

DR. CHIH-YUNG CHIU (Orcid ID : 0000-0002-6454-968X)

DR. WEN-HUNG CHUNG (Orcid ID : 0000-0002-1929-4008)

DR. SHIH-CHI SU (Orcid ID : 0000-0002-1107-9805)

Article type : Original Article: Asthma and Lower Airway Disease

Integration of metagenomics-metabolomics reveals specific signatures and functions of airway microbiota in mite-sensitized childhood asthma

Running title: Airway microbiome and host metabolome of childhood asthma

Chih-Yung Chiu^{1,2}, Hsin-Cheng Chou³, Lun-Ching Chang⁴, Wen-Lang Fan⁵, Michael Cong Vinh Dinh⁶, Yu-Lun Kuo⁷, Wen-Hung Chung⁸, Hsin-Chih Lai^{9,10}, Wen-Ping Hsieh³, Shih-Chi Su^{8,#}

¹Division of Pediatric Pulmonology, Chang Gung Memorial Hospital at Linkou, College of Medicine, Chang Gung University, Taoyuan, Taiwan

²Clinical Metabolomics Core Laboratory, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

³Institute of Statistics, National Tsing-Hua University, Hsinchu, Taiwan

⁴Department of Mathematical Sciences, Florida Atlantic University, Florida, US

⁵Genomic Medicine Core Laboratory, Chang Gung Memorial Hospital, Linkou, Taiwan

⁶Department of Computer Science, Florida Atlantic University, Florida, US

⁷Biotoools, Co., Ltd, New Taipei City, Taiwan

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ALL.14438](https://doi.org/10.1111/ALL.14438)

This article is protected by copyright. All rights reserved

⁸Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan

⁹Department of Medical Biotechnology and Laboratory Science, and Microbiota Research Center, College of Medicine, Chang Gung University, Taoyuan, Taiwan

¹⁰Central Research Laboratory, XiaMen Chang Gung Hospital, XiaMen, China.

#Corresponding author: Shih-Chi Su, PhD

Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan.

222, Mai-Chin Road, Keelung 20401, Taiwan; Phone: 886-2-24329292-3388; Fax: +886-2-27191623; Email: ssu1@cgmh.org.tw

Word count: 3439

Abstract

Background: Childhood asthma is a multifactorial inflammatory condition of the airways, associated with specific changes in respiratory microbiome and circulating metabolome. **Methods:** To explore the functional capacity of asthmatic microbiome and its intricate connection with the host, we performed shotgun sequencing of airway microbiome and untargeted metabolomics profiling of serum samples in a cohort of children with mite-sensitized asthma and non-asthmatic controls. **Results:** We observed higher gene counts and sample-to-sample dissimilarities in asthmatic microbiomes, indicating a more heterogeneous community structure and functionality among the cases than in controls. Moreover, we identified airway microbial species linked to changes in circulating metabolites and IgE responses of the host, including a positive correlation between *Prevotella sp. oral taxon 306* and dimethylglycine that were both decreased in patients. Several control-enriched species (*Eubacterium sulci*, *Prevotella pallens*, and *Prevotella sp. oral taxon 306*) were inversely correlated with total and allergen-specific IgE levels. Genes related to microbial

carbohydrate, amino acid, and lipid metabolism were differentially enriched, suggesting that changes in microbial metabolism may contribute to respiratory health in asthmatics. Pathway modules relevant to allergic responses were differentially abundant in asthmatic microbiome, such as enrichments for biofilm formation by *Pseudomonas aeruginosa*, membrane trafficking, histidine metabolism, and glycosaminoglycan degradation, and depletions for polycyclic aromatic hydrocarbon degradation. Further, we identified metagenomic and metabolomic markers (e.g. *Eubacterium sulci*) to discriminate cases from the non-asthmatic controls. **Conclusions:** Our dual-omics data reveal the connections between respiratory microbes and circulating metabolites perturbed in mite-sensitized pediatric asthma, which may be of etiological and diagnostic implications.

Keyword: Airway microbiome, circulating metabolites, pediatric asthma

Introduction

Asthma, with a sharp increase in the global incidence, morbidity, and mortality, remains a huge healthcare burden over the past decades, particularly in children ¹. Numerous risk factors, including gender, genetic inheritance, allergen exposure, infections, tobacco smoking, obesity, and perinatal factors, have been demonstrated to be interrelated and imperative to estimate the etiology and prognosis of childhood asthma ²⁻⁵. Among these, sensitization to food and inhalant allergens is recognized as a significant environmental determinant in the development of atopic diseases for young children ². However, such complex pathophysiology augments the clinical heterogeneity of the condition and introduces certain challenges for better prevention and treatment of asthma in early life.

As a chronic inflammatory disease of the airways, asthma is characterized by a wide variety of cellular and molecular mediators resulting in a broad spectrum of possible metabolic products ⁶. Even though results for individual metabolites were not fully concordant across metabolomics studies of asthma, there was consistency in the enriched pathways, such as amino acid metabolism, lipid metabolism, inflammation, and immune response ⁷, revealing the clinical utility of metabolic profiles from serum or other samples to characterize asthma. Other than host metabolism and immunity, an intricate interaction between the human microbiota and the host is central to the development and progression of asthma ⁸. In addition to intestinal microflora that shaped asthma risk, pathogenesis, and clinical presentation ^{9,10}, emerging evidence has related atopy and asthma to the microbial composition and function of the airway ^{11,12}, the most relevant site of interactions between respiratory epithelia and the environment. It has been reported that airway bacteria, such as members of the *Haemophilus*, *Moraxella*, and *Neisseria* genera, were consistently enriched in severe or childhood asthma ¹³⁻¹⁸. Recently, our group has demonstrated an inverse correlation between airway microbial diversity and mite-sensitized asthma in early childhood ¹⁹. Together, these findings suggest that circulating metabolome, linked to mite-sensitized pediatric asthma, may be in part microbially derived or modified.

Although 16S rRNA amplicon sequencing has made important strides in clinical microbiology, deep shotgun metagenome sequencing often provides more in-depth taxonomic characterization^{20,21} and functional insights into the human microbiomes²² than 16S rRNA gene sequencing. In this study, we took an unbiased approach to link information on the airway microbiome with serum metabolite data by performing shotgun metagenome sequencing of throat swabs and untargeted nuclear magnetic resonance (NMR) spectroscopy metabolomic profiling of serum samples in a cohort of children with mite-sensitized asthma and non-asthmatic controls. Our integrative analyses identified several dysbiotic genes/pathways and species-level clades in robust associations with circulating metabolites and IgE responses to dust mites, which may be indicative of biological mechanisms underlying pediatric asthma related to mite sensitization.

Materials and Methods

Study cohorts and sample collection

In this cross-sectional study, children (ages: 3–5 years) diagnosed with mite-sensitized asthma for the first time at outpatient clinics and healthy controls without mite sensitization were enrolled in the department of pediatrics, Chang Gung Memorial Hospital, with the approval by the institutional review board. Subjects sharing the same ancestry (largely Han Chinese), culture, religion and language were accrued from 2013 to 2017, and written informed consent was obtained from the parents or guardians of all the study participants. Asthma was diagnosed as ever having asthma, with the occurrence of recurrent wheeze in the last 12 months, or current use of asthma medication, based on the guidelines of the Global Initiative for Asthma²³. Age-matched children without a history of asthma or other atopic conditions or infections were enrolled as the control group. Detailed information on potential confounding variables for atopic diseases including child's gender, age, maternal atopy, passive smoking, and older siblings at birth was collected. Sample (whole blood and throat swab) collection in detail can be found in supplemental methods.

Shotgun metagenome sequencing and data analysis

DNA from the throat swab was extracted by using the FastDNA SPIN Kit for Soil (MP Biomedical) and fragmented by using a Covaris M220 sonicator (Covaris, Inc.). The library construction was carried out using a TruSeq DNA sample preparation kit following a standard protocol from Illumina, and 10 Gb of 150 bp pair-end (PE) reads per sample were sequenced on an Illumina HiSeq instrument initially for determining the proportion of microbial sequences to the total reads. For samples containing insufficient microbial reads, multiple rounds of sequencing were performed. Sequence data were processed using an assembly-based pipeline, in which quality control checks, filtering, metagenome assembly (by MEGAHIT)²⁴, contig binning (by MetaBAT)²⁵, gene prediction (by MetaGeneMark)²⁶ and functional annotation of gene clusters and comparison with the Kyoto Encyclopedia of Genes and Genomes (KEGG) protein database (by GhostKOALA)²⁷ were performed. Detailed methods regarding metagenomic data processing and analysis can be found in supplemental methods.

¹H-Nuclear Magnetic Resonance (NMR) spectroscopy and data processing

Serum samples were mixed with phosphate buffer which contains 0.1% 3-(trimethylsilyl)-propionic-2,2,3,3-d₄ acid sodium salt (TSP) as an internal chemical shift reference standard prior to spectrum acquisition. ¹H-NMR spectra were acquired at 300K on a Bruker Avance-600 MHz spectroscopy (Bruker Daltonics Inc., Fremont, CA) equipped with a 5 mm CPTCI ¹H probe. All 1D spectra were applied for analysis before Fourier transformation with zero-filled to exponential line-broadenings of 0.3 Hz. ¹H-NMR spectra acquired were manually phased, baseline-corrected, and referenced to the chemical shift of TSP ($\delta=0.0$ ppm) using TopSpin 3.2 software (Bruker BioSpin, Rheinstetten, Germany), followed by spectral bucket, spectral region exclusion and spectral normalization using AMIX version 3.9.12 (Bruker BioSpin, Rheinstetten, Germany). Serum metabolites were identified using Chenomx NMR Suite 8.2 professional software (Chenomx Inc., Edmonton AB, Canada).

Statistical analysis

Differences in the baseline characteristics between the cases and controls were determined using analysis of variance (ANOVA) or student's t test. To evaluate gene richness in the control and asthmatic samples, rarefaction curves were generated as previously described²⁸. Briefly, for a given

number of samples, we performed a randomized sampling 28 and 27 times in the control samples and asthmatic samples, respectively, and estimated the total number of genes that could be identified from these samples. α diversity (within-sample) estimated by the Shannon index based on the taxonomic profile of each sample and β diversity (between-sample) assessed by Bray-Curtis distance were compared by using Wilcoxon rank sum test. Abundance of metabolites was normalized using a log transformation and Pareto scaling. The connection of representative microbes to asthma-related metabolites and clinical indices was assessed by Spearman's rank correlation, and the importance was corrected by using the Benjamini-Hochberg procedure. Pearson correlation was used to perform the large-scale association discovery of bacterial species with serum metabolites and IgE levels, and p values were adjusted by using the Benjamini-Hochberg correction for multiple tests. To possibly rule out the effect largely derived from outliers, species with $<0.001\%$ abundance in $>80\%$ of samples were excluded in the large-scale discovery. Data were analyzed using SPSS 22.0 for Windows XP (SPSS Inc., Chicago, IL). Statistically significant biomarkers (bacterial species or serum metabolites) were evaluated by the linear discriminant analysis (LDA) of effect size (LEfSe) analysis, which employed the non-parametric factorial Kruskal-Wallis test, Wilcoxon rank sum test and LDA to identify differentially abundant biomarkers between two metadata classes ²⁹. Differentially enriched KEGG modules were identified according to their reporter score ³⁰, calculated from the Z-scores of individual KO groups. A module with reporter score $Z > 1.88$ ($> 90\%$ confidence according to a normal distribution) was considered as a significant dysbiosis module. With a particular focus on those immune-related pathways and amino acid metabolism, abundance of KO genes of seven selected dysbiosis modules, including biofilm formation by *Pseudomonas aeruginosa*, membrane trafficking, glycosaminoglycan (GAG) degradation, polycyclic aromatic hydrocarbon (PAH) degradation, polyketide sugar unit biosynthesis, histidine metabolism, and valine, leucine, isoleucine degradation, were tested for association with metabolomic and clinical parameters by Pearson correlation adjusted using the Benjamini-Hochberg procedure. To determine a panel of airway bacterial and circulating metabolic markers for detecting mite-sensitized childhood asthma, two steps were conducted. In the 1st step, Random Forests ³¹ models were used to rank each type of profiles independently and validated by 10-fold stratified cross-validation testing (500 trees, balanced class weight, max features=square root of all features). The optimal number of discriminatory markers for

each type of profiles was determined using the recursive feature elimination method ³² with default parameters using five different random seeds. For each test, the performance of the model was examined by receiver operating characteristic (ROC) curves constructed by pROC ³³ using five-fold cross-validation. In the 2nd step, discriminatory makers of two profiles identified in the previous step were pooled. Ranking, determination of the optimal number of features, and testing for the performance of the combination of respiratory bacterial and serum metabolic markers were conducted in the same method. Unless otherwise stated, all statistical tests are two-tailed, and a $p < 0.05$ is considered statistically significant.

