EBioMedicine 3 (2016) 172–179

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Allergy associations with the adult fecal microbiota: Analysis of the American Gut Project

Xing Hua, James J. Goedert *, Angela Pu, Guoqin Yu, Jianxin Shi

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 13 October 2015
Received in revised form 18 November 2015
Accepted 23 November 2015
Available online 27 November 2015

Keywords:
Human microbiome
Feces
Adults
Hygiene hypothesis
Allergy

ABSTRACT

Background: Alteration of the gut microbial population (dysbiosis) may increase the risk for allergies and other conditions. This study sought to clarify the relationship of dysbiosis with allergies in adults.
Methods: Publicly available American Gut Project questionnaire and fecal 16S rRNA sequence data were analyzed. Fecal microbiota richness (number of observed species) and composition (UniFrac) were used to compare adults with versus without allergy to foods (peanuts, tree nuts, shellfish, other) and non-foods (drug, bee sting, dander, asthma, seasonal, eczema). Logistic and Poisson regression models adjusted for potential confounders. Odds ratios and 95% confidence intervals (CI) were calculated for lowest vs highest richness tertile. Taxonomy associations considered 122 non-redundant taxa (of 2379 total taxa) with ≥0.1% mean abundance.
Results: Self-reported allergy prevalence among the 1879 participants (mean age, 45.5 years; 46.9% male) was 81.5%, ranging from 2.5% for peanuts to 40.5% for seasonal. Fecal microbiota richness was markedly lower with total allergies ($P = 10^{-9}$) and five particular allergies ($P \leq 10^{-4}$). Richness odds ratios were 1.7 (CI 1.3–2.2) with seasonal, 1.8 (CI 1.3–2.5) with drug, and 7.8 (CI 2.3–26.5) with peanut allergy. These allergic participants also had markedly altered microbial community composition (unweighted UniFrac, $P = 10^{-4}$ to 10^{-7}). Total food and non-food allergies were significantly associated with 7 and 9 altered taxa, respectively. The dysbiosis was most marked with nut and seasonal allergies, driven by higher *Bacteroidales* and reduced *Clostridiales* taxa.
Interpretation: American adults with allergies, especially to nuts and seasonal pollen, have low diversity, reduced *Clostridiales*, and increased *Bacteroidales* in their gut microbiota. This dysbiosis might be targeted to improve treatment or prevention of allergy.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Allergies, specifically type I hypersensitivity disorders, are clinically important and increasingly prevalent. In the US population from 1988–1994 to 2005–2006, self-reported prevalence of physician-diagnosed seasonal pollen allergy (hay fever), for example, increased from 8.8% to 11.3% (Salo et al., 2011; Sheikh et al., 2003). Asthma prevalence in the US population in 2005–2006 was estimated to be 14.1% (Liu et al., 2010). Modern hygiene has been postulated to contribute to the increasing prevalence of allergies, based on both functional and observational studies. Children who have fewer early life exposures, such as in small families, are more likely to develop seasonal pollen allergy or eczema (Strachan, 2000). Children in households with at least 2 dogs or cats are 70% less likely to develop serologic or skin prick test reactivity to common respiratory antigens (Ownby et al., 2002). In Europe and other modern societies, risk for allergy and asthma are lower for children on farms (Ege et al., 2011). And asthma prevalence increases with migration from a less to a more highly industrialized country (Gibson et al., 2003; Tobias et al., 2001).

Functionally, allergy results from inappropriate T-helper type 2 (Th2) immune response to generally innocuous protein. As reviewed in Arrieta et al. (2014), and elaborated in murine models (Bowman and Holt, 2001; Hrncir et al., 2008; Olszak et al., 2012), maturation of the Th2 responses that predominate at birth to Th1 predominance in infancy and adulthood is conditional on the presence of commensal gut bacteria. More recently, Ohnmacht et al. demonstrated in mice that the microbial population of the gut (the microbiota) controls systemic Th2 responses by inducing enteric Th17 and regulatory T cells (Ohnmacht et al., 2015).

Fujimura and Lynch comprehensively reviewed the relationship between the microbiota and risk for allergy and asthma, particularly in infancy and in murine models (Fujimura and Lynch, 2015). In the nasopharynx, predominance by *Moraxella*, *Streptococcus*, and *Haemophilus* during the first few months of life predicted development of childhood

Abbreviations: AGP, American Gut Project; FDR, false discovery rate; MiRKAT, Microbiome Regression-based Kernel Association Test; NHANES, National Health And Examination Survey; PD, phylogenetic diversity; PCoA, principal coordinate analysis; QIIME, Quantitative Insights Into Microbial Ecology; RA, relative abundance; 16S rRNA, 16S ribosomal RNA.

* Corresponding author at: 9609 Medical Center Drive, Room 6E106 MSC 9767, Bethesda, MD 20892, USA.
E-mail address: goedertj@mail.nih.gov (J.J. Goedert).

http://dx.doi.org/10.1016/j.ebiom.2015.11.038
2352-3964/Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

asthma (Teo et al., 2015). Among adolescents in Finland, sensitization to respiratory allergens was associated with low diversity of Gammaproteobacteria on the skin (Hanski et al., 2012). Gut microbial differences may also contribute to allergy risk in humans (Penders et al., 2014). In two studies, infants who had a higher fecal abundance of *Clostridium difficile* and *Escherichia coli*, respectively, had an increased risk of developing an allergy in the future (Kalliomaki et al., 2001; Penders et al., 2006). In a very small Swedish study, low fecal microbial diversity at age 1 month predicted atopic eczema by age 2 years (Abrahamsson et al., 2012), as well as asthma, but not rhinoconjunctivitis, eczema, or atopy, by age 7 years (Abrahamsson et al., 2014). In Denmark, fewer fecal bacterial taxa by molecular fingerprinting predicted allergic rhinitis but not asthma or atopic dermatitis by age 6 (Bisgaard et al., 2011).Comprehensive analysis based on next-generation sequencing has not yet clarified whether alteration of the gut microbiota (dysbiosis) is associated with allergy in infants or adults. To address this, in adults, we analyzed publicly available data from the American Gut Project, similar to a previous analysis of the microbiota with history of cesarean birth and appendectomy (Goedert et al., 2014).

2. Methods

2.1. Microbiome and Phenotypic Data

The 16S rRNA V4 region was sequenced by the American Gut Project (AGP). The operational taxonomic unit (OTU) table rarefied to 10,000 sequence reads per sample, as well as metadata, was downloaded from the AGP website (https://github.com/biocore/AmericanGut/tree/master/data/AG). Samples with less than 10,000 sequence reads were excluded from analysis. A current summary is available at http://microbio.me/AmericanGut/static/img/mod1_main.pdf, and details of the OTU picking and taxonomy assignment are available at http://nbviewer.ipython.org/github/biocore/American-Gut/blob/master/ipynb/module2_v1.0.ipynb. Richness (number of observed species), alpha diversity metrics [Shannon index, Chao1, phylogenetic diversity (PD)_whole_tree], beta diversity metrics (weighted and unweighted UniFrac distance matrices), and relative abundance of each taxon were calculated in the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (Caporaso et al., 2010).

