Supporting information

Family-based study reveals decreased abundance of sputum Granulicatella in asthmatics

Leyao Wang[1,2,*], David de Ángel Solá[3], Yixin Mao[4,5], Piotr Bielecki[6], Yong Zhu[5], Zongke Sun[4], Liang Shan[7], Richard Flavell[6,8], Alia Bazzy-Asaad[3], Andrew DeWan[2]

Methods

Patient recruitment

Recruitment, questionnaire, sputum collection and 16S ribosomal RNA (rRNA) sequencing were approved by the Yale University Human Research Protection Program. The families with both asthmatic and non-asthmatic offspring were identified from an ongoing family-based asthma study being conducted at the Yale School of Public Health. Asthma was defined as a self-reported physician diagnosed asthma and two or more asthma symptoms in the last 12 months. We contacted and screened the members of 5 families. The questionnaire gathered information on symptoms and medications used for asthma or respiratory infection over the preceding 4 weeks. Exclusion criteria included: 1) presence of other respiratory diseases in the subject; 2) having had symptoms or treatment for a respiratory infection or asthma exacerbation in the past 4 weeks, with questions based on previously validated clinical tools (1); and 3) use of antibiotics, omalizumab or other unknown substances that could potentially impact results. Three families were eligible for this study. Figure E1 shows the pedigree of each family. Table E1 shows the characteristics of the 14 subjects. Atopy was defined by reported positive skin prick testing of aeroallergens (cat, dog, dust/dust mites, cockroach, pollens, trees, grass, mold/fungi, and ragweed).

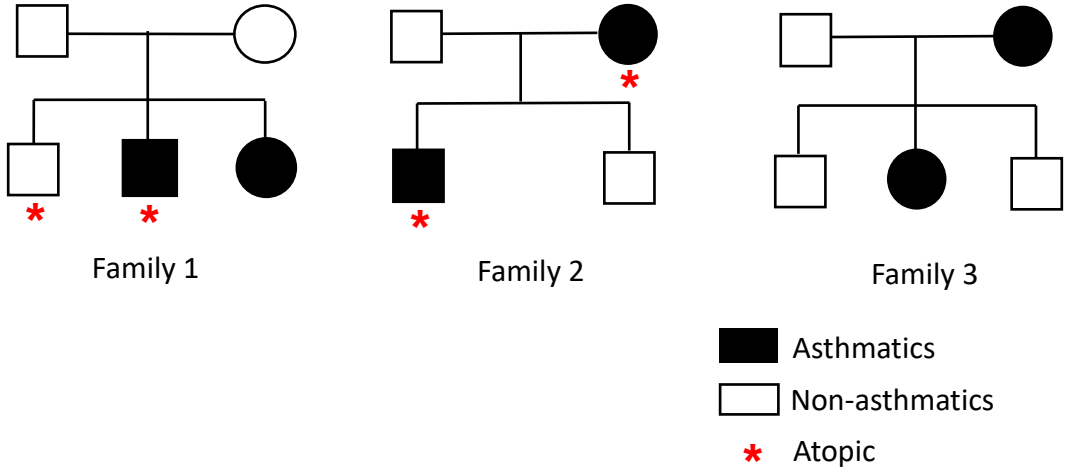


Figure E1. Pedigree drawings for each participating family. Each family has both asthmatic and non-asthmatic offspring.

Table E1. Study cohort characteristics

Variable	Asthmatics (n=6)	Non-asthmatics (n=8)	P value
Age, mean (range), y	28 (18-49)	36 (13-54)	
Sex, male, n	2	7	
White race, n	6	8	
BMI (kg/m^2)	31.0 ± 3.6	27.6 ± 2.2	0.42
Exposure to smoke, n	1	1	0.41
Atopic[1], n	3	1	0.17
FEV1 (% predicted) baseline	92 ± 6	95 ± 5	0.66
FEV1 (% predicted) minimum	82 ± 3	94 ± 5	0.09
FEV1 (% predicted) maximum	100 ± 4	101 ± 6	0.90

[1]Atopy was defined by reported positive skin prick testing of aeroallergens (cat, dog, dust/dust mites, cockroach, pollens, trees, grass, mold/fungi, and ragweed).

Statistical significance was determined by t-test for the continuous variables and Chi-squared with Yates correction for the categorical variables.

