org/10.1093/bioinformatics/btq461> (2010).
81. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336, <https://doi.org/10.1038/nmeth.f.303> (2010).
82. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267, <https://doi.org/10.1093/bioinformatics/btp636> (2010).
83. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069–5072, <https://doi.org/10.1128/AEM.03006-05> (2006).
84. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**, 610–618, <https://doi.org/10.1038/ismej.2011.139> (2012).
85. Werner, J. J. *et al.* Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J* **6**, 94–103, <https://doi.org/10.1038/ismej.2011.82> (2012).
86. Chen, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**, 2106–2113, <https://doi.org/10.1093/bioinformatics/bts342> (2012).
87. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641–1650, <https://doi.org/10.1093/molbev/msp077> (2009).
88. Li, Z. *et al.* Conditional regression based on a multivariate zero-inflated logistic normal model for microbiome relative abundance data. *Statistics in Biosciences*. <https://doi.org/10.1007/s12561-018-9219-2> (2018).
89. Zhang, C. H. Nearly Unbiased Variable Selection under Minimax Concave Penalty. *Ann Stat* **38**, 894–942, <https://doi.org/10.1214/09-Aos729> (2010).
90. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029, <https://doi.org/10.7717/peerj.1029> (2015).

## Acknowledgements

The authors are grateful to the children and families that made this study possible and to the staff of the New Hampshire Birth Cohort Study. This work was supported in part by grants from the National Institutes of Health (NIEHS P01ES022832, NIEHS P20ES018175, NIGMS R01GM123014, NIGMS P20GM104416, NLM K01LM011985 and NLM R01LM012723) and the US Environmental Protection Agency (RD-83544201 and RD-83459901).

## Author Contributions

Design and conduct of research performed by J.C.M. and M.R.K.; collection, management, analysis, and interpretation of data were conducted by A.G.H., J.C.M., Z.L., M.C., S.N.L., H.G.M., T.P., B.J. and M.R.K.; preparation, review, or approval of the manuscript were performed by A.G.H., J.C.M., Z.L., M.C., S.N.L., H.G.M., T.P., B.J., M.L.S., K.L.C. and M.R.K.; all authors reviewed, revised, and participated in the decision to submit the manuscript for publication.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-30581-9>.

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018