Results

Study design and subject characteristics

To explore the association of airway microbes with host circulating metabolites in mite-sensitized childhood asthma, we collected oropharyngeal swabs from a cross-sectional cohort of 69 subjects, comprising 32 cases and 37 non-asthmatic controls for shotgun metagenome sequencing (**Figure S1**). Among them, serum samples of 61 participants were subjected to NMR spectroscopy to obtain non-targeted metabolomic profiles. Since obtaining sufficient microbial sequencing data is challenging for non-fecal samples but also a crucial requirement for metagenomics studies, followed by stringent QC and filtering, only 55 samples with at least 4 Gb of microbial reads (**Figure S2**) were used for profiling of microbial taxonomic composition and functional potential as well as further integrative analyses. Comparisons of the demographic and clinical data associated with these 55 samples (27 cases, 28 controls) reveal that atopic indices including total IgE levels and allergen sensitization to house dust mites and egg white were higher in children with asthma than in controls. (**Table 1**). There was no difference in the distribution of sex, parental atopy, passive smoking, older siblings and household income between two groups.

Metagenomic and metabolomic signatures of mite-sensitized pediatric asthma

We first compared airway microbial diversity between the case and control group. Although no difference in within-sample species richness (α diversity) was detected between two groups, we observed higher gene counts and sample-to-sample dissimilarities (β diversity) in asthmatic

microbiomes (**Figure 1A-C**), indicating a more heterogeneous community structure and functionality among the cases than in controls. To dissect the detailed taxonomic and metabolomic features, a total of 129 species-level clades in 55 swab samples and 44 serum metabolites in 61 serum samples were profiled. An overview of asthma-specific changes in individuals' microbial and metabolic profiles determined by taking statistical significance and biological consistency into consideration using LEfSe is given in **Figure 1D-E**. Despite variations across individuals, we observed a predominance of *Neisseria elongata* in the upper airway of the asthma group over the controls. On the contrary, samples in the non-asthmatic controls were significantly enriched for *Eubacterium sulci*, *Leptotrichia wadei*, and several *Prevotella* spp. (*P. pallens*, *P. salivae*, and *P. sp. oral taxon 306*) as compared with those in the asthma group. In addition, proline-glutamate chimera and serine were among circulating metabolites exhibiting significant enrichments in the asthma group, while dimethylamine and dimethylglycine were increased in the controls.

Associations of airway microbial species with circulating metabolites and clinical indices

Integrated molecular profiles of host and gut microbial activities have improved understanding of perturbations of the microbiome-metabolome interface in human diseases^{34,35}. To examine whether specific respiratory microbes are responsible for a mechanistic relationship relating the abundance of microorganism and metabolite that is perturbed during the development of pediatric asthma, the link of six differentially abundant species to four asthma-related metabolites aforementioned and clinical indices was examined. We identified a positive correlation between *Prevotella sp. oral taxon 306* and dimethylglycine (DMG) ($p = 0.006$) (**Figure 2A**). Several control-enriched species (*E. sulci*, *P. pallens*, and *P. sp. oral taxon 306*) were inversely correlated with total and allergen-specific IgE levels (**Figure 2B**). These results reveal a microbe-metabolite connection in mite-sensitized childhood asthma. To comprehensively explore such connection, a large-scale association discovery of 129 species with metabolites and clinical indices was further assessed. We identified a set of species that is positively correlated with the levels of numerous essential amino acids (such as branched-chain amino acids and histidine) and energy-related carboxylic acids (such as citrate, lactate, and pyruvate) but inversely associated with that of glucose and pyridoxine (**Figure 3**). Notably, another set of species was found to be anti-correlated with fumarate. In addition to circulating metabolites, several

species were significantly correlated with IgE responses to allergens. Among them, *Streptococcus intermedius* was in strong associations with total and mite allergen-specific IgE levels in children with mite-sensitized asthma.

Functional characterization of airway microbiome in mite-sensitized pediatric asthma

We next investigated the functional capacities of asthmatic microbiome. By mapping our bacterial gene catalog onto the KEGG database, we obtained 4,761 KEGG orthologs (KOs) present in at least five samples for subsequent analyses. A total of 34 differentially enriched KEGG modules (18 asthma-enriched and 16 control-enriched) were identified (**Figure 4A**), of which many pathway modules are related to microbial carbohydrate, amino acid, and lipid metabolism, suggesting that changes in microbial metabolism may contribute to respiratory health in children with mite-sensitized asthma. Moreover, of particular relevance in orchestrating inflammatory and allergic responses, genes involved in biofilm formation by *Pseudomonas aeruginosa*, membrane trafficking, and glycosaminoglycan (GAG) degradation were enriched in the asthmatic microbiome, while those associated with polycyclic aromatic hydrocarbon (PAH) degradation and polyketide sugar unit biosynthesis were more abundant in the controls.

To further explore the role of these functional dysbioses, with a focus on those immune-related pathways and amino acid metabolism, we tested whether relative abundance of genes in these KEGG modules are mutually associated with metabolomic and clinical parameters. We observed that many genes of these dysbiosis modules were positively associated with serum levels of proline-glutamate chimera, the most significantly enriched metabolite in the asthma group and many amino acids, but anti-correlated with that of glucose and pyridoxine (**Figure 4B**). Abundance of one gene (hisN, encoding a histidinol-phosphatase, K05602) involved in histidine biosynthesis was correlated with serum levels of histidine. In addition, one gene (flgM, K02398) responsible for biofilm formation by *P. aeruginosa* (asthma-enriched) and another (pcaG, K00448) for PAH degradation (control-enriched) were found to be correlated with allergen-specific and total IgE levels. Further analyses of the predominant species that contribute to these dysbiosis modules (**Figure S4**) revealed that altered membrane trafficking, GAG degradation, and histidine metabolism were in part attributed by an

increase in the abundance of *Neisseria lactamica*, *Neisseria elongata*, and *Moraxella catarrhalis* in asthmatic microbiome (**Figure 4C**). These findings highlight an involvement of airway dysbiosis and its associated functional shift in regulating inflammatory and allergic reactions.

Dual-omics signatures differentiate childhood asthma from non-asthmatic controls

In order to test the potential of airway metagenomic and serum metabolomic parameters to serve as diagnostic markers, we constructed Random Forests regression models based on species only, metabolite only (**Figure S5**), or a combination of both (**Figure 5A**) to discriminate cases from non-asthmatic controls. Comparison of the classification potential among the three models demonstrated that the total area under the ROC curve (AUC) increased from 0.712 and 0.721 for the species and metabolite model, respectively, to 0.804 for the combination, revealing that dual-omics signatures markedly improves the discriminative ability (**Figure 5B**). Notably, since *Eubacterium sulci* was identified as the top discriminatory species by both Random Forests and LEfSe methods, its clinical validity was evaluated. In addition to an AUC of 0.73 for detecting a toddler with asthma, *E. sulci* also exhibited ideal discernibility between the non-allergic controls and children allergic to house dust mites (**Table S1**), with an AUC of 0.74 and 0.70 for distinguishing tolerant controls from children sensitized to *D. pteronyssinus* and *D. farinae*, respectively. Our results reveal promising avenues for diagnosis of mite-sensitized pediatric asthma via specific airway microbes.

Discussion

Cumulative evidence has related the microbial composition and function of the airway to atopy and asthma^{8,11,12}. Thus, in-depth taxonomic and functional characterization of asthmatic microbiome is highly warranted. Here, we identified specific airway dysbiosis at the species level and its associated functional shift in strong associations with circulating metabolites and IgE responses to dust mites. These connections highlight a potential microbe-host interaction in the development of childhood asthma related to mite sensitization.

In the composition of asthma-associated airway microbiota, we observed significant depletions for *Eubacterium sulci*, *Leptotrichia wadei*, and several *Prevotella* spp. in the oropharynx of children with

asthma, compared with that of non-asthmatic controls. Among them, *Prevotella* spp. have been reported to be decreased in asthmatic airways¹⁴, as *Eubacterium sulci* was for the first time identified as an asthma-associated species with possible prognostic importance. The reduction in these *Prevotella* spp. and *Eubacterium sulci* was inversely correlated with total and allergen-specific IgE levels. Compared with pathogenic proteobacteria whose outgrowth has been seen in asthmatic airways¹⁴, these asthma-depleted species had lower capacities to trigger inflammatory cytokines in immune and stromal cells of the lung^{36,37}. Such difference could be ascribed to alternate lipopolysaccharide (LPS) structures as proteobacteria contains specific enzymes to produce LPS with hexa-acylated lipid A moiety whereas *Prevotella* produce penta-acylated LPS, respectively³⁸. Hexa-acylated LPS (6 acyl) was demonstrated to be 100-fold more immunostimulatory when bound to toll-like receptor 4 than LPS that is penta-acylated (5 acyl)³⁹. Yet, a decrease in the abundance of *Leptotrichia wadei* has been detected in the upper airway while severe asthma was improved by the treatment of azithromycin⁴⁰. On the contrary, *Neisseria elongata* was significantly increased in our asthma cohort. Numerous studies have reported an increase of *Neisseria* spp. in childhood or severe asthma¹³⁻¹⁶. We also found that *Neisseria elongata* and *Neisseria lactamica* contributed to many altered microbial functional pathways that are enriched in the asthmatic airway (**Figure 4C**). In general, our findings are consistent with that from others but of higher resolution than those from 16S rRNA-based investigations.