Sputum induction and DNA extraction

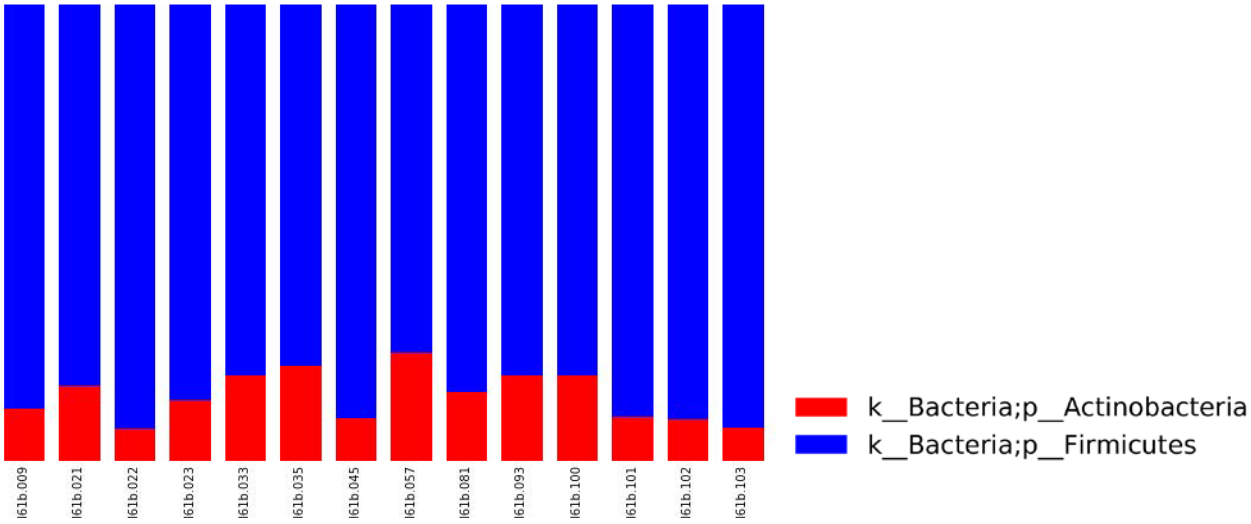
Enrolled subjects attended the Yale-New Haven Hospital Research Unit, where they signed informed consent/assent. Subjects were required to rinse their mouth with water to minimize oral contamination. Prior to beginning, a baseline FEV1 was measured; subject's respiratory status was also monitored by a physician during the procedure. If the FEV1 was ≥ 80% of predicted, procedure was started by administering nebulized 7% hypertonic saline for increasing time periods (minimum 30 s, maximum 4 min) with FEV1 measured each time based on previously described procedures (2). Predicted values for FEV1 were based on the variables and set values from NHANES III (3, 4).

Subjects were asked to expectorate sputum into storage reagent prototype provided by DNA Genotek (Ottawa, ON, Canada). The reagent keeps bacteria from phylum Firmicutes and Actinobacteria intact, but lyses most gram-negative bacteria from phylum Proteobacteria, Bacteroidetes and Fusobacteria. QIAamp DNA microbiome kit (Qiagen, Hilden, Germany) was used to extract bacterial DNA. This kit first depletes all released free DNA (including the genomic material from human cells and broken bacteria) and then applies a combination of chemical and mechanical lysis to break and extract DNA from Firmicutes and Actinobacteria.

16S rRNA sequencing and analysis

The 16S rRNA coding gene fragments were sequenced as previously described (5). Briefly the V4 region of 16S ribosomal RNA coding gene was PCR amplified (34 cycles; primer pair F515/R806) with barcodes (Phusion polymerase, New England Bioscience). After amplification, PCR products were purified (SequalPrep™, Invitrogen, Waltham, MA), resuspended in 20 µl of water and pooled. The concentration of sequencing library was assessed with qRT-PCR and final concentration of 11 pM was loaded on a MiSeq sequencer (Illumina, San Diego, CA), with sequencing conditions of 2x250 bp paired-end reads. Downstream analysis of raw data, paired end reads assembly, and UniFrac distance was analyzed with the Quantitative Insights Into Microbial Ecology (QIIME version 1.7) analysis suite (6). Reads were demultiplexed and quality filtered with a Q-score cutoff of 29. The open-reference OTU picking workflow in QIIME and the Greengenes reference database were used to cluster the reads into 97% identity Operational Taxonomic Units (OTUs). The Ribosomal Database Project classifier (RDP) and the May 2013 Greengenes taxonomy were used to assign taxonomy to representative OTUs. OTUs of less than 0.01% relative abundance, and contaminating OTUs that were also found after sequencing of 16S amplicons from PCR samples without template DNA, were filtered from OTU tables. Filtered OTU tables were rarefied to a depth of 3,000 sequences per sample for all analyses. Only the sequencing reads from Phylum Firmicutes or Actinobacteria (>98% of the total reads) were preserved for analysis. Figure E2 shows the bacterial profiling at Phylum and Genus level respectively.

A.



B.

