Indoor microbiome, microbial and plant metabolites, chemical compounds, and asthma symptoms in junior high school students: a multicentre association study in Malaysia

Yu Sun [1,2,6,7], Mei Zhang [1,6], Zheyuan Ou [1], Yi Meng [1], Yang Chen [1], Ruqin Lin [1], Jamal Hisham Hashim [3], Zailina Hashim [4], Gunilla Wieslander [5], Qingsong Chen [2], Dan Norbäck [5] and Xi Fu [2,7]

[1]Guangdong Provincial Key Laboratory of Protein Function and Regulation in Agricultural Organisms, College of Life Sciences, South China Agricultural University, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Zoonosis of the Ministry of Agriculture and Rural Affairs, Guangzhou, PR China. [2]Guangdong Provincial Engineering Research Center of Public Health Detection and Assessment, School of Public Health, Guangdong Pharmaceutical University, Guangzhou, PR China. [3]Dept of Environmental Health and Occupational Safety, Faculty of Health Sciences, Universiti Selangor, Shah Alam, Selangor, Malaysia. [4]Dept of Environmental and Occupational Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia (UPM), Serdang, Selangor, Malaysia. [5]Occupational and Environmental Medicine, Dept of Medical Science, University Hospital, Uppsala University, Uppsala, Sweden. [6]Yu Sun and Mei Zhang contributed equally to this work. [7]Yu Sun and Xi Fu contributed equally to this article as lead authors and supervised the work.

Corresponding author: Xi Fu (fuxi@gdpu.edu.cn)

Abstract

Background Indoor microbial exposure is associated with asthma, but the health effects of indoor metabolites and chemicals have not been comprehensively assessed.

Methods We collected classroom dust from 24 junior high schools in three geographically distanced areas in Malaysia (Johor Bahru, Terengganu and Penang), and conducted culture-independent high-throughput microbiome and untargeted metabolomics/chemical profiling.

Results 1290 students were surveyed for asthma symptoms (wheeze). In each centre, we found significant variation in the prevalence of wheeze among schools, which could be explained by personal characteristics and air pollutants. Large-scale microbial variations were observed between the three centres; the potential protective bacteria were mainly from phyla Actinobacteria in Johor Bahru, Cyanobacteria in Terengganu and Proteobacteria in Penang. In total, 2633 metabolites and chemicals were characterised. Many metabolites were enriched in low-wheeze schools, including plant secondary metabolites flavonoids/ isoflavonoids (isoliquiritigenin, formononetin, astragalin), indole and derivatives (indole, serotonin, 1*H*-indole-3-carboxaldehyde), and others (biotin, chavicol). A neural network analysis showed that the indole derivatives were co-occurring with the potential protective microbial taxa, including *Actinomycetospora*, *Fischerella* and *Truepera*, suggesting these microorganisms may pose health effects by releasing indole metabolites. A few synthetic chemicals were enriched in high-wheeze schools, including pesticides (2 (3*H*)-benzothiazolethione), fragrances (2-aminobenzoic acid, isovaleric acid), detergents and plastics (phthalic acid), and industrial materials (4,4-sulfonyldiphenol).

Conclusions This is the first association study between high-throughput indoor chemical profiling and asthma symptoms. The consistent results from the three centres indicate that indoor metabolites/chemicals could be a better indicator than the indoor microbiome for environmental and health assessments, providing new insights for asthma prediction, prevention and control.

Introduction

Asthma is a common allergy-related chronic respiratory disease that affects more than 350 million patients worldwide. The prevalence of asthma symptoms (wheeze and whistling) is >30% in many countries,

including Australia, Ireland and the UK, posing a severe threat to public health [1]. Epidemiology studies show that the occurrence of asthma and allergies is mainly affected by environmental exposure, including air pollution, environmental allergens and microorganisms [2–4]. A striking phenomenon is that the prevalence of asthma is significantly lower in children growing up in farming or rural areas than in urban areas [5]. Subsequent studies revealed that indoor microbial exposure is the driving factor for the variation [6]. Studies in Finland, Germany, Malaysia and China also confirmed the importance of the indoor microbiome in immune modulation and disease development [7–10]. However, it is challenging to transform the theoretical progress into practical applications, such as building an indoor microbiome indicator for environmental assessment and disease prediction. This is because the diversity of the indoor microbiome is extremely high. The total number of microbial species on Earth is approximately 1 trillion [11]. Also, the indoor microbiome shows extremely high geographic diversity. Different sets of indoor microorganisms and health-associated microorganisms are characterised in different geographical regions [7–10]. It is almost impossible to find a consensus set of health-related species across the globe and make a solid indoor microbiome–health inference. Thus, an alternative environmental assessment indicator is needed.

Indoor metabolites and chemicals could be a potential alternative for environmental assessment. Each bacterial and fungal organism can release thousands of metabolites into the living environment per hour, affecting the health of occupants. The common health-related metabolites include lipopolysaccharide (LPS), muramic acid and microbial volatile organic compounds (MVOCs) [12, 13]. However, previous studies used culture-dependent or low-throughput approaches to characterise a small set of targeted chemical exposures from microorganisms. No study used a high-throughput untargeted approach to profile comprehensive indoor metabolites and chemicals. Thus, the overall picture of metabolites/chemicals in the indoor environment is still unclear. Also, no study has conducted multi-omic analysis between indoor microorganisms and metabolites to identify the potential microbial sources of metabolites.

In this study, we surveyed 1290 junior high school students in Johor Bahru, Terengganu and Penang in Malaysia for asthma symptoms, and collected classroom dust for culture-independent high-throughput microbiome and untargeted chemical profiling. We aimed to characterise indoor metabolic/chemical exposures and uncover their relationships with health-related microorganisms. Also, we compared the environmental chemical pattern in multiple centres and tested whether it could be a better indicator than the indoor microbiome for exposure assessment.

Materials and methods
Study design and health data
We conducted classroom dust sampling and health surveys in three areas in Malaysia: Johor Bahru, Terengganu and Penang. The locations are displayed in supplementary figure S1. In each centre, eight junior high schools (four classrooms in each school) were randomly selected for dust sampling. In each class, health questionnaires in Malay were sent to 15–20 students aged 14–15 years. Health questions were obtained from the International Study of Asthma and Allergies in Childhood study [14], including an asthma symptom question: "In the last 12 months, have you had wheezing or whistling in the chest when you DID NOT have a cold or the flu?". Personal information was collected, including age and gender. The participants had no information regarding environmental data and samples collected. The study design and protocol were approved by the Medical Research and Ethics Committee of the National University of Malaysia (Selangor, Malaysia). Informed consent was obtained from all participants.