After exclusions [duplicates, diabetes, inflammatory bowel disease, age <4 years (after which the microbiota resembles that of adults (Yatsunenko et al., 2012)), missing race, specimen not feces, antibiotic used in the past month], data were analyzed for 1879 AGP participants. Each participant who provided a positive response on the AGP self-administered questionnaire was classified as having an allergy or pet. For foods, the verbatim question, which did not require validation by a physician, was: "I am allergic to ___ (mark all that apply): Peanuts, Tree nuts, Shellfish, Other, I have no food allergies that I know of." For non-foods, there were three verbatim questions: "Do you have any of the following non-food allergies? Mark all that apply: Drug (e.g. Penicillin), Pet dander, Beestings, Poison ivy/oak"; "Do you have seasonal allergies? Yes/No"; and Have you been diagnosed with any of the following conditions (check all which apply)? … (e) Asthma, Cystic Fibrosis or Lung Disease.… (v) Skin Condition.…" Thus, the allergies included four foods (peanuts, tree nuts, shellfish, other food) and six non-foods [drug, bee sting, dander, asthma, seasonal, and eczema (specified in skin conditions)]. For pets, the questions were: "Do you have a dog?" and "Do you have a cat?" Participants with an affirmative response were compared to participants without an affirmative response. In sensitivity analyses (specifically, dander allergy with dog or cat ownership in Supplemental Online Content), excluding participants with uncertain or no response reduced sample size and statistical power but had no substantive effect on the associations. We previously noted that AGP participants are widely scattered across the US and resemble the American adult population with respect to the prevalence of cesarean birth and appendectomy, but they are overwhelmingly non-Hispanic Caucasian (93%) and non-smokers (96%) (Goedert et al., 2014). In like manner for the current analysis, we compared the prevalence of allergies reported in AGP data to the prevalence of clinical allergens that were self-reported in representative samples of the US population, particularly the National Health and Nutrition Examination Survey 2005–2006 (Hoppin et al., 2011; Liu et al., 2010; Salo et al., 2011; Visness et al., 2009).

2.2. Richness, Alpha Diversity and Individual Taxa Tests

We examined allergy associations with the number of observed species (richness) and with conventional alpha diversity metrics (Shannon index, Chao1 and PD_whole_tree). Unconditional logistic regression was used to examine associations between microbiome metrics and binary allergy traits, quantified as the odds ratio (OR) and 95% confidence interval (CI). Negative binomial regression was used to examine associations with total numbers of allergy traits. All regression models were adjusted for age, sex, body mass index (BMI), season (spring, summer, fall and winter), time since last antibiotic use (2–6 months, 6–12 months, >12 months), probiotic and vitamin use. We also tested whether the associations between microbiome features and non-food allergies were confounded by food allergies by adjusting for the food allergies in the regression model.

After excluding taxa with relative abundances <0.1%, 223 taxa from the phylum level to the species level were left. Many taxa were very highly correlated, which may uncover trivial duplicate associations. We calculated pairwise Pearson correlations for relative abundances of the 223 taxa and performed pruning using Pearson correlation coefficient 0.95 as a cutoff. For a pair of highly correlated taxa at different levels, we selected the lower level taxon for association analysis. For each pair of highly correlated taxa at the same level, we randomly included one taxon for analysis. After correlation pruning, we had 122 taxa (subsequently termed "non-redundant").

2.3. Composition (Beta Diversity) Test

Weighted and unweighted UniFrac distance matrices were derived from the QIIME pipeline. For each allergy trait, we used the Microbiome Regression-based Kernel Association Test (MiRKAT) (Zhao et al., 2015), a kernel-based regression method, for testing whether microbiome composition differed between cases and controls using either the weighted or unweighted UniFrac distance matrix. The associations were adjusted for sex, age, BMI, season, time since last antibiotic use, probiotic and vitamin use. For each significant association, we used MiRKAT to run 100,000 permutations to verify the asymptotic P-value approximations. We identified significant associations by controlling false discovery rate (FDR) < 10%. We also performed principal coordinate analysis (PCoA) to derive the top three PCoA scores and examined their associations with allergy traits.

2.4. Specific Taxa Associated With Multiple Allergy Traits

We performed standard pairwise association analysis followed by false discovery rate (FDR) correction to identify significant associations between taxon/allergy pairs, which turned out to have limited statistical power because of the heavy multiple testing burden. We observed that some taxa were modestly associated with multiple allergy traits. Thus, we developed a statistical testing framework, following Siegmund et al. (2011), to identify individual taxa associated with multiple allergy traits. The test improved statistical power by aggregating weak associations across traits. The significance was evaluated by 100,000 random permutations, which automatically accounted for the correlations among allergy traits. Details are in the Supplemental Online Content. We applied the testing procedure to 122 non-redundant taxa and produced 122 P-values. We identified taxa significantly associated with multiple allergy traits by controlling FDR at 10% based on these P-values.

3. Results

3.1. Characteristics of the Population

The 1879 participants had a mean age of 45.5 years (standard deviation 15.7 years), with a majority of women (53.1%). Most participants (81.5%) self-reported at least one allergy. Each allergy's prevalence and its associations with other questionnaire variables are presented in Table 1. Approximately 3% of participants reported allergy to peanuts, tree nuts, or shellfish; and 9.1% reported allergy to other foods. Among the six non-foods, allergy prevalence ranged from 4.7% for bee sting to 40.5% for seasonal. At least one food and one non-food allergy was reported by 235 participants. Asthma, dander and drug allergy prevalence in the AGP were lower than the prevalence of these reported in the literature (Table 1), but otherwise prevalence was similar for most of the other allergies. It must be noted that the prevalence of an allergy is lower and its specificity is higher with diagnosis by a doctor, compared to when an allergy is merely self-reported (Hoppin et al., 2011). More allergies were reported by women than men, especially drug allergy ($P = 1.9 \times 10^{-7}$). Higher BMI was associated with more food and total allergies, and especially with seasonal allergy ($P = 7.6 \times 10^{-6}$). Probiotic use was associated with more total allergies ($P = 0.0008$). Otherwise, allergies had no or only modest associations with potential confounding variables (Table 1). Dog and cat ownership was reported by 578 and 555 participants, respectively. Allergy prevalence was not associated with ownership of a dog or cat (data not presented).

3.2. Characteristics of the Fecal Microbiota

Fecal microbiota profiles of the 1879 participants were mapped to 2379 distinct prokaryote taxa. These included 223 taxa with a mean relative abundance of 0.1% or higher, ranging from 48.3% for the phylum Firmicutes to 0.1% for *Brevundimonas diminuta* in the class Alphaproteobacteria (Suppl. Table 1).

3.3. Richness and Alpha Diversity with Allergies

Richness (the number of observed species) in the fecal microbiota was strongly and negatively associated with each allergy except bee sting, asthma, and eczema (Fig. 1). Low richness also was strongly associated with the number of non-food allergies ($P = 2.7 \times 10^{-6}$), food allergies ($P = 8.8 \times 10^{-7}$) and all allergies ($P = 9.1 \times 10^{-9}$). The strongest non-food associations were with drug allergy ($P = 1.2 \times 10^{-5}$) and seasonal allergy ($P = 5.1 \times 10^{-5}$), and the strongest food association was with peanut allergy ($P = 2.4 \times 10^{-7}$). Associations were similar with three estimates of alpha diversity (Fig. 1 and Suppl. Fig. 1). Table 2 presents the magnitude of the associations for each allergy across all estimates of alpha diversity. The odds ratios (OR) for lowest vs highest tertile of richness were 1.8 (CI 1.3–2.5) for drug allergy, 1.7 (CI 1.3–2.2) for seasonal allergy, and 7.8 (CI 2.3–26.5) for peanut allergy.