Moreover, we identified airway microbial species linked to fluctuations in serum metabolites of the host. Among these associations, the abundance of *Prevotella sp. oral taxon 306*, one of significantly depleted species in children with asthma, was positively correlated with the serum levels of DMG. DMG, known to increase tissue oxygen uptake and decrease lactate accumulation, was significantly reduced in urine metabolome of asthma exacerbation⁴¹ and in the circulation of our asthma cohort, suggesting a hypoxic scenario linked to mite-sensitized pediatric asthma. A well-recognized hypoxia-induced change was documented in asthma⁴², with significantly lower levels of glucose and elevated levels of lactate in the serum metabolome⁴³. Notably, we identified a set of anaerobic microbes, whose abundance was positively associated with the levels of lactate but anti-correlated with that of glucose. Although many of these associations involve metabolites and microorganisms that were not

significantly altered in our analysis likely due to the substantial heterogeneity and limited statistical power, such connections may reflect the mechanistic linkage within the microenvironment of asthmatic airways, instead of a direct trigger in the pathogenesis of asthma. In addition, the abundance of many microbial species and genes was found to be negatively associated with serum levels of pyridoxine. Pyridoxine, also known as vitamin B6, serves as a regulator of host immunity and a cofactor for a variety of cellular metabolic processes, including amino acid, lipid, and carbohydrate metabolism ⁴⁴. Lack of vitamin B6 interfered the Th1-Th2 balance, leading to an excessive Th2 response ⁴⁵. Impaired vitamin B6 metabolism and depressed plasma vitamin B6 have previously been observed in asthmatics ⁴⁶, but the reason remains unidentified. Our results may implicate additional role of airway dysbiosis in modulating vitamin-B6-dependent metabolic and immunological responses.

The functional capacities of airway microbiome in asthma are largely unexplored since obtaining sufficient bacterial reads is especially challenging for shotgun metagenome sequencing of non-fecal samples. Here, we found that microbial genes associated with PAH degradation were significantly enriched in the controls. Exposure to a variety of PAHs has been associated with the development of asthma in young children ⁴⁷. Our finding indicates poor turnover rates for PAHs mediated by respiratory microbes in asthmatic airways. In addition, bacterial genes related to degradation of GAGs, which have been associated with bronchial hyperresponsiveness, tissue inflammation and airway remodeling in pulmonary diseases ⁴⁸, were more abundant in the asthma group than in controls. Also enriched in the asthma group is the proportion of genes associated with membrane trafficking, which represents a bidirectional movement of macromolecules (e.g. lipopolysaccharides) in or out of the microbes. For Gram-negative bacteria, membrane trafficking is mediated by bacterial outer membrane vesicles and plays a vital role in interbacterial and host-microbe interaction, endotoxic shock, and activation of innate immunity ⁴⁹. Moreover, these membrane vesicles were proposed to promote bacterial adhesion to host epithelial cell surface ⁵⁰. Intriguingly, the abundance for a gene involved in biofilm formation by *Pseudomonas aeruginosa* (a case-enriched module) was positively correlated with mite-specific IgE levels in our investigation. These point out a pathogenic role for the virulence factors of Gram-negative microbes in asthma. Of note, another asthma-enriched function identified is histidine metabolism. Histidine is the precursor of histamine, which is increasingly produced by

specific gut microbiota of asthma patients ^{51,52} and remains a prominent contributor for broncospasm, edema, inflammation, and mucus secretion in asthmatics ⁵³. Collectively, our findings in this study recapitulate a functional involvement of airway dysbiosis in orchestrating inflammatory and allergic reactions in mite-sensitized childhood asthma.

To define the functionality of airway microbiome and its association with the host in asthmatics, additional work is needed to address several limitations of the present study. One drawback is that a replication cohort or samples collected at different time points in a longitudinal design are lacking for validation although many lines of our metagenomic and metabolomic results are consistent with previous findings from others. Since samples were collected while the cases were diagnosed with mite-sensitized asthma for the first time at outpatient clinics, data concerning asthma exacerbations and therapies were not obtainable. Also unavailable is information regarding the body mass index and diet, as these factors may perturb airway microbiome ^{54,55} and circulating metabolome ^{56,57} in asthma. Another issue is that in spite of a global and unbiased view of the samples, our untargeted metabolomics approach had low analytical sensitivity to metabolites with low abundance (< 100 nmol/L, e.g. histamine) and failed to conduct the absolute quantification of the samples. This may hinder the detection of their association with other variables. In addition, the details of many associations identified remain to be interpreted and determined due to limited knowledge from relevant studies to date. Ideally, validation of these microbe-metabolite interactions and their role in asthma pathogenesis can be addressed through further *in vivo* experiments. Such attempts are often laborious; thus, computational exploration of putative links detected in the study by integration of dual-omics profiling could be an important aid.

In conclusion, our findings establish a comprehensive understanding in the relationship between the airway microbiota at the species level and host metabolism and IgE responses, providing potential avenues for microbiome-based diagnosis and modality of pediatric asthma.

Acknowledgement

The authors are grateful to the Tissue Bank at Chang Gung Memorial Hospital, Keelung, Taiwan for preparing clinical specimens. We thank Mr. Rhian Resnick for his aid with the use of high performance computing (HPC) service at Florida Atlantic University, Florida, US. This study was supported by research grants from the Chang Gung Memorial Hospital (BMRPE97) and from the Ministry of Science and Technology, Taiwan (MOST-107-2321-B-182A-001).

Abbreviations

NMR: nuclear magnetic resonance; ORF: open reading frame; KEGG: Kyoto Encyclopedia of Genes and Genomes; KO: KEGG orthologue; LEfSe: linear discriminant analysis of effect size; ROC: receiver operating characteristic; DMG: dimethylglycine; GAG: glycosaminoglycan; PAH: polycyclic aromatic hydrocarbon; AUC: area under the ROC curve

Author contributions

C-Y.C., W-P.H., and S-C.S. performed conceptualization. C-Y.C., and W-H.C. prepared sampling. X-C.Z., L-C.C., W-L.F., M.C.V.D. and Y-L.K. performed data analysis. X-C.Z., L-C.C., and Y-L.K. performed data curation. C-Y.C. and S-C.S. performed writing. H-C.L., W-P.H., and S-C.S. performed supervision.

Conflict of interest

All the authors declared no competing interests.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Reference

1. Braman SS. The global burden of asthma. *Chest*. 2006;130(1 Suppl):4S-12S.
2. Chiu CY, Huang YL, Tsai MH, et al. Sensitization to food and inhalant allergens in relation to atopic diseases in early childhood: a birth cohort study. *PloS one*. 2014;9(7):e102809.

- Accepted Article
3. Ehrlich RI, Du Toit D, Jordaan E, et al. Risk factors for childhood asthma and wheezing. Importance of maternal and household smoking. *American journal of respiratory and critical care medicine*. 1996;154(3 Pt 1):681-688.
 4. Herzog R, Cunningham-Rundles S. Pediatric asthma: natural history, assessment, and treatment. *Mt Sinai J Med*. 2011;78(5):645-660.
 5. Cookson W. The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat Rev Immunol*. 2004;4(12):978-988.
 6. Ishmael FT. The inflammatory response in the pathogenesis of asthma. *J Am Osteopath Assoc*. 2011;111(11 Suppl 7):S11-17.
 7. Kelly RS, Dahlin A, McGeachie MJ, et al. Asthma Metabolomics and the Potential for Integrative Omics in Research and the Clinic. *Chest*. 2017;151(2):262-277.
 8. Abdel-Aziz MI, Vijverberg SJH, Neerincx AH, Kraneveld AD, Maitland-van der Zee AH. The crosstalk between microbiome and asthma: Exploring associations and challenges. *Clin Exp Allergy*. 2019.
 9. Stokholm J, Blaser MJ, Thorsen J, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun*. 2018;9(1):141.
 10. Kozik AJ, Huang YJ. The microbiome in asthma: Role in pathogenesis, phenotype, and response to treatment. *Ann Allergy Asthma Immunol*. 2019;122(3):270-275.
 11. Huang YJ. The respiratory microbiome and innate immunity in asthma. *Current opinion in pulmonary medicine*. 2015;21(1):27-32.
 12. Chung KF. Airway microbial dysbiosis in asthmatic patients: A target for prevention and treatment? *The Journal of allergy and clinical immunology*. 2017;139(4):1071-1081.
 13. Huang YJ, Nariya S, Harris JM, et al. The airway microbiome in patients with severe asthma: Associations with disease features and severity. *The Journal of allergy and clinical immunology*. 2015;136(4):874-884.
 14. Hilty M, Burke C, Pedro H, et al. Disordered microbial communities in asthmatic airways. *PLoS one*. 2010;5(1):e8578.
 15. Goleva E, Jackson LP, Harris JK, et al. The effects of airway microbiome on corticosteroid responsiveness in asthma. *Am J Respir Crit Care Med*. 2013;188(10):1193-1201.
 16. Green BJ, Wiriyaichaiyorn S, Grainge C, et al. Potentially pathogenic airway bacteria and neutrophilic inflammation in treatment resistant severe asthma. *PLoS one*. 2014;9(6):e100645.

17. Kim BS, Lee E, Lee MJ, et al. Different functional genes of upper airway microbiome associated with natural course of childhood asthma. *Allergy*. 2018;73(3):644-652.
18. McCauley K, Durack J, Valladares R, et al. Distinct nasal airway bacterial microbiotas differentially relate to exacerbation in pediatric patients with asthma. *The Journal of allergy and clinical immunology*. 2019;144(5):1187-1197.
19. Chiu CY, Chan YL, Tsai YS, et al. Airway Microbial Diversity is Inversely Associated with Mite-Sensitized Rhinitis and Asthma in Early Childhood. *Scientific reports*. 2017;7(1):1820.
20. Nayfach S, Pollard KS. Toward Accurate and Quantitative Comparative Metagenomics. *Cell*. 2016;166(5):1103-1116.
21. Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550(7674):61-66.
22. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490(7418):55-60.
23. Kroegel C. Global Initiative for Asthma (GINA) guidelines: 15 years of application. *Expert Rev Clin Immunol*. 2009;5(3):239-249.
24. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674-1676.
25. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. 2015;3:e1165.
26. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res*. 2010;38(12):e132.
27. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*. 2016;428(4):726-731.
28. Yang SF, Huang HD, Fan WL, et al. Compositional and functional variations of oral microbiota associated with the mutational changes in oral cancer. *Oral oncology*. 2018;77:1-8.
29. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60.
30. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(8):2685-2689.

31. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32.
32. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1-26.
33. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
34. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569(7758):655-662.
35. Yachida S, Mizutani S, Shiroma H, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med*. 2019;25(6):968-976.
36. Larsen JM, Steen-Jensen DB, Laursen JM, et al. Divergent pro-inflammatory profile of human dendritic cells in response to commensal and pathogenic bacteria associated with the airway microbiota. *PloS one*. 2012;7(2):e31976.
37. Larsen JM, Musavian HS, Butt TM, Ingvorsen C, Thysen AH, Brix S. Chronic obstructive pulmonary disease and asthma-associated Proteobacteria, but not commensal Prevotella spp., promote Toll-like receptor 2-independent lung inflammation and pathology. *Immunology*. 2015;144(2):333-342.
38. Brix S, Eriksen C, Larsen JM, Bisgaard H. Metagenomic heterogeneity explains dual immune effects of endotoxins. *J Allergy Clin Immunol*. 2015;135(1):277-280.
39. Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature*. 2009;458(7242):1191-1195.
40. Lopes Dos Santos Santiago G, Brusselle G, Dauwe K, et al. Influence of chronic azithromycin treatment on the composition of the oropharyngeal microbial community in patients with severe asthma. *BMC Microbiol*. 2017;17(1):109.
41. Loureiro CC, Duarte IF, Gomes J, et al. Urinary metabolomic changes as a predictive biomarker of asthma exacerbation. *The Journal of allergy and clinical immunology*. 2014;133(1):261-263 e261-265.
42. Waddell JA, Emerson PA, Gunstone RF. Hypoxia in bronchial asthma. *Br Med J*. 1967;2(5549):402-404.
43. Jung J, Kim SH, Lee HS, et al. Serum metabolomics reveals pathways and biomarkers associated with asthma pathogenesis. *Clin Exp Allergy*. 2013;43(4):425-433.
44. Yoshii K, Hosomi K, Sawane K, Kunisawa J. Metabolism of Dietary and Microbial Vitamin B Family in the Regulation of Host Immunity. *Front Nutr*. 2019;6:48.
45. Qian B, Shen S, Zhang J, Jing P. Effects of Vitamin B6 Deficiency on the Composition and Functional Potential of T Cell Populations. *J Immunol Res*. 2017;2017:2197975.