Dust sampling and environmental characteristic measurements
We sampled classroom dust on floors, desks, chairs, bookshelves and curtains with a vacuum cleaner in Johor Bahru and Terengganu [7, 15]. The vacuum procedure was maintained at 4 min: 2 min on the floor and 2 min on other surfaces above floor level, including student desks, chairs, bookshelves and curtains. The dust was collected in a sampler (ALK-Abelló, Copenhagen, Denmark) with a filter pore size of 6 µm. In schools in Penang, we collected settled dust on the upper frame of the blackboard with a metal spoon. The vacuumed and settled dust was sieved to fine dust through a metal mesh screen (pore size 0.3 mm). The fine dust was stored in a freezer at −80°C.

Indoor relative humidity and carbon dioxide were measured by a Q-Trak indoor air quality monitor (TSI, St Paul, MN, USA). Indoor nitrogen dioxide was sampled by a diffusion sampler from IVL Swedish Environmental Research (Gothenburg, Sweden).

High-throughput amplicon sequencing
Bacterial and fungal amplicon sequencing was performed by using dust samples. In brief, total microbial DNA was extracted from 10 mg fine dust by an E.Z.N.A. Soil DNA Kit (D5625-01; Omega Bio-Tek,

Norcross, GA, USA) and a DNA SPIN Kit (MP Biomedicals, Santa Ana, CA, USA). DNA quality was assessed by a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and agarose gel electrophoresis. Bacterial 16S rRNA genes and fungal internal transcribed spacer regions were amplified, and sample-specific barcode sequences were added during the library preparation step. The amplicons were sequenced by Illumina MiSeq and PacBio platforms. Raw sequence data were deposited in the QIITA microbial study management platform (https://qiita.ucsd.edu; 12875) and the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa; CRA002825, CRA002876, CRA005646 and CRA005647) [16]. The absolute bacterial and fungal concentration was quantified by quantitative PCR with universal primers [15].

Profiling chemical compounds in classroom dust

Chemical compounds in classroom dust were assessed by untargeted liquid chromatography-mass spectrometry (LC-MS) (BioNovoGene, Suzhou, China). 10 mg fine dust was added to 0.6 mL 2-chlorophenylalanine in methanol and centrifuged at 12 000 rpm at 4°C for 10 min. 300 µL supernatant was filtered through a 0.22 µm membrane. Chromatographic separation was performed by Acquity UPLC HSS T3 columns (2.1×150 mm, 1.8 µm; Waters, Milford, MA, USA) at 40°C at a flow rate of 0.25 mL·min^{-1}. A Vanquish HPLC systems Q Exactive HF-X hybrid quadrupole–orbitrap mass spectrometer was used for LC-MS detection (Thermo Fisher Scientific). Electrospray ionisation MS experiments were performed with spray voltages of 3.5 kV and −2.5 kV in positive and negative modes. Sheath and auxiliary gas and capillary temperature were set at 30 and 10 arbitrary units and 325°C. The analyser scanned a mass range of m/z 81–1000 at a mass resolution of 60 000. Chemicals were annotated by searching against mzCloud (www.mzcloud.org), Human Metabolome Database (www.hmdb.ca), MoNA (https://mona. fiehnlab.ucdavis.edu), METLIN (https://metlin.scripps.edu) and MassBank (www. massbank.jp).

High-throughput data analysis

Microbiome data processing and analysis were conducted on the QIIME 2 platform [17]. Raw reads were assigned to samples according to the barcode information. Low-quality and chimeric reads were removed. Sequence taxonomy was annotated by the Silva (release 115) and UNITE (release 5) databases [18, 19]. LEfSe (linear discriminant analysis (LDA) effect size) analysis was conducted to characterise the enriched microbial taxa in different groups (LDA >2) [20]. Mmvec was used to estimate microbe–metabolite interaction co-occurrence probabilities [21] and the results were visualised by TBtools [22]. The key features of the chemical compounds and Globally Harmonized System (GHS) classification were obtained from the PubChem database [23].

Results

Prevalence of asthma symptoms in three centres in Malaysia

In total, 1290 junior high school students from Johor Bahru, Terengganu and Penang were surveyed for asthma symptoms/wheeze. The three centres were located in the south, northeast and northwest Malaysia (supplementary figure S1), and eight junior high schools were randomly selected in each centre. We found large-scale variation in the prevalence of asthma symptoms among schools in Johor Bahru, Terengganu and Penang (p=0.0007, p=0.0004 and p=0.002, respectively) (table 1). For example, the prevalence of wheeze was 12.7%, 20.0%, 14.3%, 29.9%, 8.9%, 8.7%, 5.6% and 9.1% in the eight schools in Terengganu. In each centre, the top four schools were defined as "high-wheeze" schools and the bottom four schools were defined as "low-wheeze" schools.

We further explored the personal and environmental characteristics that could explain the variation of asthma symptoms among schools. Students' age and gender did not differ between high- and low-wheeze schools (p>0.05, Mann–Whitney test). The indoor environmental characteristics, including temperature, relative humidity and carbon dioxide concentration, were similar among the three centres (table 2). Indoor nitrogen dioxide concentration was lower in Terengganu than in Johor Bahru and Penang. In all centres, these environmental characteristics did not differ between high- and low-wheeze schools (p>0.05), failing to explain the variation of asthma symptoms.