3.4. Composition (Beta Diversity) With Allergies

Composition of the fecal microbial population was significantly altered with all allergy traits except asthma (Fig. 2). The strongest associations with unweighted UniFrac distance were found with peanut ($P = <1 \times 10^{-7}$), shellfish ($P = 5.1 \times 10^{-5}$), tree nut ($P = 2.5 \times 10^{-4}$), seasonal ($P = 9.1 \times 10^{-5}$), and drug ($P = 4.7 \times 10^{-6}$) allergies. In contrast, weighted UniFrac distance only showed weak associations with seasonal, tree nut, and peanut allergies, and no association with other allergies. By unweighted UniFrac distance, composition was increasingly altered with a larger number of allergies, both to foods and non-foods (Fig. 2).

To further understand the associations between microbiota composition and allergies, we derived the top three PCoA scores based on the unweighted UniFrac distance matrix and tested the associations

Table 1
Selected characteristics of 1879 participants in the American Gut Project and associations with allergy traits.[a]

Allergy traits	Prevalence		Associations with allergy traits															
	AGP	U.S. adults	Age		Sex		BMI		Cesarean birth		Probiotic use		Vitamin use		Time since last antibiotic use			
															2 ~ 6 months		7 ~ 12 months	
			Mean = 45.5, Sd = 15.7		Male = 807		Mean = 23.9, Sd = 5.2		N = 176		N = 227		N = 915		N = 225		N = 261	
			β	P	β	P	β	P	β	P	β	P	β	P	β	P	β	P
Peanuts	2.5%	1.3% (Liu et al., 2010)	−0.015	0.20	−0.65	0.08	−0.061	0.15	0.20	0.70	1.17	0.002	0.44	0.20	0.13	0.79	0.28	0.55
Tree nuts	3.2%	NE	−0.014	0.17	−0.35	0.26	−0.056	0.15	−0.59	0.33	0.05	0.91	0.19	0.54	0.10	0.84	0.51	0.19
Shellfish	2.6%	1.0% (Liu et al., 2010)[b]	0.007	0.53	−0.47	0.15	0.071	0.0002	0.39	0.45	0.60	0.13	0.27	0.40	0.36	0.39	−0.11	0.82
Other food	9.1%	NE	0.0006	0.93	−0.50	0.007	−0.001	0.95	−0.08	0.80	0.49	0.04	0.32	0.09	0.29	0.27	0.37	0.12
Asthma	8.4%	14.1% (Liu et al., 2010)	−0.013	0.05	−0.83	0.00006	0.057	0.00005	0.35	0.18	0.30	0.124	−0.01	0.98	0.38	0.15	0.32	0.24
Bee sting	4.7%	19% (Wood et al., 2014)	0.009	0.30	−0.58	0.02	0.017	0.38	−0.89	0.14	−0.70	0.14	−0.24	0.32	−0.19	0.63	−0.69	0.11
Dander	14.6%	~32% (Wegienka et al., 2011)[c]	0.0007	0.90	−0.04	0.78	0.018	0.18	−0.08	0.76	0.62	0.002	0.20	0.17	−0.18	0.45	0.25	0.22
Eczema	17.8%	~17% (Hanifin et al., 2007)[e]	−0.002	0.68	−0.32	0.02	0.028	0.02	−0.34	0.18	0.34	0.09	0.10	0.48	0.36	0.07	0.22	0.27
Drug	18.8%	33% (Wood et al., 2014)	0.018	0.0001	−0.74	***1.9×10^{-7}***	−0.007	0.59	0.19	0.42	0.00	1.00	0.10	0.46	0.53	0.006	0.11	0.58
Seasonal	40.6%	~40% (Hoppin et al., 2011)[d]	0.004	0.31	−0.01	0.90	0.049	***7.6×10^{-6}***	−0.16	0.40	0.30	0.07	0.26	0.02	0.17	0.30	0.21	0.19
Non-foods	NE	NE	0.003	0.14	−0.24	0.00003	0.021	0.00001	0.20	0.01	0.12	0.036	−0.08	0.40	0.12	0.13	0.17	0.04
Foods	NE	NE	−0.006	0.30	−0.47	0.0036	−0.001	0.96	0.52	0.01	0.31	0.049	−0.08	0.77	0.29	0.19	0.28	0.23
All allergies	81.5%	53% (Hoppin et al., 2011)	0.002	0.42	−0.26	0.00001	0.018	0.00047	0.25	0.00	0.15	0.013	−0.08	0.45	0.14	0.10	0.18	0.04

[a] Highly significant associations ($P < 10^{-5}$) in ***bold italic***; moderately significant associations ($P < 0.001$) in underline. Sd, standard deviation. NE, prevalence not evaluated for general U.S. population.
[b] Shrimp allergy.
[c] In Wood et al. (2014) 8.5% reported severe allergic reaction to animals.
[d] Doctor-diagnosed hay fever (10.8%) plus rhinitis in the past 12 months without cold or influenza (33.4%).
[e] Based on a self-administered questionnaire sent to a sample of households (N = 60,000) representative of the US population. Of the population studied, 17.1% reported at least one of four eczematous symptoms; empirically defined eczema was found in 10.7%, and empirically defined atopic dermatitis was found in 6%. Physician-diagnosed eczema was 9.1% in NHANES 2005–2006 (Liu et al., 2010).