46. Reynolds RD, Natta CL. Depressed plasma pyridoxal phosphate concentrations in adult asthmatics. *Am J Clin Nutr*. 1985;41(4):684-688.
47. Miller RL, Garfinkel R, Horton M, et al. Polycyclic aromatic hydrocarbons, environmental tobacco smoke, and respiratory symptoms in an inner-city birth cohort. *Chest*. 2004;126(4):1071-1078.
48. Karakiulakis G, Roth M. Muscarinic receptors and their antagonists in COPD: anti-inflammatory and antiremodeling effects. *Mediators Inflamm*. 2012;2012:409580.
49. Ellis TN, Kuehn MJ. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. *Microbiol Mol Biol Rev*. 2010;74(1):81-94.
50. Halhoul N, Colvin JR. The ultrastructure of bacterial plaque attached to the gingiva of man. *Arch Oral Biol*. 1975;20(2):115-118.
51. Barcik W, Pugin B, Westermann P, et al. Histamine-secreting microbes are increased in the gut of adult asthma patients. *The Journal of allergy and clinical immunology*. 2016;138(5):1491-1494 e1497.
52. Barcik W, Pugin B, Bresco MS, et al. Bacterial secretion of histamine within the gut influences immune responses within the lung. *Allergy*. 2019;74(5):899-909.
53. White MV. The role of histamine in allergic diseases. *The Journal of allergy and clinical immunology*. 1990;86(4 Pt 2):599-605.
54. Michalovich D, Rodriguez-Perez N, Smolinska S, et al. Obesity and disease severity magnify disturbed microbiome-immune interactions in asthma patients. *Nat Commun*. 2019;10(1):5711.
55. Sokolowska M, Frei R, Lunjani N, Akdis CA, O'Mahony L. Microbiome and asthma. *Asthma Res Pract*. 2018;4:1.
56. Barber D, Villasenor A, Escribese MM. Metabolomics strategies to discover new biomarkers associated to severe allergic phenotypes. *Asia Pac Allergy*. 2019;9(4):e37.
57. Rodriguez-Coira J, Delgado-Dolset MI, Obeso D, et al. Troubleshooting in Large-Scale LC-ToF-MS Metabolomics Analysis: Solving Complex Issues in Big Cohorts. *Metabolites*. 2019;9(11).

Figure legends

Figure 1. Childhood asthma is associated with disturbances in upper airway microbiome and circulating metabolome. (A) Rarefaction curves based on gene count in cases and non-asthmatic controls. (B-C) α -diversity (Shannon index; B) and β -diversity (Bray-Curtis distance; C) of the two groups at the species level. *** $p < 0.001$, Wilcoxon rank sum test. (D-E) Relative abundances of species normalized by a log transformation (D) and normalized percentage concentrations of metabolites with a log transformation and Pareto scaling (E) discriminatory between the asthma and control group. Significance shown in the right box was determined by using LEfSe. Blue dashed lines indicate a p value of 0.05. Effect size of variables expressed as a log10 transformed LDA score is shown in the left box where bars in red and blue denote enrichment of variables in cases and non-asthmatic controls, respectively.

Figure 2. Airway microbes are associated with asthma-linked metabolite and clinical indices. (A) Correlation between *Prevotella sp. oral taxon 306* and N,N-dimethylglycine (DMG) ($n=53$, Spearman's $r = 0.37$, $p=0.006$, $q=0.15$, the importance of correlation corrected by using the Benjamini-Hochberg procedure, with 24 correlations tested). The loess regression curve indicates the trend of correlation. (B) Associations of differentially abundant species with IgE responses in mite-sensitized childhood asthma (* $q < 0.05$, the importance of correlation corrected by using the Benjamini-Hochberg procedure, with 30 correlations tested). Species in red and blue denote asthma-enriched and control-enriched markers determined by LEfSe.

Figure 3. Airway microbiota in association with host metabolism and clinical indices of asthma. Pearson correlation of airway microbes with serum metabolites and clinical indices, specifically those selected by Random Forests model (RF), LEfSe (LDA, $p < 0.4$), or both against the disease status ($n=53$, * $q < 0.05$, ** $q < 0.01$, *** $q < 0.001$, adjusted by using the Benjamini-Hochberg procedure). Metabolite in red denotes the asthma-enriched marker determined by LEfSe. Glyceril-PC-EtOH, Glycerol-Sn-Glycero-3-phosphocholine-Ethanol; DMSO2-Malonate, Dimethyl-Sulfone-Malonate; PyroGlu-3-OHB, Pyroglutamate-3-Hydroxybutyric acid; D.ptero-IgE, *Dermatophagoides*

pteronysinus-specific immunoglobulin E; *D.farinae*-IgE, *Dermatophagoides farinae*-specific immunoglobulin E; Imp., Model of Importance Variable.

Figure 4. Asthma-associated changes in airway microbial function are linked to host metabolism and IgE responses. (A) Comparisons of the relative abundance of KEGG modules between asthmatics and controls by using the reporter score. Pathway modules with a reporter score $Z > 1.88$ ($> 90\%$ confidence according to a normal distribution) were considered as significantly dysbiotic functions. (B) Abundance of KO genes involved in amino acid metabolism and representative pathway modules showed significant correlation with the levels of circulating metabolites and total and allergen-specific IgE responses ($n=53$, * $q < 0.1$, ** $q < 0.05$, *** $q < 0.01$, by Pearson correlation adjusted using the Benjamini-Hochberg procedure). Metabolites marked in red denote asthma-enriched markers determined by LEfSe, and are labeled as discriminatory by using by Random Forests model (RF), LEfSe (LDA, $p < 0.4$), or both against the disease status. DMSO2-Malonate, Dimethyl-Sulfone-Malonate; PyroGlu-3-OHB, Pyroglutamate-3-Hydroxybutyric acid; *D.ptero*-IgE, *Dermatophagoides pteronyssinus*-specific immunoglobulin E; *D.farinae*-IgE, *Dermatophagoides farinae*-specific immunoglobulin E; Imp., Model of Importance Variable. (C) Relative abundance of bacterial species associated with representative KEGG modules. Top 15 dominant bacterial species are denoted by color (right), and the remaining ones are assigned as others.

Figure 5. Dual-omics signatures discriminate childhood asthma from non-asthmatic controls.

(A) Metagenomic and metabolomic markers for detecting patients with mite-sensitized asthma ($n=25$) from the controls ($n=28$) identified from Random Forests classifiers based on the combination of dual-omics markers. Markers are ranked in descending order of their importance to the accuracy of the model. The boxes represent 25th–75th percentiles, and black lines indicate the median. (B) ROC curves depict trade-offs between true and false positive rates for detecting patients with mite-sensitized asthma as classification stringency varies. AUC, the total area under the ROC curve.

Table 1. Baseline characteristics of 55 children investigated in this study.

Characteristics	Control (n = 28)	Asthma (n = 27)	<i>P</i> value
Age	4.54 ± 0.35	4.32 ± 0.85	0.210
Sex, male	14 (50.0%)	19 (70.4%))	0.123
Parental atopy	20 (71.4%)	20 (87.0%)	0.305
Parental smoking	11 (39.3%)	10 (45.5%)	0.661
Household income			0.957
Low, ≤ 500,000 NTD	10 (35.7%)	8 (32.0%)	0.776
Medium, 500,000-1,000,000 NTD	15 (53.6%)	14 (56.0%)	0.859

High, > 1,000,000 NTD	3 (10.7%)	3 (12.0%)	1.000
Older sibling	12 (42.9%)	14 (58.3%)	0.266
Breastfeeding > 6 mo			0.148
Exclusive	16 (57.1%)	8 (32.0%)	0.051
Partial	5 (17.9%)	5 (20.0%)	1.000
Formula	7 (25.0%)	12 (48.0%)	0.104
Allergen-specific IgE, kU/L			
<i>D. pteronyssinus</i>	0.53 ± 1.24	22.48 ± 28.87	0.003
<i>D. farinae</i>	0.43 ± 1.32	15.45 ± 17.97	<0.001
Egg white	0.15 ± 0.18	1.01 ± 1.34	0.001
Cow's milk	0.18 ± 0.24	0.59 ± 0.87	0.119
Total serum IgE, kU/L	61.73 ± 87.13	467.45 ± 498.84	<0.001
IgE sensitization (> 100 kU/L)	6 (21.4%)	21 (77.8%)	<0.001
Mite sensitization	9 (32.1%)	26 (96.3%)	<0.001
Food sensitization	4 (14.3%)	9 (36.0%)	0.067

Data shown are mean ± SD or number (%) of patients as appropriate. yr, year; NTD, New Taiwan Dollar; IgE, immunoglobulin E.