Indoor microorganisms enriched in the three centres and high/low-wheeze schools

High-throughput sequencing was applied to characterise the abundance of indoor microorganisms in Johor Bahru, Terengganu and Penang (supplementary tables S1–S8). We found strong geographic variation in indoor bacteria and fungi among centres. For example, bacterial genera *Ralstonia*, *Ochrobactrum* and *Craurococcus* were present in high abundance (7.8%, 2.7% and 1.3%, respectively) in Penang, but were presented in low abundance in Johor Bahru and Terengganu (0% and 0.06%, 0.02% and 0.42%, and 0.05% and 0.74%, respectively) (supplementary table S7). Fungal genera *Wallemia* and *Candida* were present in high abundance in Johor Bahru and Terengganu (6.6% and 12.2%, and 1.3% and 1.8%,

TABLE 1 Prevalence of asthma symptoms (wheeze) among junior high school students in three centres in Malaysia								
	Students (n)	Prevalence (%)	Total students (n)	Prevalence (%)	p-value	Male (%)	Female (%)	p-value
Johor Bahru								
Schools with high prevalence	57	15.8	236	14.8	0.0007	45.5	54.5	0.051
	60	13.3						
	59	11.9						
	60	18.3						
Schools with low prevalence	56	7.1	226	5.3		54.5	45.5	
	54	1.9						
	56	3.6						
	60	8.3						
Terengganu								
Schools with high prevalence	63	12.7	241	19.5	0.0004	40.7	59.3	0.41
	55	20.0						
	56	14.3						
	67	29.9						
Schools with low prevalence	45	8.9	222	8.1		36.9	63.1	
	46	8.7						
	54	5.6						
	77	9.1						
Penang								
Schools with high prevalence	56	23.2	158	21.6	0.002	49.4	50.6	0.21
	35	20.0						
	22	18.2						
	44	22.7						
Schools with low prevalence	52	13.5	210	10.0		42.9	57.1	
	47	10.6						
	42	9.5						
	67	7.5						

In each centre, the top four schools were defined as schools with a high prevalence of asthma symptoms ("high-wheeze") and the bottom four schools were defined as schools with a low prevalence of asthma symptoms ("low-wheeze"). p-values were calculated by the Chi-squared test.

respectively) but in low abundance in Penang (0.06% and 0.03%, respectively) (supplementary table S8). Common mould species *Penicillium* and *Cladosporium* were present in high abundance only in Johor Bahru (10.2% and 7.8%, respectively). *Fusarium* was present in high abundance in Terengganu but not in Johor Bahru and Penang (2.9% *versus* 0.3% and 0.009%). *Aspergillus* was present in high abundance in all centres (20.7%, 12.7% and 12.1%, respectively).

We further explored the potential indoor microorganisms enriched in high/low-wheeze schools in each centre. In Johor Bahru, 15 bacterial taxa were enriched in low-wheeze schools and more than half were from phylum Actinobacteria, including *Rubrobacter, Actinomyces, Blastococcus, Janibacter, Actinomycetospora, Pseudokineococcus* and *Marmoricola* (figure 1 and supplementary figure S2). However, in Terengganu, the bacteria enriched in low-wheeze schools were mainly from phylum

TABLE 2 Environmental characteristics in junior high schools in three centres in Malaysia			
	Johor Bahru	Terengganu	Penang
Indoor temperature (°C)	29.0 (27.3–30.9)	29.5 (28.3–32.8)	28.1 (25.4–30.3)
Indoor relative humidity (%)	69.9 (62.9–76.9)	72.4 (63.8–79.9)	78.9 (69.7–88.2)
Indoor carbon dioxide (ppm)	492 (376–689)	432 (397–600)	425 (360–720)
Indoor nitrogen dioxide (µg·m^{-3})	23.4 (13.0–41.7)	8.2 (2.8–14.6)	22.9 (20.5–29.9)

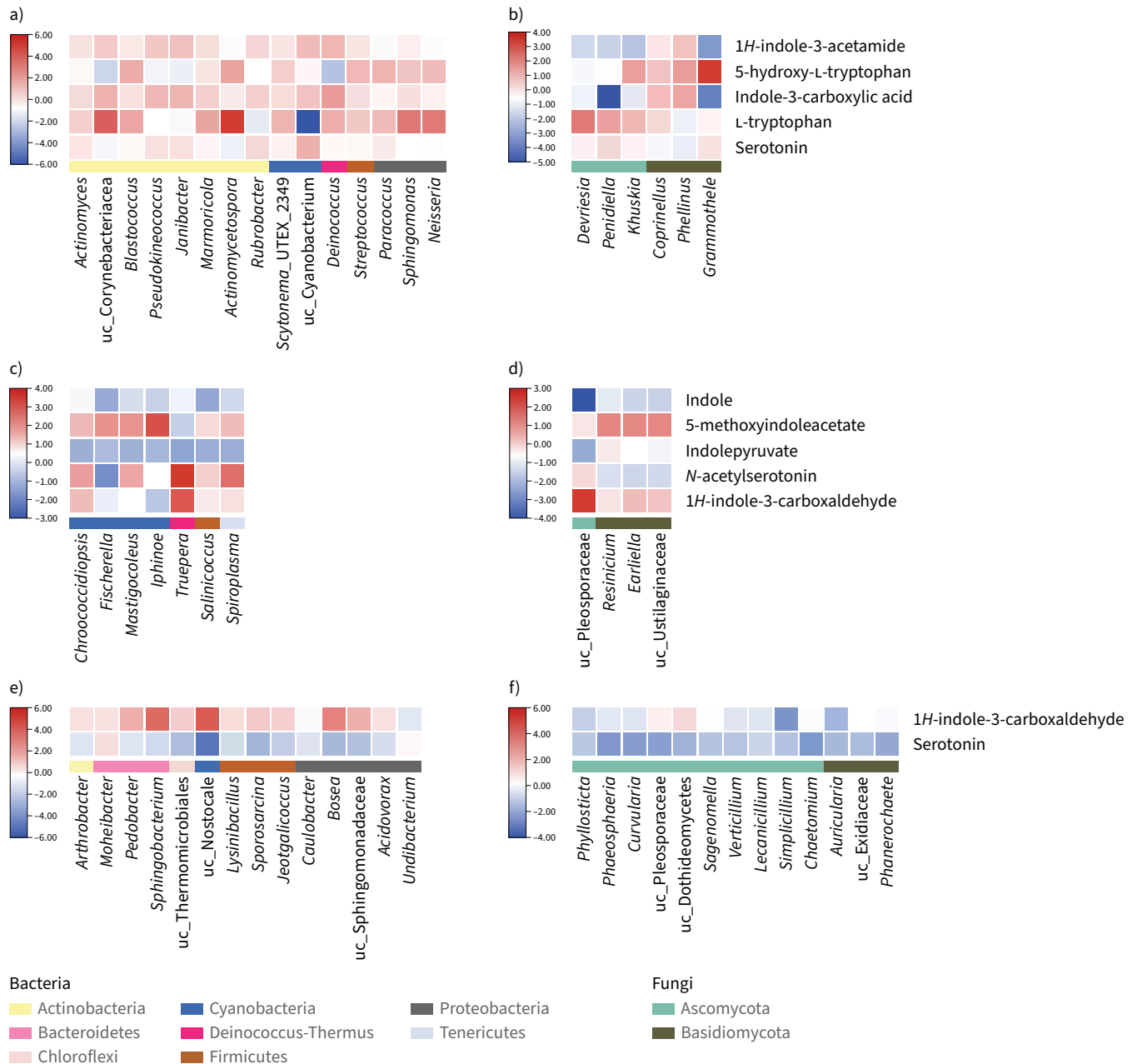Data are presented as mean (minimum–maximum).