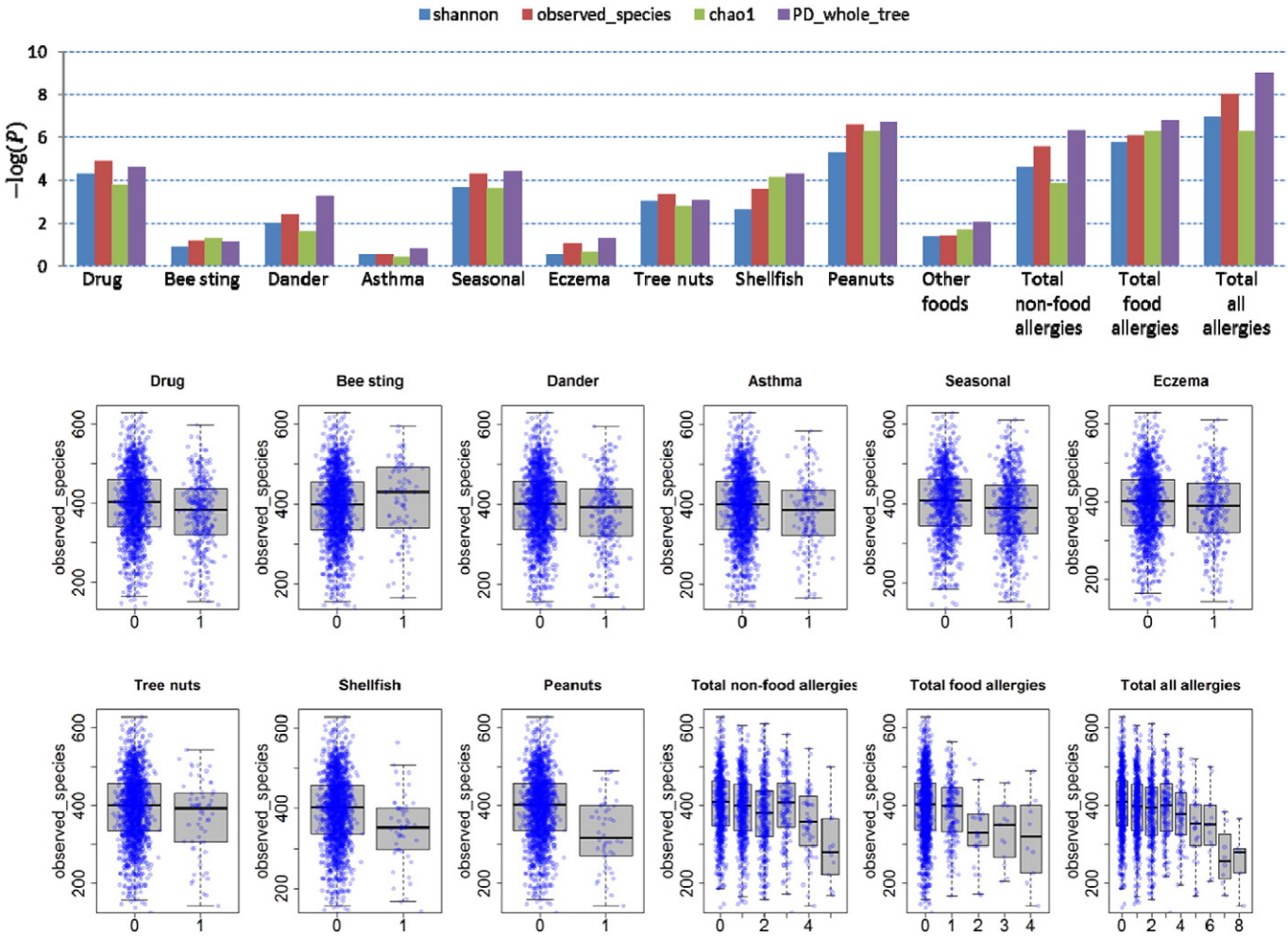

Fig. 1. Associations between richness (observed species), alpha diversity and allergies. Upper panel: Association P-values were derived by unconditional logistic regression (for each individual allergy) or by negative binomial regressions (for total allergies), adjusting for age, sex, body mass index (BMI), time since last antibiotics, season, probiotic and vitamin usage. Lower panels: Box plots for the associations of richness (observed_species). Box plots for Shannon index, Chao1 and PD_whole_tree are reported in Fig. E1. In each box plot, "0" and "1" represent the group without and with the specified allergy, respectively.

between these PCoA scores and the allergy traits (Fig. 2 for PCoA1; Suppl. Fig. 2 for PCoA2 and PCoA3). The three PCoA scores explained 17.5%, 4.4% and 3.2% of the model variance, respectively. The associations with PCoA scores were consistent with but weaker than those with the UniFrac distance matrix, suggesting that the top PCoA scores contain useful but incomplete information.

Table 2
Odds ratio (OR) and 95% confidence interval (CI) for associations of ten allergies with fecal microbiota richness and alpha diversity.

OR (95% CI) by tertile[a]		Drug	Bee sting	Dander	Asthma	Seasonal	Eczema	Tree nuts	Shellfish	Peanuts	Other food
Shannon index	M	1.72 (1.22–2.41)	0.37 (0.20–0.68)	1.18 (0.82–1.69)	1.03 (0.642–1.64)	1.27 (0.981–1.64)	1.01 (0.73–1.41)	2.00 (0.79–5.03)	2.00 (0.74–5.44)	4.15 (1.15–15)	1.52 (0.96–2.40)
	L	***2.00 (1.43–2.79)***	0.69 (0.41–1.17)	1.72 (1.22–2.43)	1.35 (0.85–2.13)	***1.71 (1.32–2.21)***	1.19 (0.86–1.65)	***3.62 (1.54–8.52)***	5.02 (2.02–12.40)	***8.57 (2.54–28.9)***	1.87 (1.2–2.92)
Richness	M	1.57 (1.12–2.20)	0.48 (0.27–0.86)	1.44 (1.0–2.07)	1.68 (1.04–2.71)	1.43 (1.1–1.85)	1.09 (0.78–1.52)	2.29 (0.93–5.64)	3.11 (1.11–8.71)	4.51 (1.28–15.9)	1.27 (0.82–1.99)
	L	***1.83 (1.31–2.54)***	0.66 (0.38–1.13)	1.83 (1.29–2.61)	1.52 (0.94–2.47)	***1.73 (1.33–2.24)***	1.22 (0.88–1.70)	***3.19 (1.34–7.62)***	***5.41 (2.03–14.5)***	***7.78 (2.29–26.5)***	1.45 (0.94–2.24)
Chao1	M	1.33 (0.95–1.86)	0.40 (0.22–0.73)	1.08 (0.76–1.55)	1.58 (0.99–2.52)	1.15 (0.888–1.49)	1.20 (0.86–1.67)	1.30 (0.56–3.01)	3.08 (1.1–8.61)	2.80 (0.87–8.95)	1.51 (0.96–2.38)
	L	1.68 (1.22–2.33)	0.61 (0.36–1.04)	1.42 (1.01–2.00)	1.31 (0.82–2.11)	1.67 (1.29–2.16)	1.29 (0.93–1.80)	2.24 (1.04–4.83)	***5.48 (2.06–14.6)***	***6.79 (2.3–20)***	1.73 (1.11–2.70)
PD whole tree	M	1.78 (1.27–2.50)	0.46 (0.26–0.82)	1.27 (0.884–1.83)	1.49 (0.92–2.42)	1.50 (1.16–1.94)	1.00 (0.72–1.41)	1.70 (0.70–4.11)	2.33 (0.79–6.81)	2.35 (0.73–7.64)	1.45 (0.91–2.30)
	L	***1.93 (1.38–2.71)***	0.68 (0.40–1.17)	1.77 (1.25–2.52)	1.52 (0.95–2.44)	1.65 (1.27–2.15)	1.34 (0.97–1.86)	2.86 (1.25–6.54)	***6.29 (2.38–16.6)***	***6.56 (2.23–19.3)***	1.86 (1.19–2.90)

[a] Odds ratio [OR, and 95% confidence interval (CI)] estimates for middle (M) and lowest (L) tertile, versus highest tertile, for fecal microbiota richness and alpha diversity estimates (Shannon index, Chao1, and PD whole tree). All models were adjusted for age, sex, body mass index (BMI), season (spring, summer, fall and winter), time since last antibiotics use (2–6 months, 6–12 months, >12 months), probiotic and vitamin use. Highly significant associations (lower CI >1.3) in ***bold italic***; moderately significant associations (lower CI 1.0–1.3) in underline.