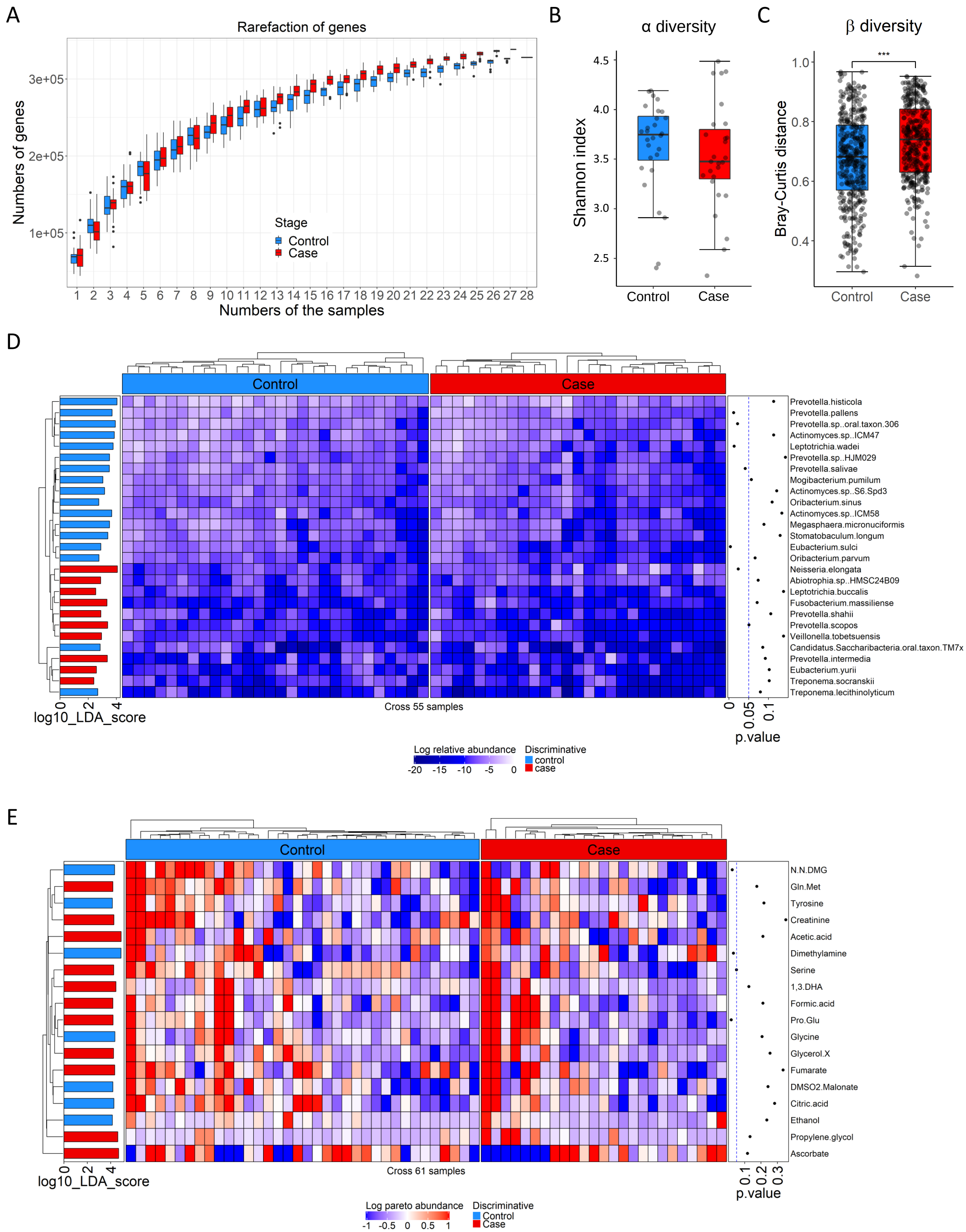
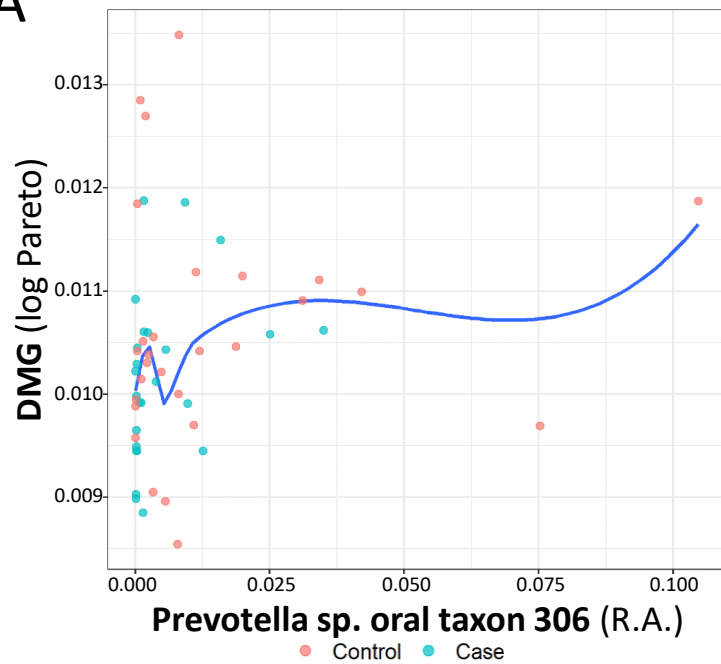


Fig 1

A



B

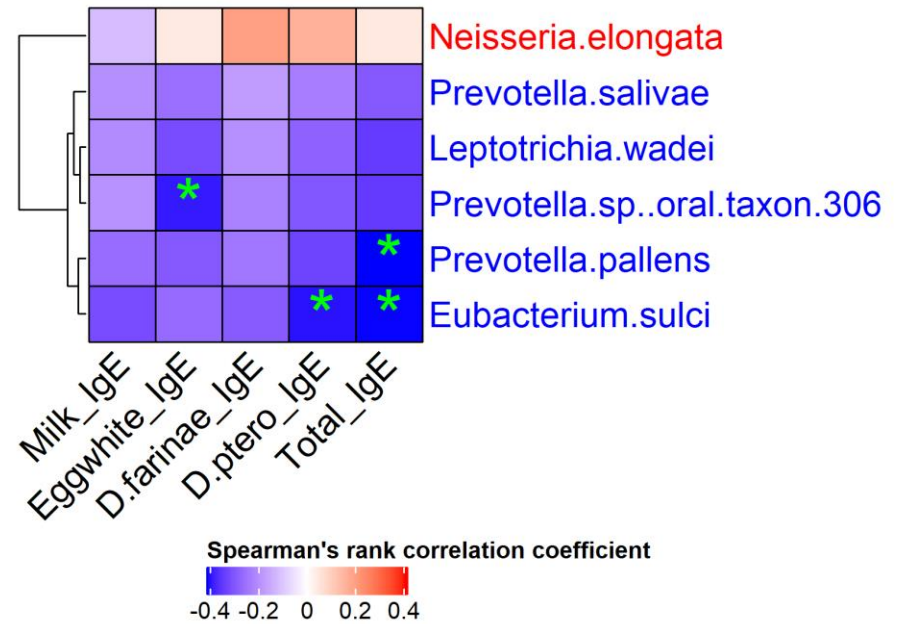


Fig 2

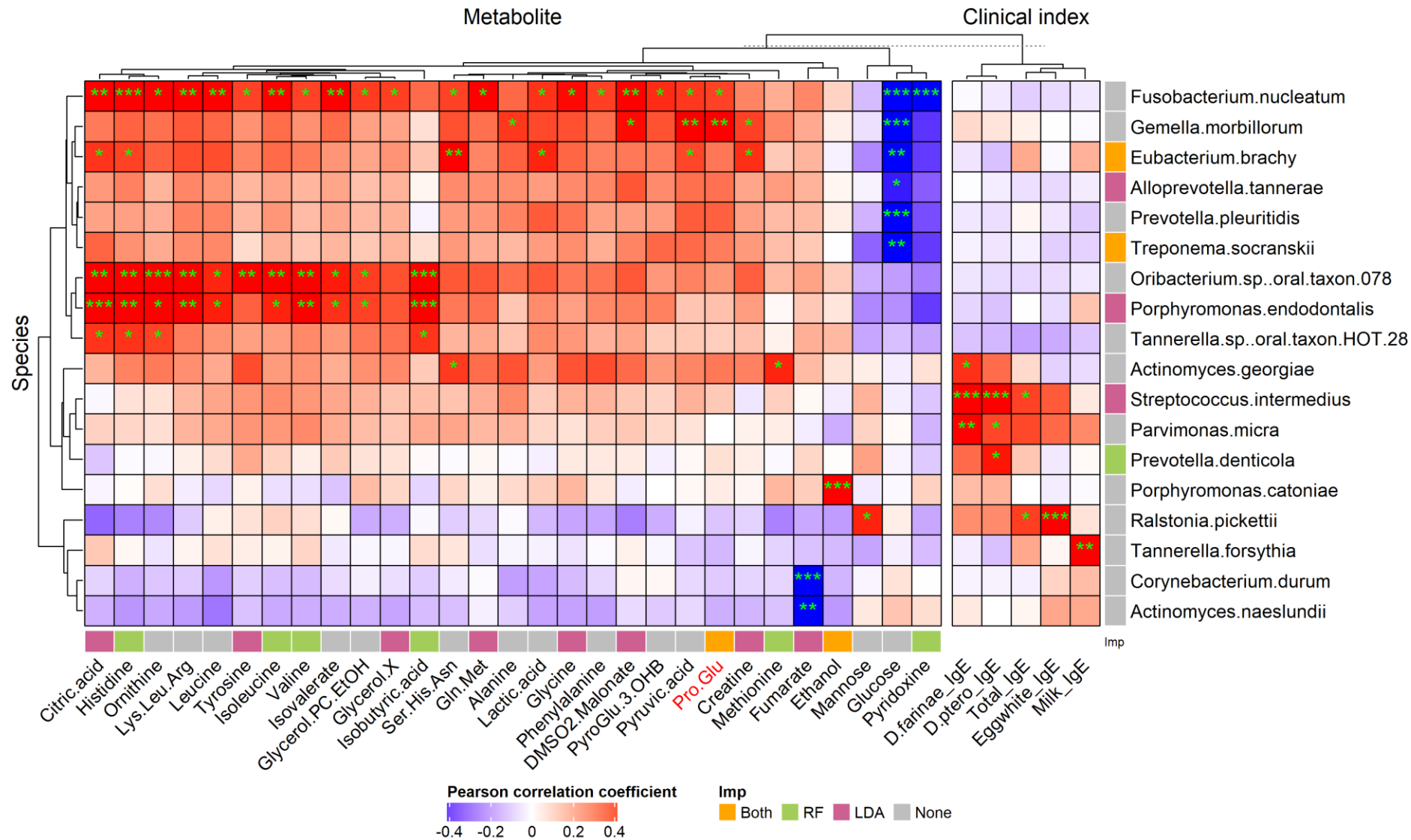
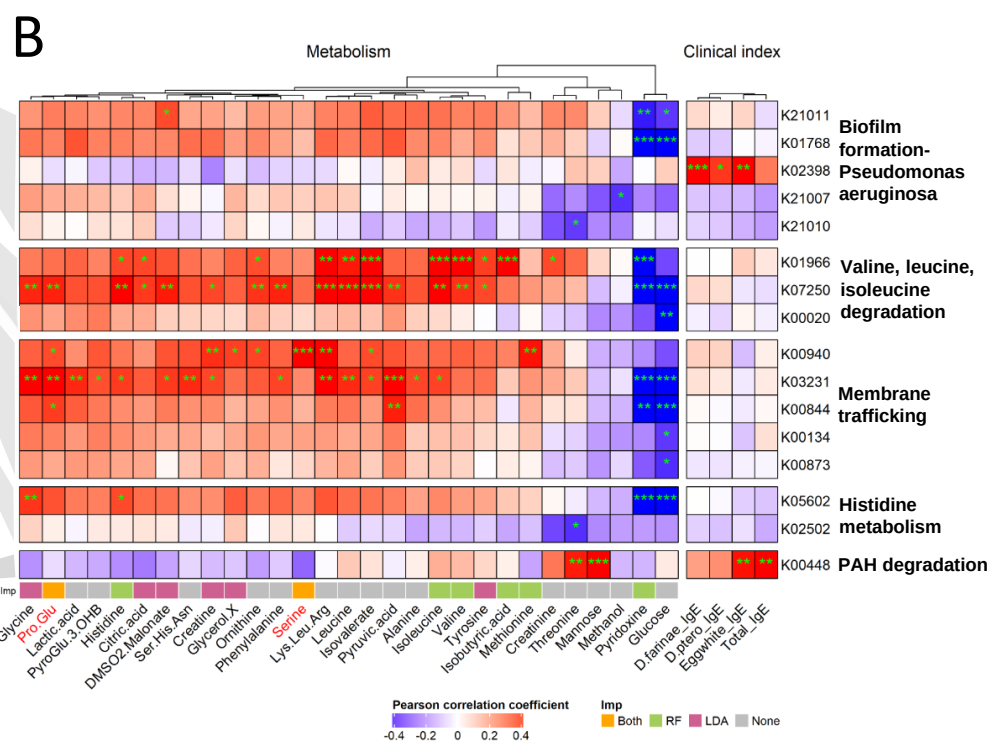
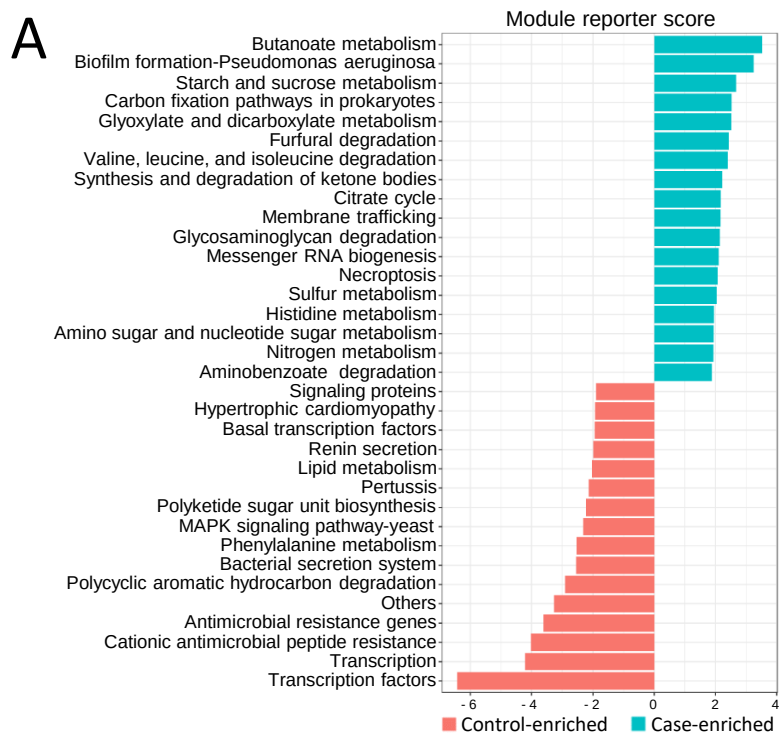
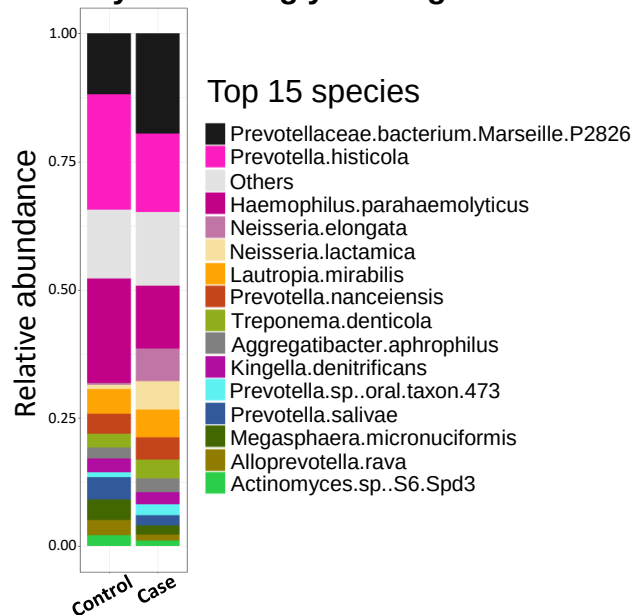