FIGURE 1 Co-occurrence probability of potential protective microbial taxa and indoles and derivatives in three centres in Malaysia: a) bacterial and b) fungal taxa and metabolites in Johor Bahru; c) bacterial and d) fungal taxa and metabolites in Terengganu; and e) bacterial and f) fungal taxa and metabolites in Penang. The potential protective microbial taxa are presented at the x-axes of the heat plots and the potential protective metabolites are presented on the y-axes. The potential protective taxa were defined as bacterial and fungal taxa enriched in schools with a low asthma prevalence (linear discriminant analysis (LDA) >3 in Johor Bahru; LDA >2 in Terengganu and Penang). The potential protective metabolites were defined as indoles and derivatives enriched in schools with a low asthma prevalence (p<0.01, false discovery rate <0.1, fold change >2). uc: uncharacterised.

Cyanobacteria, including *Chroococcidiopsis*, *Fischerella*, *Mastigocoleus* and *Iphinoe*. In Penang, the bacteria enriched in low-wheeze schools were mainly from phylum Proteobacteria, including *Caulobacter*, *Bosea*, *Acidovorax*, *Undibacterium* and uncharacterised (uc)_Sphingomonadaceae. The results indicate that different centres have a unique set of potential protective bacterial taxa. The bacterial taxa enriched in high-wheeze schools also differed among centres. *Catellicoccus* and *Ignatzschineria* were enriched in high-wheeze schools in Johor Bahru, *Methylobacterium*, *Bacillus*, *Sphingomonas* and *Pantoea* were

enriched in Terengganu, and *Roseomonas* was enriched in Penang (supplementary figure S2). The potential protective fungal taxa were from phyla Ascomycota and Basidiomycota (figure 1).

Quantitative PCR was also conducted in Johor Bahru and Terengganu (supplementary tables S9 and S10). The absolute concentration of indoor bacteria and fungi did not differ between high- and low-wheeze schools ($p > 0.1$, t-test).

Potential protective metabolites and associations with microorganisms

In total, 2633 chemicals were characterised by LC-MS. The geographic pattern was also observed for chemical compounds (supplementary figure S3). Indoor chemical composition in Johor Bahru, Terengganu and Penang is located, respectively, in the right, upper-left and lower-left side of the ordination plot (supplementary figure S3b). However, general rules were also observed. Three classes of metabolites were almost exclusively enriched in low-wheeze schools ($p < 0.01$, false discovery rate (FDR) < 0.1, fold change > 2), including flavonoids, isoflavonoids, and indole and derivatives (table 3 and figure 1). Flavonoids and isoflavonoids are important classes of plant secondary metabolites, widely found in various plants, fruits and vegetables. Three flavonoids and isoflavonoids were enriched in two or more centres, including isoliquiritigenin, formononetin and 6-hydroxydaidzein. Other flavonoid metabolites, including baicalin, astragalin, tangeritin, daidzein, luteolin and procyanidin B2, were enriched in one centre (table 3). Indole and derivatives form a class of common small signalling molecules in microorganisms, plants and animals. Serotonin and indole-3-carboxaldehyde were enriched in two or more centres, and indole, L-tryptophan, 1*H*-indole-3-acetamide and indolepyruvate were enriched in one centre. The explicit enrichment of

	Classification	Name	Molecular formula	Representative plants	Suggested health effects
Commonly enriched in Johor Bahru, Terengganu and Penang	Flavonoids	Isoliquiritigenin	$C_{15}H_{12}O_4$	*Glycyrrhiza uralensis*, *Mongolian glycyrrhiza*, *Glycyrrhiza glabra*	Antioxidant, anti-inflammatory, GABA modulator
	Isoflavonoids	Formononetin	$C_{16}H_{12}O_4$	*Trifolium pratense*	Antioxidant
		6-hydroxydaidzein	$C_{15}H_{10}O_5$	Soybean	Anti-inflammatory
Johor Bahru	Flavonoids	Afzelechin	$C_{15}H_{14}O_5$	*Bergenia ligulata*	
		Amentoflavone	$C_{30}H_{18}O_{10}$	*Ginkgo biloba*, *Hypericum perforatum*	Antiallergic
		Astragalin	$C_{21}H_{20}O_{11}$	*Phytolacca americana*, *Phegopteris connectilis*	Anti-inflammatory, antioxidant, antiatopic dermatitis
		Baicalin	$C_{21}H_{18}O_{11}$	Chinese herbal medicine *Scutellaria baicalensis*	Antiallergic, attenuated serum IgE and effector T-cells
		Diosmetin	$C_{16}H_{12}O_6$	Citrus	Antiallergic
		Epicatechin	$C_{15}H_{14}O_6$	*Pentace burmanica*	Antioxidant, reduces resistance to insulin
		Tangeritin	$C_{20}H_{20}O_7$	Citrus, tangerines	Antioxidant
	Isoflavonoids	Glycitein	$C_{16}H_{12}O_5$	Soybean	Improves sleep quality
Terengganu	Flavonoids	(2*S*)-liquiritigenin	$C_{15}H_{12}O_4$	*Glycyrrhizae uralensis*	Antioxidant
	Isoflavonoids	Daidzein	$C_{15}H_{10}O_4$	Soy plants	Anti-inflammatory
Penang	Flavonoids	Luteolin	$C_{15}H_{10}O_6$	Pteridophyta, Bryophyta, Magnoliophyta, Pinophyta	Anti-inflammatory, immune system modulator
		Procyanidin B2	$C_{30}H_{26}O_{12}$	*Cinchona pubescens*, *Cinnamomum verum*, *Crataegus monogyna*, *Uncaria guianensis*, *Vitis vinifera*, *Litchi chinensis*, *Ecdysanthera utilis*	Therapeutic for acute lung injury
		Quercetin 3-(6-malonyl-glucoside)	$C_{24}H_{22}O_{15}$	*Lactuca sativum*	
	Isoflavonoids	Ononin	$C_{22}H_{22}O_9$	*Trifolium pratense*	