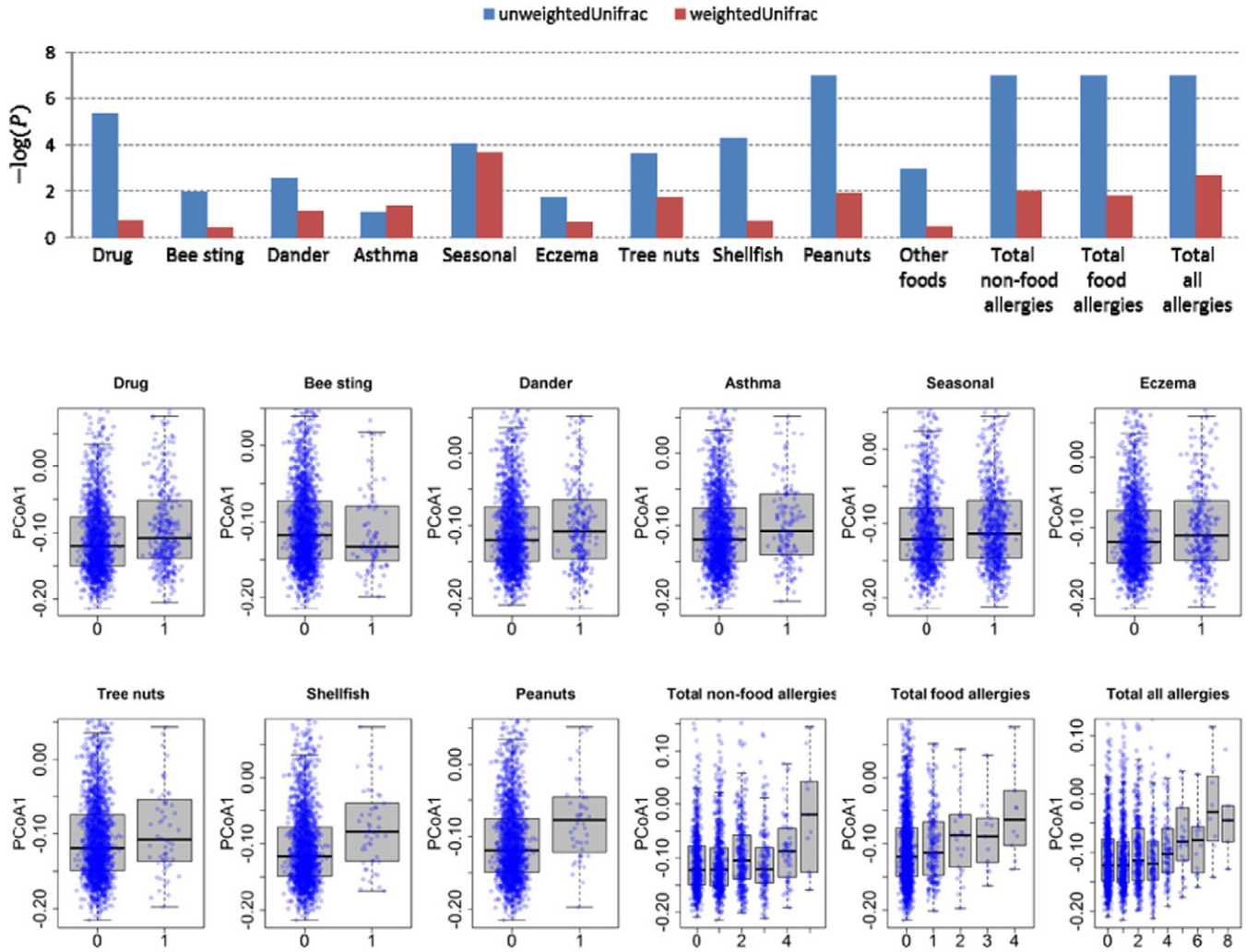

Fig. 2. Association between microbiome composition (beta diversity) and allergies. Upper panels: Association P-values were calculated by MiRKAT (Zhao et al., 2015) using unweighted and weighted UniFrac distance matrices. Associations were adjusted for age, sex, BMI, time since last antibiotics, season, probiotic and vitamin usage. Lower panels: Box plots of the top PCoA scores based on unweighted UniFrac distance matrix. P-values were based on logistic regression for each individual allergy and by negative binomial regressions for total allergies. Box plots for PCoA2 and PCoA3 are reported in Fig. E2. The top three PCoA scores explained 17.5%, 4.4% and 3.2% of the variance.

3.5. Correlations and Sensitivity Analyses

There was a moderate correlation between total food and non-food allergies (R = 0.325), stronger correlation between peanut and tree nut allergies (R = 0.44), and weaker correlations for the other food allergy pairs (range R = 0.18–0.27). Except for a moderate correlation of dander with seasonal allergies (R = 0.34) and with asthma (R = 0.26), all other correlation pairs were weak (R ≤ 0.20) or essentially null (R < 0.10, Suppl. Table 2).

Suppl. Table 3 presents sensitivity analyses for possible confounding or mediation by cesarean birth, ownership of a dog or cat, and food allergies. Each of these variables was added to the richness, alpha diversity, and beta diversity models. Based on attenuation of P-values, little or no evidence of confounding or mediation was found, with one exception. P-value increased about 2 logs when food allergies were added to the total non-food allergy models for alpha diversity and richness (e.g., $P = 2.67 \times 10^{-6}$ increased to $P = 5.21 \times 10^{-4}$). Likewise, when food allergies were added to the total non-food allergy model for unweighted UniFrac distance, $P < 10^{-8}$ increased to $P = 2.1 \times 10^{-5}$.

3.6. Specific Taxa with Multiple Allergies

Standard analysis for associations between the 1220 pairs of 10 allergies and 122 non-redundant taxa identified only four pairs that were significant at FDR < 10% (nominal P < 0.00035, Suppl. Table 4). To test the associations of specific taxa with multiple allergies at FDR < 10%, we aggregated the allergies and used 100,000 permutations (see Methods). As shown in Fig. 3 and Suppl. Table 5, both drug and bee sting allergies were associated, albeit in opposite directions, with *g__Paraprevotella* and two Alphaproteobacteria taxa. Peanut and tree nut allergies were positively associated with *g__Bacteroides* and with two closely related taxa (*g__Bacteroides*;*s__fragilis* and *o__Bacteroidales*), and negatively associated with *o__Clostridiales*, *g__Prevotella*;*s__*, and *f__Ruminococcaceae* (Fig. 3). Seasonal allergy was associated with ten taxa, resembling the nut allergy taxa. Subsets of the seasonal taxa were also associated, although more weakly, with asthma, eczema and dander allergy. With negative binomial regression models to adjust for all covariates, 8 taxa were associated with total food allergies, and 8 taxa were associated with total non-food allergies (Fig. 3), corroborating the multiple trait results.

4. Discussion

In this American Gut Project population, 4% of whom were ages 4–17 and two-thirds of whom were between ages 30–62, the prevalence of one or more self-reported allergies was 81.5%. About 3% reported peanut or tree nut allergy, and nearly 41% reported seasonal allergy. Women reported more drug allergy, and obese participants reported

