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 ($P \le 10^{-4}$). 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.

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

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.