TABLE 3 Flavonoids and isoflavonoids enriched in schools with a low asthma prevalence in three centres in Malaysia

The potential protective flavonoids and isoflavonoids were defined as enriched in schools with a low asthma prevalence compared with schools with a high asthma prevalence ($p < 0.01$, false discovery rate < 0.1, fold change > 2). The representative plants and suggested health effects were mainly acquired from PubChem. It should be noted that these are only representative plants; not all plants can produce these flavonoids and isoflavonoids. GABA: γ-aminobutyric acid.

flavonoids and indole derivatives in the dust of low-wheeze schools suggests their potential anti-inflammatory and antiallergic effects. A neural network analysis showed that the indole and its derivatives were co-occurring with many potential protective microbial taxa, such as L-tryptophan with *Actinomycetospora* and uc_Corynebacteriacea, *N*-acetylserotonin and indole with *Truepera*, indole with uc_Pleosporaceae, and 5-methoxyindoleacetate with *Fischerella*. The results suggest that these microorganisms may produce these metabolites. A literature search showed that *Actinomycetospora* and *Fischerella* were capable of producing indole derivatives [24, 25], supporting the *in silico* association results.

Other potential protective metabolites were also identified, including biotin, chavicol, ecgonine, dihydrocortisol, *etc.* (figure 2), which belong to different metabolic classes. Biotin was closely associated with many protective bacteria, including *Actinomyces*, *Paracoccus* and *Sphingomonas*. A literature search in laboratory experiments showed that these taxa could produce biotin [26–28], consistent with the co-occurrence analysis.

Potential risk chemicals were all synthetic chemicals

Potential risk environmental chemicals were defined as chemicals significantly enriched in high-wheeze schools (p<0.01, FDR <0.1, fold change >2). One (2(3*H*)-benzothiazolethione), one (2-aminobenzoic acid) and three (isovaleric acid, phthalic acid and 4,4-sulfonyldiphenol) hazardous chemicals were characterised in Johor Bahru, Terengganu and Penang, respectively. These chemicals included pesticides, cleaning detergents, perfumes and industrial materials (table 4). The GHS has classified them as hazardous chemicals; adverse health effects include dermatitis, hypersensitivity, inflammation, and eye and respiratory tract irritation. More hazardous chemicals were detected in Penang, which may explain the overall high prevalence of asthma symptoms compared with Johor Bahru.

Discussion

This is the first study to use a high-throughput untargeted approach to profile indoor chemical exposure and asthma. Natural metabolites, including microbial and plant metabolites, were protective for asthma, and synthetic chemicals, including pesticides, detergents and industrial solvents, were risk factors for asthma. Also, this is the first study to assess the interactions between indoor microbiomes and metabolites, revealing that indoor microorganisms may produce protective metabolites. In addition, three centres in Malaysia with large geographic separation were surveyed to support our results and conclusions. We found large-scale indoor microbiome variation, indicating an overall different microbiome exposure in each centre. However, general rules were observed for indoor chemical compounds. Plant metabolites from flavonoids and isoflavonoids and microbial metabolites from indole derivatives showed potential protective effects, whereas synthetic chemicals showed adverse health effects. The results suggest that indoor chemicals could be a more solid and consistent indicator in exposure assessments for asthma.

There are also limitations in this study. First, dust in Johor Bahru and Terengganu was collected from vacuuming dust from floors, desks, tables, bookshelves and curtains, whereas dust in Penang was collected from the blackboard frame, which may produce sampling bias. However, we argue that the two approaches should be comparable. Bookshelves, curtains and blackboard frames are seldom or never cleaned in these classrooms, and thus both approaches collect dust representing long-term exposure. Second, we applied second-generation amplicon sequencing in this study, only resolving taxonomic resolution at the genus level. Third, as only the marker gene is sequenced, the abundance and health associations for function genes cannot be assessed. However, our study profiled indoor microbial metabolites, which provide more direct evidence for metabolic exposures than functional gene assessment.