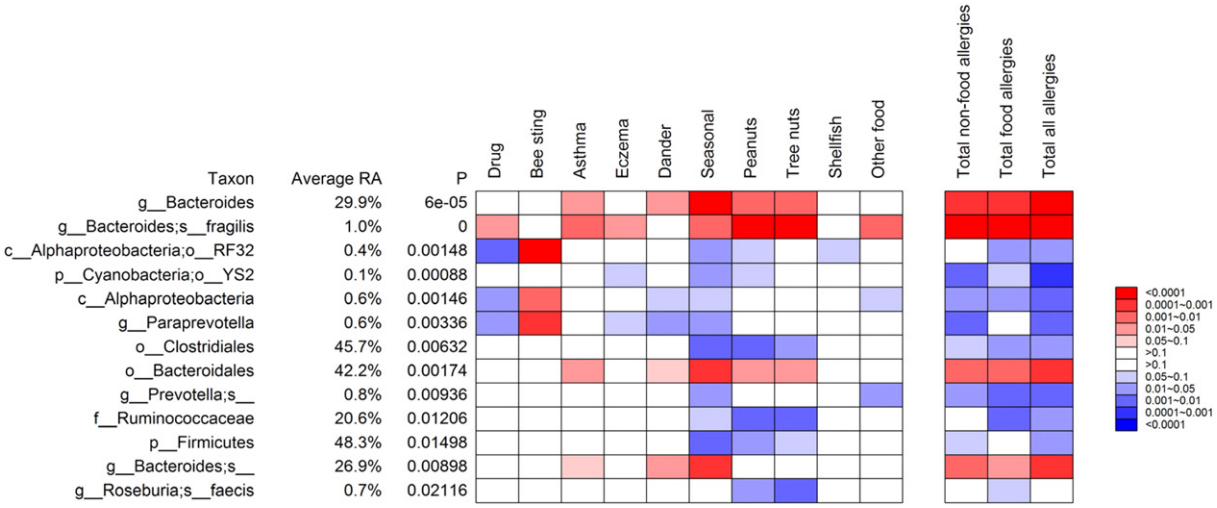

Fig. 3. Taxa associated with multiple allergy traits. We identified 13 taxa [false discovery rate (FDR) < 10%] associated with multiple food or non-food allergies. Each taxon's shorthand name, average relative abundance (RA), and a P-value for testing multiple allergy traits based on 100,000 permutations are presented. The heat map shows statistically significant, covariate-adjusted P-values (red for positive, blue for negative) for associations with total allergies (by negative binomial models) and with each allergy (by logistic regression). The Z-scores and P-values for individual taxon associations are in Table E5.

more seasonal allergy. Pet ownership was irrelevant for our adults, which is consistent with previous null or ambiguous associations that may reflect exposure to dogs or particularly to cats outside the home or in early childhood (Dharmage et al., 2012; Simpson and Custovic, 2005). Correlations between the allergies were null or weak, except between dander and seasonal ($R = 0.34$) and between peanut and tree nut ($R = 0.44$). Independent of these within-subject correlations, we found statistically significant fecal dysbiosis across multiple allergies. Specifically, reduced richness and altered composition was found with all allergies except asthma, bee sting, and eczema. The dysbiosis was most marked with allergies to nuts and seasonal pollen, and it was driven by higher abundance of *Bacteroidales* and reduced abundance of *Clostridiales*.

The source of allergy-associated dysbiosis is unknown. One possibility is cesarean birth (Penders et al., 2014). In a longitudinal, fecal microbiome study of Swedish infants, cesarean delivery was associated with significantly delayed *Bacteroides* colonization and lower blood levels of Th1 cytokines (CXCL10 and CXCL11) but no difference in Th2 cytokines (Jakobsson et al., 2014). Cesarean delivered infants also had low diversity of *Bacteroidetes* taxa. By age 24 months, *Bacteroides* and *Clostridia* taxa predominated irrespective of delivery route (Jakobsson et al., 2014). The Swedish study is consistent with reports that cesarean-delivered children have a higher risk for developing asthma (1.2-fold), atopic sensitization (1.7-fold), and allergic rhinitis (2-fold) (Kolokotroni et al., 2012; Pistiner et al., 2008; Thavagnanam et al., 2008). In a Danish study based on 16S rRNA fingerprinting, lower fecal bacterial diversity at age 1 month was not associated with cesarean delivery, but it was associated with increased likelihood of allergic rhinitis (1.3-fold) by age 6 years (Bisgaard et al., 2011). Consistent with the Danish study, allergy was not associated with cesarean birth in the current study. Moreover, although we found that both cesarean birth and allergies were associated with low richness, they were associated with different taxa (Goedert et al., 2014). These findings suggest that the dysbiosis of allergy in adults develops postnatally.

In our study, the gut microbiota was not associated asthma or eczema, and only narrowly associated with bee sting allergy. The AGP questionnaire asked "Have you been diagnosed with … Asthma, Cystic Fibrosis or Lung Disease", implying validation by a health care professional. However, grouping with cystic fibrosis and lung disease would clearly reduce specificity for asthma. Moreover, these respiratory conditions were not explicitly linked to allergy. Non-specificity would be expected to inflate asthma prevalence, which was not seen, as prevalence was lower in the AGP (8.4%) than reported for the US population (14.1%) (Liu et al., 2010). Specificity may also have been poor for eczema, as this was one of several "skin conditions" in the AGP questionnaire. In addition, even true atopic eczema, with 3% estimated prevalence in developed countries (Eyerich and Novak, 2013), is not a simple Th2-mediated disease but also involves genetic and non-genetic dermal barriers (Eyerich and Novak, 2013). Systemic reaction to bee or wasp sting, which also has a prevalence of approximately 3% (Golden, 2013), was unrelated to atopy in three previous studies (Birnbaum et al., 1994; Fernandez et al., 1999; Golden et al., 1989). We found that bee sting allergy was significantly associated with higher abundance of Alphaproteobacteria and *g__Paraprevotella* but not with global microbiota metrics.

The implications of the hygiene hypothesis extend beyond allergy per se, as exemplified by the reduced risk of non-Hodgkin lymphoma for adults who were the first-born or only child in their family, and for those who developed allergy to pollen and perhaps to foods (Cozen et al., 2007; Grulich et al., 2005; Smedby et al., 2007). There are similar birth-order associations with young-adult Hodgkin lymphoma (Chang et al., 2004; Gutensohn et al., 1975; Mack et al., 2015; Westergaard et al., 1997); survivors of this malignancy had low diversity of the fecal microbiota (Cozen et al., 2013), an association compatible with the hygiene hypothesis but also with cancer and its treatment. Human microbiota studies of other conditions linked to the hygiene hypothesis are lacking.

Our study had several important weaknesses. Some of the allergies were probably misclassified, because the questionnaire data were self-reported with no validation by a physician or an objective test. This would bias toward the null, by introducing noise into a true association between a condition and an exposure. Our cross-sectional design cannot distinguish whether the observed dysbiosis preceded or followed the development of allergies. The specimens represented only one time point, but this also would bias toward the null, reducing the chance to detect associations with time-varying fecal microbiome metrics. We did not investigate possible associations with seasonality, but rather adjusted for it. With this approach, we would not have identified transient changes as reported for pollen season in Japan (Odamaki et al., 2007). Avoidance of particular foods could alter the microbiota, but intake of these foods (peanuts, tree nuts, and shellfish) is probably too low to affect the microbiota of the general population. Antihistamines or other medications may have altered the allergic participants' microbiota, although this has not been reported. We adjusted for reported use of

probiotics and vitamins, based on the possibility that their use may have altered the microbiota. We excluded participants who used an antibiotic within one month, and adjusted analyses for earlier antibiotic use.

In conclusion, American adults with allergies, especially but not exclusively to nuts and seasonal pollen, have lower richness and altered composition of their gut microbiota. This observation of an allergy-associated dysbiosis supports the hygiene hypothesis, but the origin of the dysbiosis is unknown. Also unknown is whether prevention or amelioration of the dysbiosis can modify allergy prevalence or severity (Costa et al., 2014; Singh et al., 2013). Clinical trials and other longitudinal studies that incorporate fecal microbiota characterization will be needed to address these questions.

Author contributions

JJG and JS conceived and designed the study, interpreted the data, and drafted the manuscript; XH developed the analytic pipeline, led the analysis, and generated the tables and figures; GY processed the raw data and generated the OTU table; AP cleaned the phenotype data, analyzed the pet data, and assisted with the Introduction; all authors contributed to and approved the final version of the manuscript.