Fig 3

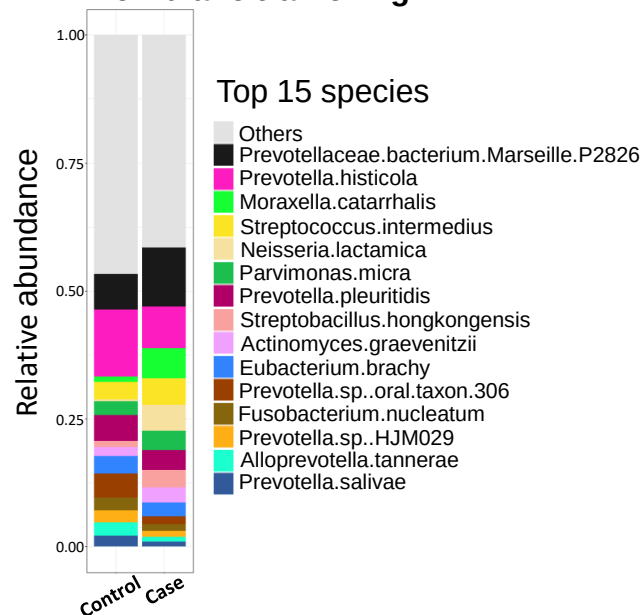


C

Glycosaminoglycan degradation



Membrane trafficking



Histidine metabolism

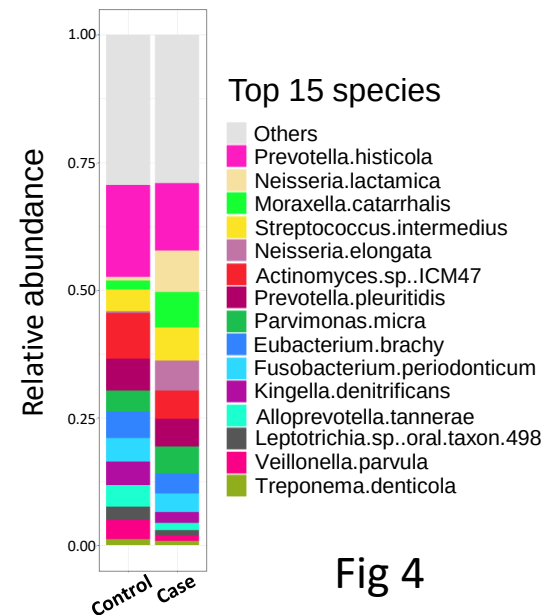
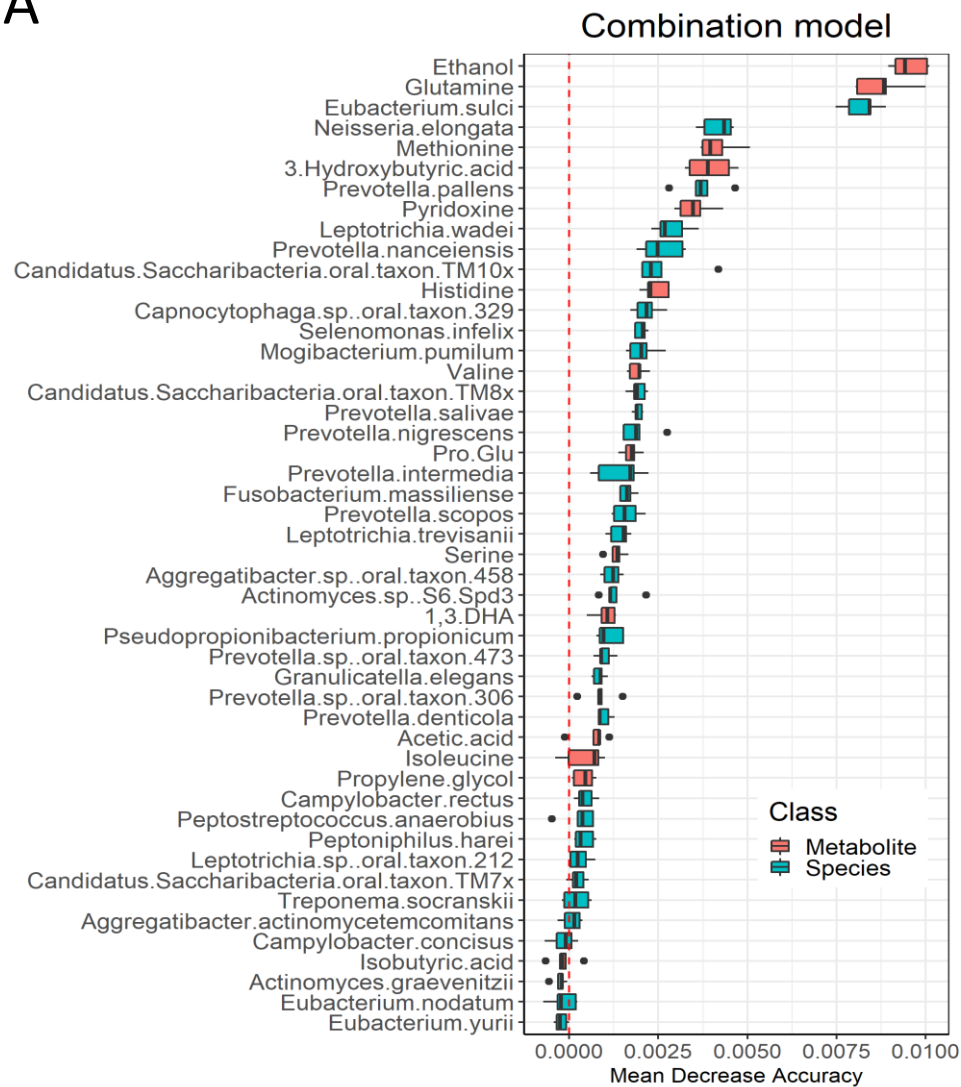


Fig 4

A



B

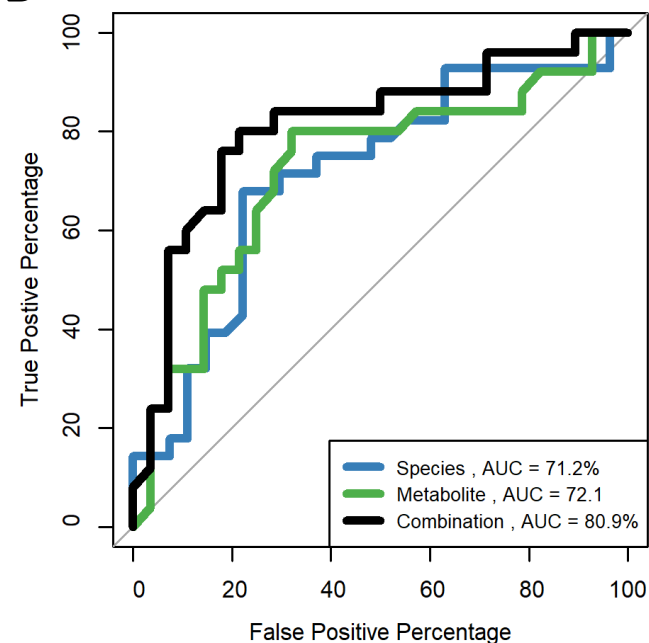


Fig 5

Integration of metagenomics-metabolomics reveals specific signatures and functions of airway microbiota in mite-sensitized childhood asthma

Chih-Yung Chiu^{1,2}, MD, PhD; Hsin-Cheng Chou³, BS; Lun-Ching Chang⁴, PhD; Wen-Lang Fan⁵, PhD; Michael Cong Vinh Dinh⁶, BS; Yu-Lun Kuo⁷, PhD; Wen-Hung Chung⁸, MD, PhD; Hsin-Chih Lai^{9,10}, PhD; Wen-Ping Hsieh³, PhD; Shih-Chi Su^{8,#}, PhD

¹Division of Pediatric Pulmonology, Chang Gung Memorial Hospital at Linkou, College of Medicine, Chang Gung University, Taoyuan, Taiwan

²Clinical Metabolomics Core Laboratory, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

³Institute of Statistics, National Tsing-Hua University, Hsinchu, Taiwan

⁴Department of Mathematical Sciences, Florida Atlantic University, Florida, US

⁵Genomic Medicine Core Laboratory, Chang Gung Memorial Hospital, Linkou, Taiwan

⁶Department of Computer Science, Florida Atlantic University, Florida, US

⁷Biotoools, Co., Ltd, New Taipei City, Taiwan

⁸Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan

⁹Department of Medical Biotechnology and Laboratory Science, and Microbiota Research Center, College of Medicine, Chang Gung University, Taoyuan, Taiwan

¹⁰Central Research Laboratory, XiaMen Chang Gung Hospital, XiaMen, China.