Flavonoids and isoflavonoids are plant secondary metabolites with a polyphenolic structure. Flavonoids are found in many plants and plant-derived food, and isoflavonoids are predominantly found in soybeans and leguminous plants [29]. Flavonoids and isoflavonoids have anti-inflammatory, antioxidative, anticarcinogenic and antiallergic properties for humans and animals (table 3). Several flavonoids identified have shown protective effects for asthma in previous studies. For example, isoliquiritigenin suppresses interleukin (IL)-4 and IL-5 production in a dose-dependent manner *in vitro* [30], and astragalin reduces IL-4, IL-5 and IL-13 levels and inhibits eosinophil infiltration in mice [31]. Also, formononetin alleviates lung inflammation and cytokine levels and reduces oxidative stress in an ovalbumin-sensitised mouse model [32]. However, the previous studies mainly reported the health effects of flavonoids and isoflavonoids in laboratory animals and cell lines. This is the first study to show their beneficial effects as inhaled exposure for human populations, providing a novel perspective on respiratory disease.
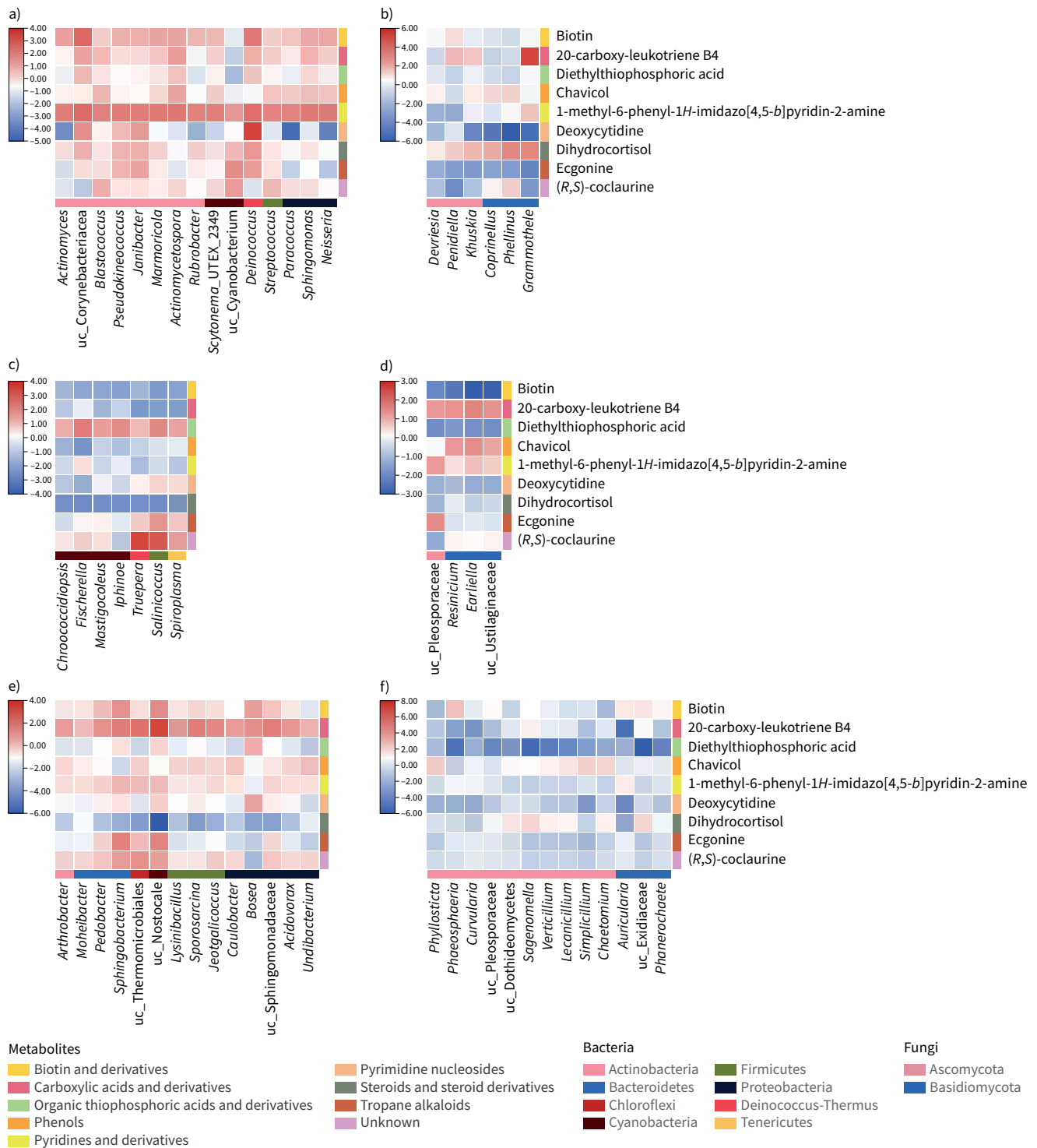
FIGURE 2 Co-occurrence probability of potential protective microbial taxa and other protective metabolite classes in three centres in Malaysia: a) bacterial and b) fungal taxa and metabolites in Johor Bahru; c) bacterial and d) fungal taxa and metabolites in Terengganu; and e) bacterial and f) fungal taxa and metabolites in Penang. The potential protective microbial taxa are presented on the x-axis of the heat plot and the potential protective metabolites are presented on the y-axis. The potential protective taxa were defined as bacterial and fungal taxa enriched in schools with a low asthma prevalence (linear discriminant analysis (LDA) >3 in Johor Bahru; LDA >2 in Terengganu and Penang). The potential protective microbial metabolites were defined as metabolites enriched in schools with a low asthma prevalence compared with schools with a high asthma prevalence (p<0.01, false discovery rate <0.2, fold change >2) in each centre. The potential protective metabolites were categorised into Human Metabolome Database class and the metabolite classes are also presented. uc: uncharacterised.

TABLE 4 Potential risk environmental chemicals in three centres in Malaysia

	Environmental chemicals	Molecular formula	Key features and manufacturing products	GHS hazard statements	Symptoms, disorders and diseases	Experimental evidence
Johor Bahru	2(3H)-benzothiazolethione	$C_7H_5NS_2$	Industrial materials, paints, pesticide	H317: May cause an allergic skin reaction	Dermatitis, hypersensitivity	[47]
Terengganu	2-aminobenzoic acid	$C_7H_7NO_2$	Dyes, perfumes	H318 (42.55%): Causes serious eye damage H319 (57.45%): Causes serious eye irritation	Nausea, allergic reaction, skin irritation, itching and dermatitis	
Penang	Isovaleric acid	$C_5H_{10}O_2$	Fragrance, perfumes	H314 (100%): Causes severe skin burns and eye damage H318 (84.07%): Causes serious eye damage	Inflammatory bowel disease, autism	[48]
	Phthalic acid	$C_8H_6O_4$	Cleaning products, laundry, pesticides, plastic products	H315 (96.4%): Causes skin irritation H319 (94.96%): Causes serious eye irritation H335 (95.68%): May cause respiratory irritation	Eyes, skin, respiratory tract and mucous membrane irritation	
	4,4-sulfonyldiphenol	$C_{12}H_{10}O_4S$	Electroplating solvent, washfastening agent, metabolite and endocrine disruptor	H319 (15.38%): Causes serious eye irritation	Obesity, glucose intolerance	[49]

The potential risk environmental chemicals were defined as chemicals enriched in schools with a high asthma prevalence compared with schools with a low asthma prevalence (p<0.01, false discovery rate <0.1, fold change>2) in each centre. The characteristics information for chemicals was searched in the PubChem database from the National Center for Biotechnology Information [23]. Chemicals annotated as drug or common human, animal or plant metabolites were not included in the table. Globally Harmonized System (GHS) classifications, including GHS hazard code, hazard statements and notified classification ratio, are presented (GHS hazard codes between H310 and H336 are presented as these hazard statements mainly relate to allergic irritations).

