ENVIRONMENTAL MICROBES, CLIMATE, AND ALLERGIC DISEASE

BACKGROUND/ RATIONALE: Allergic diseases such as asthma, atopic dermatitis, and allergic rhinitis have doubled in prevalence over the last 50 years and now affect a third of the global population, constituting a major component of the global disease burden[1,2]. This rapid increase has been most prominent in industrialized and urban areas, and patterns of geographic variation suggest a clear role for protective environmental factors such as proximity to farm animals[3,4]. Environmental microbes, including high levels of bacterial and fungal diversity in household dust, have been repeatedly shown to correlate with lower rates of allergic disease, and prevent allergic sensitization in animal models[5-9]. A body of research showing that microbial exposures are important for proper immune system development provides a likely explanation for these findings[10]. Despite this knowledge, there remains a substantial gap in our understanding of how the environmental microbiome impacts allergic disease. Given the complexity of microbial exposures, large sample sizes and advanced analytic techniques are needed to identify the most relevant exposures and translate these into improvements in health.

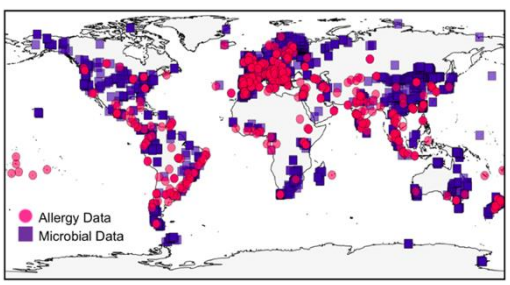
We propose to investigate how the environmental microbiome predicts variation in allergic disease by pairing robust data from a study of nearly two million children with nearby soil, built environment, and livestock samples from six publicly available global microbiome sampling initiatives. Our preliminary data show that soil-related microbial diversity is associated with lower prevalence of lifetime symptoms of asthma, atopic dermatitis, and allergic rhinoconjunctivitis (**Table**). Numerous studies have highlighted how changes in plant and animal communities lag behind climate change, and our prior work suggests this phenomenon also applies to soil fungi and bacteria[11]. Utilizing high-resolution global climate maps, we also plan to assess how environmental microbes equilibrate with changing climate and forecast future rates of allergic disease. <u>The overarching objective of this proposal is to examine associations between environmental microbes, climate, and allergic disease at a global scale.</u>

	Lifetime symptoms of asthma		Lifetime symptoms of atopic dermatitis		Lifetime symptoms of allergic rhinoconjunctivitis	
	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI
Soil-related Shannon microbial diversity score	0.65	0.50 – 0.84	0.79	0.59 – 1.05	0.92	0.74 – 1.13
Animal-related Shannon microbial diversity score	0.80	0.70 – 0.92	0.76	0.61 – 0.95	0.88	0.78 – 0.98

Aim 1: Determine the extent to which microbial communities (including bacteria, archaea, fungi, nematodes, protozoa, and viruses) from soil, the built environment, and livestock are associated with allergic disease prevalence among children. *We hypothesize that greater microbial diversity will be associated with lower rates of asthma, allergic rhinitis, and atopic dermatitis. In addition, we aim to identify the taxa and/or species most closely associated with each subtype of allergic disease.*

Aim 2: Predict future allergy rates based on climate-driven shifts in microbial distributions. *We will develop models based on temporal disequilibria with climatic factors from up to seventy years ago and use lags in equilibration to predict future allergy rates.*

APPROACH: We plan to integrate and analyze three sources of existing publicly available data to achieve the above aims. *<u>Microbiome data</u>:* We have compiled a spatially explicit global database of microbial taxa and genes (using published 16S and18S rRNA, gene-based taxonomic profiles and also metagenome data for some of the samples). This database includes data from 10,957 samples of fungi from the built environment[12], fungi, bacteria, archaea, and nematodes from soils[13-16]; bacteria, archaea, and protozoa from domesticated animals[17], and associated metadata. These sources were selected because of their scope, ease of normalization, and the high quality of data.[18] *<u>Allergic disease data</u>:* The International Study of Asthma and Allergies in Childhood (ISAAC)[19], one of the largest global epidemiologic efforts to date, included standardized assessments of



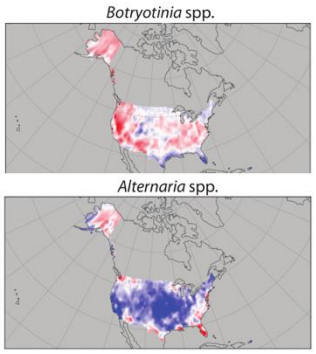
atopic dermatitis, asthma, and rhinitis from 1,962,480 children at 306 sites around the world that overlap with the microbial data (**Figure**). *Climate data:* We will use global maps that include ten key climate variables (e.g., mean temperature, precipitation, potential evapotranspiration) at one-degree resolution for each month beginning in 1950[20].