#Corresponding author: Shih-Chi Su, PhD

Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan.

222, Mai-Chin Road, Keelung 20401, Taiwan; Phone: 886-2-24329292-3388; Fax: +886-2-27191623; Email: ssu1@cgmh.org.tw

Running head: Airway microbiome and host metabolome of childhood asthma

Keyword: Airway microbiome, circulating metabolites, childhood asthma

Supplemental methods

Sample collection

5 mL of whole blood was withdrawn for determining serum IgE concentrations and further metabolomics analyses. Levels of total and allergen-specific circulating IgE against two inhalants (*Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*) that caused the sensitization in the vast majority of children in Taiwan ¹ as well as a mixture of three most frequent food allergens (egg white, milk, and wheat) were measured by using commercial kits for IgE (ImmunoCAP Phadiatop Infant; Phadia). Allergic sensitization is defined as ImmunoCAP Phadiatop Infant values ≥ 0.35 kU/L ². Throat swabs were collected without touching any surface (i.e. tongue, mouth, and teeth) other than the oropharynx by a physician using sterile cotton swabs before inhaled or nasal administration of corticosteroids for regular daily treatment. None of the subjects has received antibiotics for at least four weeks prior to the sampling. The swab was immediately kept in the collection tube and stored at -80°C .

Metagenomic data processing and analysis

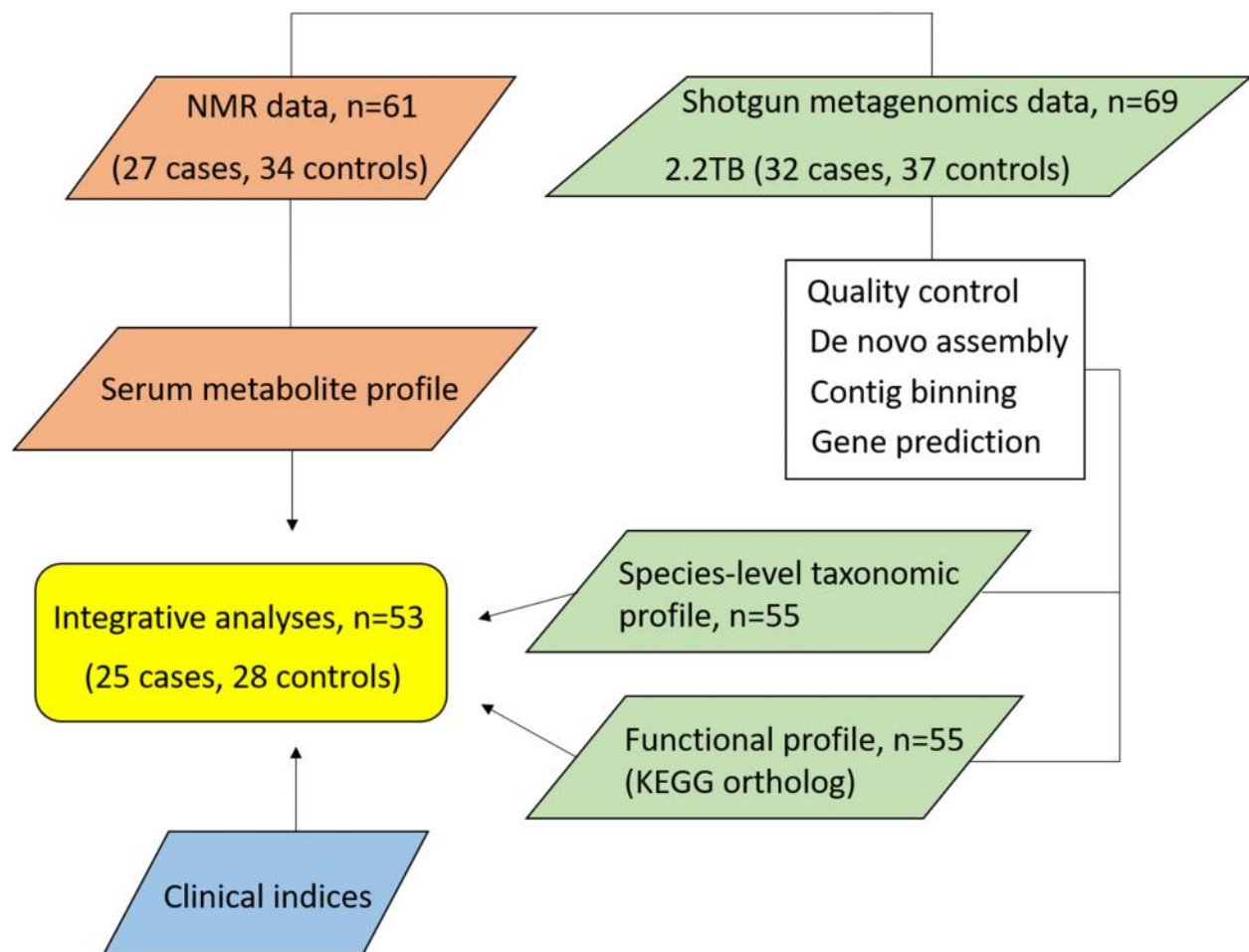
The raw sequence reads were trimmed using in-house scripts to remove the Illumina adapters, and trimmed sequences with the length of shorter than 60 bp or low quality (containing more than 50 % of bases with a Q score ≤ 5) were discarded. Contamination of host sequences were then removed using BowTie2 ³ to map the reads against the reference genome (hg19). The remaining reads were pooled and subject to metagenome assembly using MEGAHIT ⁴ to generate contigs. Clean reads were mapped to contigs using BowTie2, and the mapping output was used for contig binning through MetaBAT ⁵. The resulting bins with sufficient quality ($\geq 50\%$ completeness, $\leq 5\%$ contamination, determined by CheckM ⁶) were considered for

further analysis. Ab initio gene prediction was carried out from assembled contigs using MetaGeneMark ⁷. After dereplication, all non-redundant genes were used to construct the gene catalogue. Gene abundance in each sample was determined by alignment of the reads using BWA MEM against the gene catalogue and measuring the number of reads mapped to each gene sequence. The taxonomic assignment of individual open reading frames (ORFs) was used to annotate bins. A bin was assigned to the taxon to which more than 50% of its ORFs belong to. For constructing the species-level taxonomic profile, the relative abundance of a species was computed per sample, based on the number of reads assigned to the species divided by the total number of aligned reads in the sample. If multiple bins were annotated to the same species, their relative abundance was added up. Functional annotation was performed by mapping annotated ORFs to the Kyoto Encyclopedia of Genes and Genomes (KEGG) protein database using GhostKOALA ⁸ for labeling non-redundant genes with KEGG orthologue (KO) number. The relative abundance of genes from the same KO was summed up and used as the content of this KO in a sample to generate the KO profile.