Acknowledgments

The authors are grateful to Drs. Howard Young and Emily Vogtmann for comments on the manuscript and especially to the participants. This project was supported by the Intramural Research Program of the National Cancer Institute (ZIA-CP-010214). As noted at http://microbio.me/americangut/FAQ.psp, the American Gut Project is a donation-supported academic effort and collaboration between the Earth Microbiome Project and the Human Food Project, all of which are primarily run out of the laboratory of Rob Knight at the University of California San Diego. The website notes that all data except information that needs to be kept confidential for privacy reasons is openly and freely released into public databases.

Appendix A. Supplementary methods and data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2015.11.038.

References

Abrahamsson, T.R., Jakobsson, H.E., Andersson, A.F., Bjorksten, B., Engstrand, L., Jenmalm, M.C., 2012. Low diversity of the gut microbiota in infants with atopic eczema. J. Allergy Clin. Immunol. 129, 434–440 (440 e431-432).

Abrahamsson, T.R., Jakobsson, H.E., Andersson, A.F., Bjorksten, B., Engstrand, L., Jenmalm, M.C., 2014. Low gut Microbiota diversity in early infancy precedes asthma at school age. Clin. Exp. Allergy 44, 842–850.

Arrieta, M.C., Stiemsma, L.T., Amenyogbe, N., Brown, E.M., Finlay, B., 2014. The intestinal microbiome in early life: health and disease. Front. Immunol. 5, 427.

Birnbaum, J., Vervloet, D., Charpin, D., 1994. Atopy and systemic reactions to hymenoptera stings. Allergy Proc. 15, 49–52.

Bisgaard, H., Li, N., Bonnelykke, K., Chawes, B.L., Skov, T., Paludan-Muller, G., Stokholm, J., Smith, B., Krogfelt, K.A., 2011. 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 (e641-645).

Bowman, L.M., Holt, P.G., 2001. Selective enhancement of systemic th1 immunity in immunologically immature rats with an orally administered bacterial extract. Infect. Immun. 69, 3719–3727.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., et al., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336.

Chang, E.T., Montgomery, S.M., Richiardi, L., Ehlin, A., Ekbom, A., Lambe, M., 2004. Number of siblings and risk of Hodgkin's lymphoma. Cancer Epidemiol. Biomark. Prev. 13, 1236–1243.

Costa, D.J., Marteau, P., Amouyal, M., Poulsen, L.K., Hamelmann, E., Cazaubiel, M., Housez, B., Leuillet, S., Stavnsbjerg, M., Molimard, P., et al., 2014. Efficacy and safety of the probiotic *Lactobacillus paracasei* LP-33 in allergic rhinitis: a double-blind, randomized, placebo-controlled trial (GA2LEN Study). Eur. J. Clin. Nutr. 68, 602–607.

Cozen, W., Cerhan, J.R., Martinez-Maza, O., Ward, M.H., Linet, M., Colt, J.S., Davis, S., Severson, R.K., Hartge, P., Bernstein, L., 2007. The effect of atopy, childhood crowding, and other immune-related factors on non-Hodgkin lymphoma risk. Cancer Causes Control 18, 821–831.

Cozen, W., Yu, G., Gail, M.H., Ridaura, V.K., Nathwani, B.N., Hwang, A.E., Hamilton, A.S., Mack, T.M., Gordon, J.I., Goedert, J.J., 2013. Fecal microbiota diversity in survivors of adolescent/young adult Hodgkin lymphoma: a study of twins. Br. J. Cancer.

Dharmage, S.C., Lodge, C.L., Matheson, M.C., Campbell, B., Lowe, A.J., 2012. Exposure to cats: update on risks for sensitization and allergic diseases. Curr. Allergy Asthma Rep. 12, 413–423.

Ege, M.J., Mayer, M., Normand, A.C., Genuneit, J., Cookson, W.O., Braun-Fahrlander, C., Heederik, D., Piarroux, R., von Mutius, E., Group, G.T.S., 2011. Exposure to environmental microorganisms and childhood asthma. N. Engl. J. Med. 364, 701–709.

Eyerich, K., Novak, N., 2013. Immunology of atopic eczema: overcoming the Th1/Th2 paradigm. Allergy 68, 974–982.

Fernandez, J., Blanca, M., Soriano, V., Sanchez, J., Juarez, C., 1999. Epidemiological study of the prevalence of allergic reactions to Hymenoptera in a rural population in the Mediterranean area. Clin. Exp. Allergy 29, 1069–1074.

Fujimura, K.E., Lynch, S.V., 2015. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. Cell Host Microbe 17, 592–602.

Gibson, P.G., Henry, R.L., Shah, S., Powell, H., Wang, H., 2003. Migration to a western country increases asthma symptoms but not eosinophilic airway inflammation. Pediatr. Pulmonol. 36, 209–215.

Goedert, J.J., Hua, X., Yu, G., Shi, J., 2014. Diversity and composition of the adult fecal microbiome associated with history of cesarean birth or appendectomy: analysis of the American Gut Project. EBioMedicine 1, 167–172.

Golden, D.B., 2013. Advances in diagnosis and management of insect sting allergy. Ann. Allergy Asthma Immunol. 111, 84–89.

Golden, D.B., Marsh, D.G., Kagey-Sobotka, A., Freidhoff, L., Szklo, M., Valentine, M.D., Lichtenstein, L.M., 1989. Epidemiology of insect venom sensitivity. JAMA 262, 240–244.

Grulich, A.E., Vajdic, C.M., Kaldor, J.M., Hughes, A.M., Kricker, A., Fritschi, L., Turner, J.J., Milliken, S., Benke, G., Armstrong, B.K., 2005. Birth order, atopy, and risk of non-Hodgkin lymphoma. J. Natl. Cancer Inst. 97, 587–594.

Gutensohn, N., Li, F.P., Johnson, R.E., Cole, P., 1975. Hodgkin's disease, tonsillectomy and family size. N. Engl. J. Med. 292, 22–25.

Hanifin, J.M., Reed, M.L., Eczema, P., Impact Working, G., 2007. A population-based survey of eczema prevalence in the United States. Dermatitis 18, 82–91.

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Makela, M.J., et al., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc. Natl. Acad. Sci. U. S. A. 109, 8334–8339.

Hoppin, J.A., Jaramillo, R., Salo, P., Sandler, D.P., London, S.J., Zeldin, D.C., 2011. Questionnaire predictors of atopy in a US population sample: findings from the National Health and Nutrition Examination Survey, 2005–2006. Am. J. Epidemiol. 173, 544–552.

Hrncir, T., Stepankova, R., Kozakova, H., Hudcovic, T., Tlaskalova-Hogenova, H., 2008. Gut microbiota and lipopolysaccharide content of the diet influence development of regulatory t cells: studies in germ-free mice. BMC Immunol. 9, 65.

Jakobsson, H.E., Abrahamsson, T.R., Jenmalm, M.C., Harris, K., Quince, C., Jernberg, C., Bjorksten, B., Engstrand, L., Andersson, A.F., 2014. Decreased gut microbiota diversity, delayed bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. Gut 63, 559–566.

Kalliomaki, M., Kirjavainen, P., Eerola, E., Kero, P., Salminen, S., Isolauri, E., 2001. Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing. J. Allergy Clin. Immunol. 107, 129–134.