Indole and derivatives form a group of aromatic heterocyclic organic compounds widely distributed in bacteria, plants and animals [33]. The health effects of indoles were mainly studied in the human gut, with few studies on environmental microorganisms. A large variety of gut microorganisms can produce indole and derivatives, including *Clostridium novyi*, *Escherichia coli*, *Fusobacterium*, *Enterococcus faecalis* and *Corynebacterium acnes* [34, 35]. Indole and derivatives can improve human intestinal epithelial barrier integrity and reduce gut inflammation by decreasing the expression of pro-inflammatory cytokine NF-κB and increasing anti-inflammatory IL-10 [36]. Indolepyruvate and indole-3-acetamide were identified as potential protective metabolites in our study. In human gut studies, these indole derivatives can activate the expression of the aryl hydrocarbon receptor gene [37], which has an anti-inflammatory role in blocking pro-inflammatory T-cells in asthma development [38]. The inhaled exposure of indole metabolites could have a similar mechanism by activating aryl hydrocarbon receptors in the lung and respiratory tract.

Besides flavonoids and indoles, biotin and chavicol were enriched in low-wheeze schools. Biotin, a B7 vitamin, is an essential nutrient for humans. A previous indoor metagenomics survey reported that a higher abundance of biotin metabolism pathways was associated with a lower prevalence of sick building syndrome [39]. Chavicol is a natural phenylpropene found in *Piper betle*, and is used in the traditional herbal medicine of China and India. Chavicol analogues can attenuate interferon-γ expression in T-helper cells, and modulate inflammation and immune responses [40], supporting their roles in reducing asthma symptoms.

Only a few chemicals were significantly enriched in high-wheeze schools after removing drugs and common human and plant metabolites, and all of them were synthetic chemicals, including pesticides, paints, fragrances and industrial solvents. Phthalate exposure is reported to associate with asthma. A meta-analysis of 43 studies reported that benzyl phthalate increased the odds of childhood asthma by 39–41% [41]. A home survey in China also reported that a high concentration of phthalic acid esters increased childhood diagnosed asthma [42]. The other potential risk chemicals are not reported to associate with asthma by scientific publications, but the GHS classification, developed by the United Nations, indicated that these chemicals might have adverse health effects, including dermatitis, hypersensitivity and respiratory tract irritation. Thus, future environmental surveys and asthma epidemiology should also consider these chemicals.

Previous indoor metabolite studies mainly surveyed microbial metabolites by low-throughput approaches. ARAKI et al. [12] and CHOI et al. [43] reported that many MVOCs were positively associated with asthma and rhinitis. In our study, only three MVOCs (bornyl acetate, 2-heptanone and estragole) were detected in vacuum dust and none of them were significantly enriched in high/low-wheeze schools. It is likely that most volatile chemicals cannot be detected by vacuum dust sampling. The total LPS concentration seems to be mainly protectively associated with asthma [44]. Muramic acid was reported to be positively or negatively associated with asthma [45, 46]. In this study, the untargeted LC-MS also detected muramic acid and LPS (tridecanoic acid, hydroxy hexadecanoic acid and myristic acid), but none reached significance after the FDR adjustment.

Conclusions

In this study, we found large-scale variation in the microbiome composition and health-related microorganisms in three centres in Malaysia. This could be due to the extremely high diversity of environmental microorganisms [11]. Thus, it is challenging to use the indoor microbial composition to build a universal reference catalogue for health assessments and disease prediction. However, consistent associations were observed for indoor chemical compounds, suggesting they could be used as an environmental assessment indicator for disease prediction, providing new insights and strategies for disease prevention and control.

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: 733–743.

2   von Mutius E, Smits HH. Primary prevention of asthma: from risk and protective factors to targeted strategies for prevention. Lancet 2020; 396: 854–866.

3   Lu C, Norbäck D, Li Y, et al. Early-life exposure to air pollution and childhood allergic diseases: an update on the link and its implications. Expert Rev Clin Immunol 2020; 16: 813–827.

4   Lu C, Norbäck D, Zhang Y, et al. Furry pet-related wheeze and rhinitis in pre-school children across China: associations with early life dampness and mould, furry pet keeping, outdoor temperature, PM_{10} and $PM_{2.5}$. Environ Int 2020; 144: 106033.

5   von Mutius E, Vercelli D. Farm living: effects on childhood asthma and allergy. Nat Rev Immunol 2010; 10: 861–868.

6   Ege MJ, Mayer M, Normand AC, et al. Exposure to environmental microorganisms and childhood asthma. N Engl J Med 2011; 364: 701–709.

7   Fu X, Norbäck D, Yuan Q, et al. Indoor microbiome, environmental characteristics and asthma among junior high school students in Johor Bahru, Malaysia. Environ Int 2020; 138: 105664.

8   Fu X, Li Y, Meng Y, et al. Derived habitats of indoor microbes are associated with asthma symptoms in Chinese university dormitories. Environ Res 2021; 194: 110501.

9   Kirjavainen PV, Karvonen AM, Adams RI, et al. Farm-like indoor microbiota in non-farm homes protects children from asthma development. Nat Med 2019; 25: 1089–1095.

10   Fu X, Ou Z, Zhang M, et al. Indoor bacterial, fungal and viral species and functional genes in urban and rural schools in Shanxi Province, China – association with asthma, rhinitis and rhinoconjunctivitis in high school students. Microbiome 2021; 9: 138.

11   Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci USA* 2016; 113: 5970.

12   Araki A, Kanazawa A, Kawai T, *et al.* The relationship between exposure to microbial volatile organic compound and allergy prevalence in single-family homes. *Sci Total Environ* 2012; 423: 18–26.

13   Norbäck D, Markowicz P, Cai G-H, *et al.* Endotoxin, ergosterol, fungal DNA and allergens in dust from schools in Johor Bahru, Malaysia – associations with asthma and respiratory infections in pupils. *PLoS One* 2014; 9: e88303.

14   Beasley R. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. *Lancet* 1998; 351: 1225–1232.

15   Fu X, Meng Y, Li Y, *et al.* Associations between species-level indoor microbiome, environmental characteristics, and asthma in junior high schools of Terengganu, Malaysia. *Air Qual Atmos Health* 2022; 15: 1043–1055.