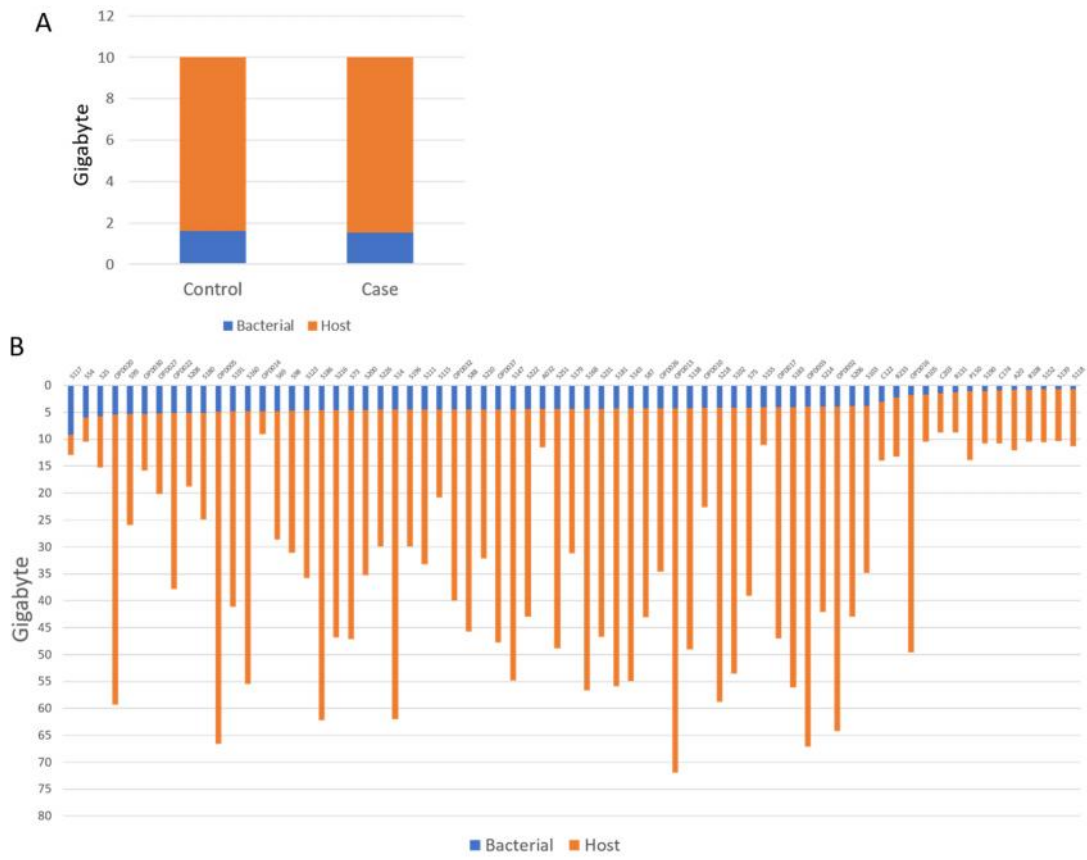
Supplemental results

Shotgun metagenome sequencing of throat swabs and taxonomic profiling

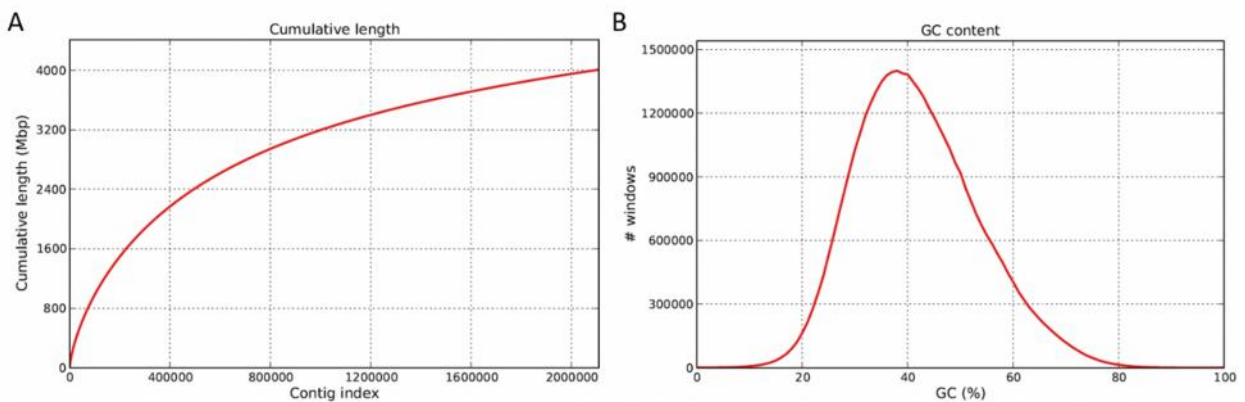
A crucial requirement in the metagenomics studies is that a sufficient volume of sequencing data is obtained to enable reliable quantitative comparisons among samples, as the number of genes that are detected in any given sample increases with the volume of sequencing data until saturation. Yet, obtaining sufficient sequencing data is especially challenging for non-fecal samples. To address this, we initially sequenced 69 DNA samples from throat swabs to 10 gigabyte per sample. Following quality control, the sequencing reads from the host were filtered. The proportion of microbial reads to the total in each sample varied and averaged at only 16.08 and 15.33% for the control and case group, respectively (**Figure S2A**). Additional rounds of sequencing (2.2 terabyte in total) were performed to bring the sequencing depth to at least 4 gigabyte of bacterial sequences per sample for 55 samples (**Figure S2B**). After sufficient sequencing data were collected, the bacterial sequencing reads from all 69 samples were pooled and *de novo* assembled into a set of 2109988 contigs that together comprise the metagenome (**Table S2**), with a N50 of 3009 bp. The total length of assembled contigs is 4008.3 Mbp (**Figure S3A**), and the GC content of which is 41.85% (**Figure S3B**). Clustering of assembled contigs generated 318 phylogenetic bins. A total of 156 bins was assigned to the species level, from which 129 unique species-level clades were annotated (**Table S3**) and profiled (**Table S4**).



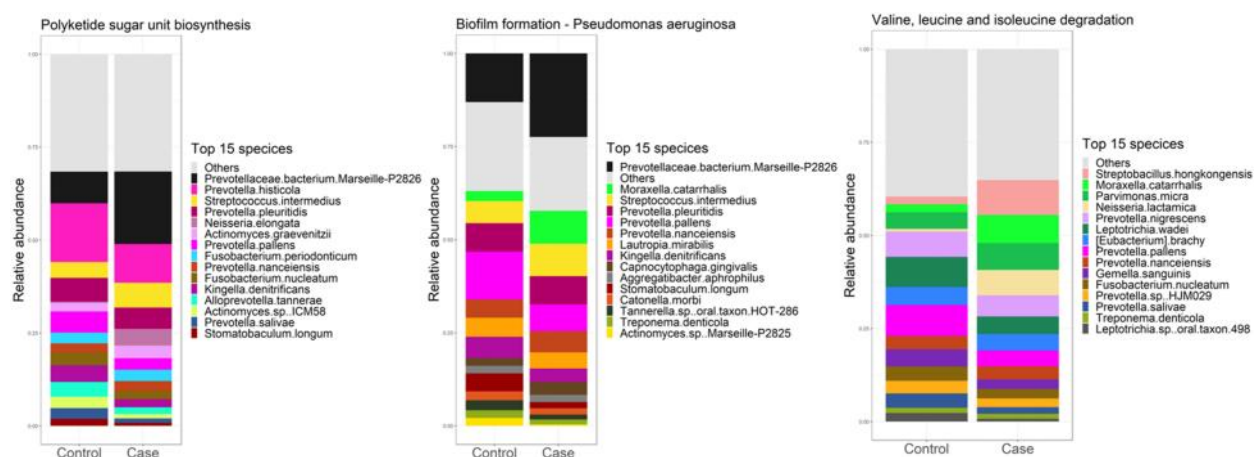
Supplemental figure 1. Overview of the study. Throat swabs from 69 subjects were used to obtain deep shotgun sequencing data, from which functional and species-level taxonomic profiles of 55 samples were generated after stringent QC. In addition, serum samples from 61 subjects were used to conduct NMR spectroscopy analysis to generate metabolite profiles. Samples from 53 subjects (25 cases, 28 controls) were available for both the sequencing and metabolomics data analyses. QC, quality control; NMR, nuclear magnetic resonance.



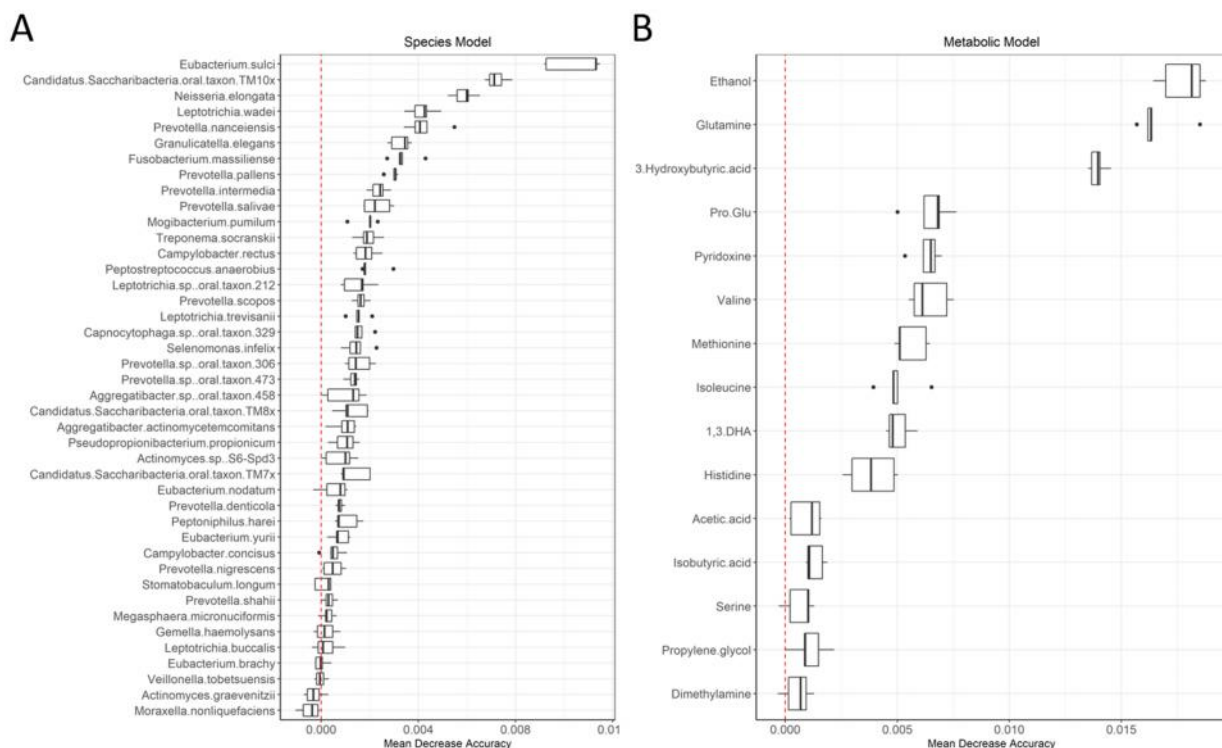
Supplemental figure 2. (A) The averaged ratio of bacterial reads to the host reads in the control and case group. **(B)** The amounts of bacterial reads to the host reads in each sample after multiple rounds of sequencing.



Supplemental figure 3. (A) Cumulative length of assembled contigs **(B)** Distribution of GC content of assembled contigs.



Supplemental figure 4. Relative abundance of bacterial species associated with representative KEGG modules in mite-sensitized childhood asthma. Top 15 dominant bacterial species are denoted by color (right), and the remaining ones are assigned as others.



Supplemental figure 5. Metagenomic and metabolomic markers for detecting patients with mite-sensitized asthma from the controls identified from Random Forests classifiers based on the species (A) and metabolite profile (B). Markers are ranked in descending order of their importance to the accuracy of the model. The boxes represent 25th–75th percentiles, and black lines indicate the mean.

Supplemental table 1.

Clinical validity for *E. sulci* in detecting childhood asthma and mite sensitization.

	Asthma (27) vs non-asthmatic control (28)	IgE sensitization (27) vs tolerant control (28)	D. ptero sensitization (26) vs tolerant control (29)	D. farinae sensitization (19) vs tolerant control (36)
AUC	0.73	0.78	0.74	0.70
P value	0.0033	0.0003	0.0037	0.018

E. sulci, *Eubacterium sulci*; IgE, immunoglobulin E; D.ptero, *Dermatophagoides pteronyssinus*; D.farinae, *Dermatophagoides farinae*; AUC, the total area under the ROC curve.

Supplemental table 2. Statistics of assembled contigs

# contigs (>= 0 bp)	2109988
# contigs (>= 1000 bp)	1116010
# contigs (>= 5000 bp)	154256
# contigs (>= 10000 bp)	30714
# contigs (>= 25000 bp)	1513
# contigs (>= 50000 bp)	187
Total length (>= 0 bp)	4008366077
Total length (>= 1000 bp)	3319023435
Total length (>= 5000 bp)	1282767641
Total length (>= 10000 bp)	448155002
Total length (>= 25000 bp)	56562633
Total length (>= 50000 bp)	14451872
# contigs	2109988
Largest contig	277185
Total length	4008366077
GC (%)	41.85
N50	3009
N75	1345
L50	342122
L75	845749
# N's per 100 kbp	0.00

All statistics are based on contigs of size ≥500 bp unless otherwise indicated.

Supplemental table 3. Annotation results of phylogenetic bins.

	Kingdom	Phylum	Class	Order	Family	Genus	Species
Annotated bin#	17	22	12	13	13	83	156
Unique taxa#	1	5	3	4	6	24	129

Reference

1. Braman SS. The global burden of asthma. *Chest* 2006;130(1 Suppl):4S-12S. doi: 10.1378/chest.130.1_suppl.4S
2. Ballardini N, Nilsson C, Nilsson M, et al. ImmunoCAP Phadiatop Infant--a new blood test for detecting IgE sensitisation in children at 2 years of age. *Allergy* 2006;61(3):337-43. doi: 10.1111/j.1398-9995.2005.00936.x
3. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357-9. doi: 10.1038/nmeth.1923
4. Li D, Liu CM, Luo R, et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31(10):1674-6. doi: 10.1093/bioinformatics/btv033
5. Kang DD, Froula J, Egan R, et al. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015;3:e1165. doi: 10.7717/peerj.1165
6. Parks DH, Imelfort M, Skennerton CT, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25(7):1043-55. doi: 10.1101/gr.186072.114
7. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* 2010;38(12):e132. doi: 10.1093/nar/gkq275
8. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* 2016;428(4):726-31. doi: 10.1016/j.jmb.2015.11.006