Aim 1: To determine how microbial exposures are associated with allergic disease, we will employ deep quantile regression along with feature selection procedures (e.g. expected gradients, approximate Shapley values)[21]. We will assess whether emergent attributes of microbial communities (e.g., α-diversity, β-diversity, ɣ-diversity) and the occurrence and relative abundance of individual microbial taxa and genes are associated with asthma, rhinitis, and atopic dermatitis individually, and with a composite outcome of any allergic disease. We will evaluate the importance of multiple potential confounders collected as part of the ISAAC data (including local socioeconomic status, population density, air pollution, ancestry, age, participant reported housing characteristics, antibiotic use, and pet and farm animal exposure) in our regression models, and use variance partitioning methods to assess whether otherwise unexplained variation in allergic disease can be attributed to environmental microbiomes.

A limitation of the available data relates to the geographic proximity of data collection sites: although most microbiome data were collected in close proximity to the school sites from which participants were recruited (90.5% of allergy samples are within 400 km of a microbial sample) the sampling locations were not coincident. We will employ three approaches to evaluate the sensitivity of our results to this limitation:

1. Compare sub-analyses including only data from sites that are in very close geographic proximity.
2. Combine the data within latitude/longitude grid cells, and repeat analyses on coarse-grained data.
3. Interpolate allergy data and microbial data using ecological niche modeling[22].

Aim 2: To assess the links between climate, environmental microbiomes, and allergies, we will consider both the climate data from the month of microbe and allergy data collection, and historic climate data averaged over five and ten-year windows going back to 1950. Our *preliminary data* from the 1000 Homes Project[23] show that geographic distributions of fungal allergens are in temporal disequilibrium with climate change (**Figure**). In particular, the distributions of *Pleospora*, *Alternaria*, and *Botryotinia* spp. appear to be associated with the distribution of climate from up to 50 years ago. Leveraging our previously successful forecasting approach[11], we will predict how microbial distributions would shift if they were to equilibrate to existing climate change and the resulting changes in allergic disease rates. We will use both the raw, untransformed climate data, and where multicollinearity between climate variables is an issue, climate data pruned using Iterative Random Forest Leave One Out Prediction (iRF-LOOP) to replace redundant variables with linear combinations thereof[24].



Projected shifts in the distribution of two key fungal allergens. Red indicates increases in relative abundance and blue indicates decreases.

INNOVATION, IMPACT AND SUSTAINABILITY: Prior research has largely focused on bacteria and/or fungi at local scales; our novel approach integrating data to examine multiple types of microbes from various environmental sources (ie household dust, soil, plants, and animals) is innovative and enables identification of patterns that may only be evident at global scales. Leveraging the expertise of our multidisciplinary team (Dr. Abuabara is an epidemiologist and physician with expertise in allergic disease, Drs. Brown and Jacobson are computational biologists with complementary expertise in machine learning and statistical modeling, Dr. Hess is a microbiologist with expertise in metagenomics and agriculture, and Dr. Ladau is an ecologist with expertise in novel mathematical approaches to analyze the climatic impact on microbial systems), we will develop computational methods that are likely to have a sustained impact on the field. For example, in recent work examining the role of the environmental microbiome in the coronavirus pandemic, we developed a novel quantile regression procedure to predict county-level variation in infection fatality ratios and found that indoor-outdoor fungal beta diversity was highly predictive of this outcome (preliminary results presented at the 2021 ASM meeting; manuscript now under review).

While analyses of host-associated microbiome data can elucidate mechanistic processes, environmental microbiome data linked to an extremely large patient population can be used to understand geographic variation in allergic disease and associations with changing climatic factors over time. We will begin with data from amplicon sequencing to identify taxonomic profiles, and subsequently draw on existing metagenomic data to develop future work more focused on functionality of microbial communities. Although we won't be able to infer causality with area-level associations, the results will enable the design of targeted sample collection. We

plan to apply for funding to collect new data on environmental microbes in the context of the Global Asthma Network (an extension of the ISAAC studies described above with detailed human microbiome, genetic, and immunologic data on participants).