16   Wang Y, Song F, Zhu J, *et al.* GSA: Genome Sequence Archive. *Genomics Proteomics Bioinformatics* 2017; 15: 14–18.

17   Bolyen E, Rideout JR, Dillon MR, *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; 37: 852–857.

18   Quast C, Pruesse E, Gerken J, *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013; 41: D590–D596.

19   Koljalg U, Nilsson RH, Abarenkov K, *et al.* Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* 2013; 22: 5271–5277.

20   Segata N, Waldron L, Ballarini A, *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 2012; 9: 811–814.

21   Morton JT, Aksenov AA, Nothias LF, *et al.* Learning representations of microbe–metabolite interactions. *Nat Methods* 2019; 16: 1306–1314.

22   Chen C, Chen H, Zhang Y, *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 2020; 13: 1194–1202.

23   Kim S, Thiessen PA, Bolton EE, *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res* 2016; 44: D1202–D1213.

24   Saito S, Oku N, Igarashi Y. Mycetoindole, an *N*-acyl dehydrotryptophan with plant growth inhibitory activity from an actinomycete of the genus Actinomycetospora. *J Antibiot* 2022; 75: 44–47.

25   Hillwig ML, Zhu Q, Liu X. Biosynthesis of ambiguine indole alkaloids in cyanobacterium *Fischerella ambigua*. *ACS Chem Biol* 2014; 9: 372–377.

26   Saito II, Honda H, Kawabe T, *et al.* Comparison of biotin production by recombinant *Sphingomonas* sp. under various agitation conditions. *Biochem Eng J* 2000; 5: 129–136.

27   Feng Y, Kumar R, Ravcheev DA, *et al. Paracoccus denitrificans* possesses two BioR homologs having a role in regulation of biotin metabolism. *MicrobiologyOpen* 2015; 4: 644–659.

28   Akishina RI, Voronkova VV, Pomortseva NV. [Effect of the nutrient medium composition on biotin biosynthesis by an *Actinomyces* species 313–152 culture]. *Prikl Biokhim Mikrobiol* 1982; 18: 339–342.

29   Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. *J Nutr Sci* 2016; 5: e47.

30   Yang N, Patil S, Zhuge J, *et al. Glycyrrhiza uralensis* flavonoids present in anti-asthma formula, ASHMI, inhibit memory Th2 responses *in vitro* and *in vivo*. *Phytother Res* 2013; 27: 1381–1391.

31   Liu J, Cheng Y, Zhang X, *et al.* Astragalin attenuates allergic inflammation in a murine asthma model. *Inflammation* 2015; 38: 2007–2016.

32   Yi L, Cui J, Wang W, *et al.* Formononetin attenuates airway inflammation and oxidative stress in murine allergic asthma. *Front Pharmacol* 2020; 11: 533841.

33   Kumar P, Lee J-H, Lee J. Diverse roles of microbial indole compounds in eukaryotic systems. *Biol Rev* 2021; 96: 2522–2545.

34   Lee J-H, Lee J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol Rev* 2010; 34: 426–444.

35   Yano JM, Yu K, Donaldson GP, *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* 2015; 161: 264–276.

36   Bansal T, Alaniz RC, Wood TK, *et al.* The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. *Proc Natl Acad Sci USA* 2010; 107: 228–233.

37   Vyhlídalová B, Krasulová K, Pečinková P, *et al.* Gut microbial catabolites of tryptophan are ligands and agonists of the aryl hydrocarbon receptor: a detailed characterization. *Int J Mol Sci* 2020; 21: 2614.

38   Poulain-Godefroy O, Bouté M, Carrard J, *et al.* The aryl hydrocarbon receptor in asthma: friend or foe? *Int J Mol Sci* 2020; 21: 8797.

39   Fu X, Ou Z, Zhang M, *et al.* Classroom microbiome, functional pathways and sick-building syndrome (SBS) in urban and rural schools – potential roles of indoor microbial amino acids and vitamin metabolites. *Sci Total Environ* 2021; 795: 148879.

40   Min HJ, Nam J-W, Yu ES, *et al.* Effect of naturally occurring hydroxychavicol acetate on the cytokine production in T helper cells. *Int Immunopharmacol* 2009; 9: 448–454.

41 Li MC, Chen CH, Guo YL. Phthalate esters and childhood asthma: a systematic review and congener-specific meta-analysis. *Environ Pollut* 2017; 229: 655–660.

42 Zhang J, Sun C, Lu R, *et al.* Associations between phthalic acid esters in household dust and childhood asthma in Shanghai, China. *Environ Res* 2021; 200: 111760.

43 Choi H, Schmidbauer N, Bornehag CG. Volatile organic compounds of possible microbial origin and their risks on childhood asthma and allergies within damp homes. *Environ Int* 2017; 98: 143–151.

44 Williams LK, Ownby DR, Maliarik MJ, *et al.* The role of endotoxin and its receptors in allergic disease. *Ann Allergy Asthma Immunol* 2005; 94: 323–332.

45 van Strien RT, Engel R, Holst O, *et al.* Microbial exposure of rural school children, as assessed by levels of *N*-acetylmuramic acid in mattress dust, and its association with respiratory health. *J Allergy Clin Immunol* 2004; 113: 860–867.

46 Karvonen AM, Hyvarinen A, Rintala H, *et al.* Quantity and diversity of environmental microbial exposure and development of asthma: a birth cohort study. *Allergy* 2014; 69: 1092–1101.

47 Remy S, Verstraelen S, Van Den Heuvel R, *et al.* Gene expressions changes in bronchial epithelial cells: markers for respiratory sensitizers and exploration of the NRF2 pathway. *Toxicol In Vitro* 2014; 28: 209–217.

48 Marchesi JR, Holmes E, Khan F, *et al.* Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *J Proteome Res* 2007; 6: 546–551.

49 Azevedo LF, Porto Dechandt CR, Cristina de Souza Rocha C, *et al.* Long-term exposure to bisphenol A or S promotes glucose intolerance and changes hepatic mitochondrial metabolism in male Wistar rats. *Food Chem Toxicol* 2019; 132: 110694.