Kolokotroni, O., Middleton, N., Gavatha, M., Lamnisos, D., Priftis, K.N., Yiallouros, P.K., 2012. Asthma and atopy in children born by caesarean section: effect modification by family history of allergies — a population based cross-sectional study. BMC Pediatr. 12, 179.

Liu, A.H., Jaramillo, R., Sicherer, S.H., Wood, R.A., Bock, S.A., Burks, A.W., Massing, M., Cohn, R.D., Zeldin, D.C., 2010. National prevalence and risk factors for food allergy and relationship to asthma: results from the National Health and Nutrition Examination Survey 2005–2006. J. Allergy Clin. Immunol. 126, 798–806 (e713).

Mack, T.M., Norman Jr., J.E., Rappaport, E., Cozen, W., 2015. Childhood determination of Hodgkin lymphoma among U.S. servicemen. Cancer Epidemiol. Biomark. Prev.

Odamaki, T., Xiao, J.Z., Iwabuchi, N., Sakamoto, M., Takahashi, N., Kondo, S., Iwatsuki, K., Kokubo, S., Togashi, H., Enomoto, T., et al., 2007. Fluctuation of fecal microbiota in individuals with Japanese cedar pollinosis during the pollen season and influence of probiotic intake. J. Investig. Allergol. Clin. Immunol. 17, 92–100.

Ohnmacht, C., Park, J.H., Cording, S., Wing, J.B., Atarashi, K., Obata, Y., Gaboriau-Routhiau, V., Marques, R., Dulauroy, S., Fedoseeva, M., et al., 2015. Mucosal immunology. The microbiota regulates type 2 immunity through RORgammat(+) T cells. Science 349, 989–993.

Olszak, T., An, D., Zeissig, S., Vera, M.P., Richter, J., Franke, A., Glickman, J.N., Siebert, R., Baron, R.M., Kasper, D.L., et al., 2012. Microbial exposure during early life has persistent effects on natural killer T cell function. Science 336, 489–493.

Ownby, D.R., Johnson, C.C., Peterson, E.L., 2002. Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age. JAMA 288, 963–972.

Penders, J., Stobberingh, E.E., Thijs, C., Adams, H., Vink, C., van Ree, R., van den Brandt, P.A., 2006. Molecular fingerprinting of the intestinal microbiota of infants in whom atopic eczema was or was not developing. Clin. Exp. Allergy 36, 1602–1608.

Penders, J., Gerhold, K., Thijs, C., Zimmermann, K., Wahn, U., Lau, S., Hamelmann, E., 2014. New insights into the hygiene hypothesis in allergic diseases: mediation of sibling and birth mode effects by the gut microbiota. Gut Microbes 5, 239–244.

Pistiner, M., Gold, D.R., Abdulkerim, H., Hoffman, E., Celedon, J.C., 2008. Birth by cesarean section, allergic rhinitis, and allergic sensitization among children with a parental history of atopy. J. Allergy Clin. Immunol. 122, 274–279.

Salo, P.M., Calatroni, A., Gergen, P.J., Hoppin, J.A., Sever, M.L., Jaramillo, R., Arbes Jr., S.J., Zeldin, D.C., 2011. Allergy-related outcomes in relation to serum IgE: results from

the National Health and Nutrition Examination Survey 2005–2006. J. Allergy Clin. Immunol. 127, 1226–1235 (e1227).

Sheikh, A., Smeeth, L., Hubbard, R., 2003. There is no evidence of an inverse relationship between TH2-mediated atopy and TH1-mediated autoimmune disorders: lack of support for the hygiene hypothesis. J. Allergy Clin. Immunol. 111, 131–135.

Siegmund, D., Yakir, B., Zhang, N.R., 2011. Detecting simultaneous variant intervals in aligned sequences. Ann. Appl. Stat. 5, 645–668.

Simpson, A., Custovic, A., 2005. Pets and the development of allergic sensitization. Curr. Allergy Asthma Rep. 5, 212–220.

Singh, A., Hacini-Rachinel, F., Gosoniu, M.L., Bourdeau, T., Holvoet, S., Doucet-Ladeveze, R., Beaumont, M., Mercenier, A., Nutten, S., 2013. Immune-modulatory effect of probiotic *Bifidobacterium lactis* NCC2818 in individuals suffering from seasonal allergic rhinitis to grass pollen: an exploratory, randomized, placebo-controlled clinical trial. Eur. J. Clin. Nutr. 67, 161–167.

Smedby, K.E., Hjalgrim, H., Chang, E.T., Rostgaard, K., Glimelius, B., Adami, H.O., Melbye, M., 2007. Childhood social environment and risk of non-Hodgkin lymphoma in adults. Cancer Res. 67, 11074–11082.

Strachan, D.P., 2000. Family size, infection and atopy: the first decade of the "hygiene hypothesis". Thorax 55 (Suppl. 1), S2–10.

Teo, S.M., Mok, D., Pham, K., Kusel, M., Serralha, M., Troy, N., Holt, B.J., Hales, B.J., Walker, M.L., Hollams, E., et al., 2015. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe 17, 704–715.

Thavagnanam, S., Fleming, J., Bromley, A., Shields, M.D., Cardwell, C.R., 2008. A meta-analysis of the association between caesarean section and childhood asthma. Clin. Exp. Allergy 38, 629–633.

Tobias, A., Soriano, J.B., Chinn, S., Anto, J.M., Sunyer, J., Burney, P., European Community Respiratory Health, S., 2001. Symptoms of asthma, bronchial responsiveness and atopy in immigrants and emigrants in Europe. European Community Respiratory Health Survey. Eur. Respir. J. 18, 459–465.

Visness, C.M., London, S.J., Daniels, J.L., Kaufman, J.S., Yeatts, K.B., Siega-Riz, A.M., Liu, A.H., Calatroni, A., Zeldin, D.C., 2009. Association of obesity with IgE levels and allergy symptoms in children and adolescents: results from the National Health and Nutrition Examination Survey 2005–2006. J. Allergy Clin. Immunol. 123, 1163–1169 (1169 e1161-1164).

Wegienka, G., Johnson, C.C., Havstad, S., Ownby, D.R., Nicholas, C., Zoratti, E.M., 2011. Lifetime dog and cat exposure and dog- and cat-specific sensitization at age 18 years. Clin. Exp. Allergy 41, 979–986.

Westergaard, T., Melbye, M., Pedersen, J.B., Frisch, M., Olsen, J.H., Andersen, P.K., 1997. Birth order, sibship size and risk of Hodgkin's disease in children and young adults: a population-based study of 31 million person-years. Int. J. Cancer 72, 977–981.

Wood, R.A., Camargo Jr., C.A., Lieberman, P., Sampson, H.A., Schwartz, L.B., Zitt, M., Collins, C., Tringale, M., Wilkinson, M., Boyle, J., et al., 2014. Anaphylaxis in America: the prevalence and characteristics of anaphylaxis in the United States. J. Allergy Clin. Immunol. 133, 461–467.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al., 2012. Human gut microbiome viewed across age and geography. Nature 486, 222–227.

Zhao, N., Chen, J., Carroll, I.M., Ringel-Kulka, T., Epstein, M.P., Zhou, H., Zhou, J.J., Ringel, Y., Li, H., Wu, M.C., 2015. Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. Am. J. Hum. Genet. 96, 797–807.