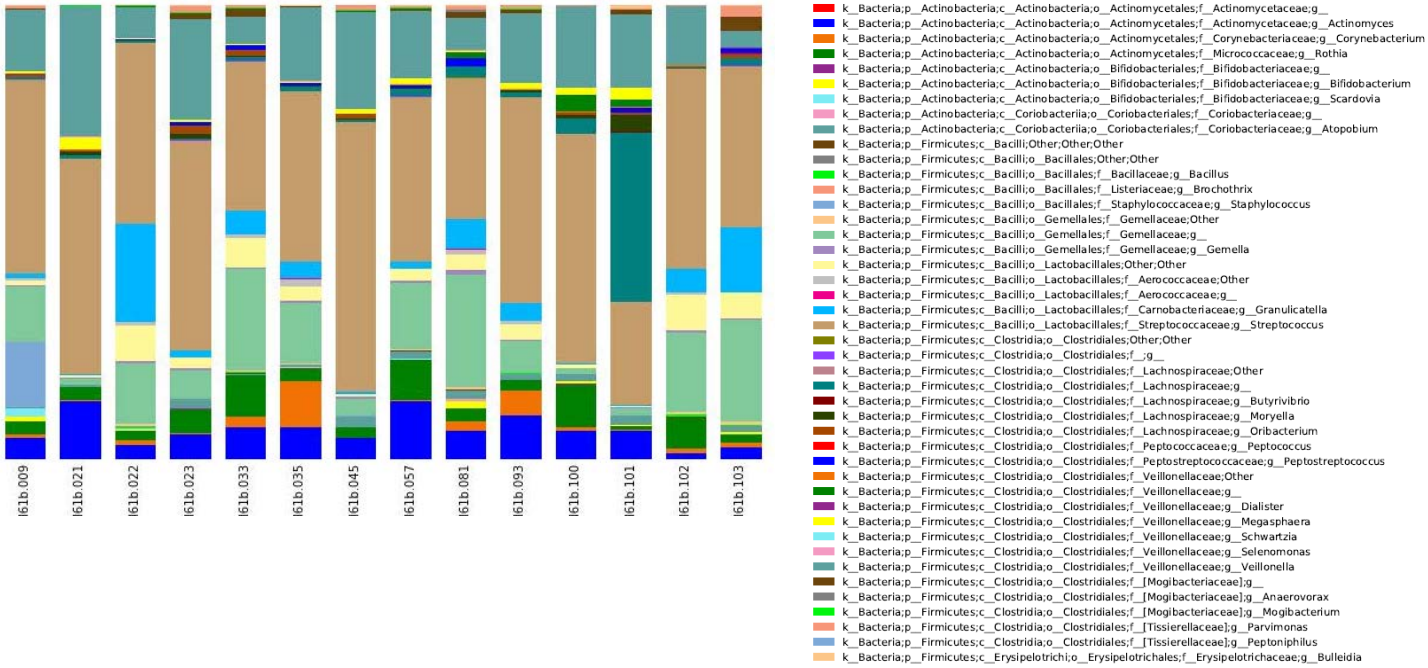


Figure E2. Bacterial profiling by 16S shown at Phylum (A) and Genus (B) taxonomic level. Each bar represents the overall sputum bacterial community in each subject.

The alpha diversity (Shannon's index) Analysis was performed in the statistical software package "ape" in R v3.4.3 (7). Bacterial genera with mean relative abundance over 1% among offspring were included in an analysis to identify if any of them displayed differences in relative

abundance between the asthmatic and non-asthmatic sibling(s). The relative abundance of each genus in case and control of each family is shown in Table E2. Red color indicates the specific genus is depleted in case compared to control within the specific family. Green color indicates the specific genus is enriched in case compared to control within the specific family.

Table E2. Relative abundance of bacterial genera

Genus	Family 1		Family 2		Family 3		P-value
	Case	Control	Case	Control	Case	Control	
Granulicatella	0.010	0.053	0.052	0.143	0.036	0.128	0.043
Veillonella	0.215	0.061	0.130	0.036	0.161	0.109	0.078
Streptococcus	0.528	0.328	0.440	0.353	0.374	0.425	0.392
Actinomyces	0.050	0.070	0.012	0.025	0.070	0.062	0.426
Atopobium	0.021	0.005	0.004	0.014	0.006	0.008	0.457
Gemella	0.018	0.066	0.082	0.058	0.037	0.059	0.542
Corynebacterium	0.002	0.022	0.011	0.011	0.100	0.032	0.609
Rothia	0.036	0.092	0.066	0.019	0.030	0.023	0.984

Relative abundance of the top abundant bacterial genera between discordant sibling(s) within each family. Mean value was used if a family has more than one case or control sibling. P-value was calculated using paired t-test.

Supporting References:

1. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol. 2004;113(1):59-65.
2. Gibson PG, Wlodarczyk JW, Hensley MJ, Gleeson M, Henry RL, Cripps AW, et al. Epidemiological association of airway inflammation with asthma symptoms and airway hyperresponsiveness in childhood. Am J Respir Crit Care Med. 1998;158(1):36-41.
3. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med. 1999;159(1):179-87.
4. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948-68.
5. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013;79(17):5112-20.
6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.
7. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20(2):289-90.