Finally, the proposed work has the potential for widespread and long-lasting impact because it addresses fundamental questions about the consequences of biodiversity loss for human health. Our results will inform future hypothesis testing of modifiable risk factors driving global variation in the environmental microbiome and disease. For example, prior research has found that rumen microbes are highly conserved across the world and primarily influenced by feeding patterns[17], and that microbial communities have shifted in a consistent manner with soil nitrogen and phosphorous additions[25]. If microbes were identified as important predictors of allergic disease, future interventions may focus on livestock feed or soil management to improve public health. In addition, our approach could be extended beyond the study of allergic disease to other immune-mediated conditions including autoimmune diseases, inflammatory diseases, and infectious disease.

REFERENCES

1. Asher MI, Montefort S, Bjorksten B, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet.* 2006;368(9537):733-743.
2. Global Burden of Disease Collaborators, Kyu HH, Pinho C, et al. Global and National Burden of Diseases and Injuries Among Children and Adolescents Between 1990 and 2013: Findings From the Global Burden of Disease 2013 Study. *JAMA Pediatr.* 2016;170(3):267-287.
3. von Mutius E, Vercelli D. Farm living: effects on childhood asthma and allergy. *Nat Rev Immunol.* 2010;10(12):861-868.
4. Schram ME, Tedja AM, Spijker R, Bos JD, Williams HC, Spuls PI. Is there a rural/urban gradient in the prevalence of eczema? A systematic review. *Br J Dermatol.* 2010;162(5):964-973.
5. von Mutius E. The microbial environment and its influence on asthma prevention in early life. *J Allergy Clin Immunol.* 2016;137(3):680-689.
6. Fujimura KE, Demoor T, Rauch M, et al. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. *Proc Natl Acad Sci U S A.* 2014;111(2):805-810.
7. Fyhrquist N, Ruokolainen L, Suomalainen A, et al. Acinetobacter species in the skin microbiota protect against allergic sensitization and inflammation. *J Allergy Clin Immunol.* 2014;134(6):1301-1309 e1311.
8. Schuijs MJ, Willart MA, Vergote K, et al. Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. *Science.* 2015;349(6252):1106-1110.
9. Ege MJ, Mayer M, Normand AC, et al. Exposure to environmental microorganisms and childhood asthma. *N Engl J Med.* 2011;364(8):701-709.
10. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. *Science.* 2016;352(6285):539-544.
11. Ladau J, Shi Y, Jing X, et al. Existing Climate Change Will Lead to Pronounced Shifts in the Diversity of Soil Prokaryotes. *mSystems.* 2018;3(5).
12. Amend AS, Seifert KA, Samson R, Bruns TD. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc Natl Acad Sci U S A.* 2010;107(31):13748-13753.
13. Vetrovsky T, Kohout P, Kopecky M, et al. A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nat Commun.* 2019;10(1):5142.
14. Bahram M, Hildebrand F, Forslund SK, et al. Structure and function of the global topsoil microbiome. *Nature.* 2018;560(7717):233-237.
15. van den Hoogen J, Geisen S, Routh D, et al. Soil nematode abundance and functional group composition at a global scale. *Nature.* 2019;572(7768):194-198.
16. Thompson LR, Sanders JG, McDonald D, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature.* 2017;551(7681):457-463.
17. Henderson G, Cox F, Ganesh S, et al. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep.* 2015;5:14567.
18. Yilmaz P, Kottmann R, Field D, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature biotechnology.* 2011;29(5):415-420.
19. Ellwood P, Asher MI, Beasley R, Clayton TO, Stewart AW, Committee IS. The international study of asthma and allergies in childhood (ISAAC): phase three rationale and methods. *Int J Tuberc Lung Dis.* 2005;9(1):10-16.
20. Harris I, Osborn TJ, Jones P, Lister D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data.* 2020;7(1):109.
21. Rodrigues F, Pereira FC. Beyond Expectation: Deep Joint Mean and Quantile Regression for Spatiotemporal Problems. *IEEE Trans Neural Netw Learn Syst.* 2020;31(12):5377-5389.
22. Franklin, J. *Mapping Species Dristributions. Spatial Inference and Prediction.* Cambridge University Press; 2010.
23. Barberan A, Ladau J, Leff JW, et al. Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci U S A.* 2015;112(18):5756-5761.

24. Cliff A, Romero J, Kainer D, Walker A, Furches A, Jacobson D. A High-Performance Computing Implementation of Iterative Random Forest for the Creation of Predictive Expression Networks. *Genes (Basel).* 2019;10(12).
25. Leff JW, Jones SE, Prober SM, et al. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc Natl Acad Sci U S A.* 2015;112(35):10967-10972.