ORIGINAL ARTICLE

WILEY

Classifying asthma control using salivary and fecal bacterial microbiome in children with moderate-to-severe asthma

Jelle M. Blankestijn[1,2,3] | Alejandro Lopez-Rincon[4,5] | Anne H. Neerincx[1] |
Susanne J. H. Vijverberg[1] | Simone Hashimoto[1,6] | Mario Gorenjak[7] |
Olaia Sardón Prado[8,9] | Paula Corcuera-Elosegui[8] | Javier Korta-Murua[8] |
Maria Pino-Yanes[10,11,12] | Uroš Potočnik[7,13] | Corinna Bang[14] | Andre Franke[14] |
Christine Wolff[15] | Susanne Brandstetter[15] | Antoaneta A. Toncheva[16] |
Parastoo Kheiroddin[16] | Susanne Harner[16] | Michael Kabesch[15,16] |
Aletta D. Kraneveld[4] | Mahmoud I. Abdel-Aziz[1,2,3,17] |
Anke H. Maitland-van der Zee[1,2,3,6] | the SysPharmPediA Consortium

[1]Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[2]Amsterdam Institute for Infection and Immunity, Amsterdam, The Netherlands

[3]Amsterdam Public Health, Amsterdam, The Netherlands

[4]Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands

[5]Department of Data Science, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

[6]Department of Pediatric Respiratory Medicine, Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands

[7]Center for Human Molecular Genetics and Pharmacogenomics, Faculty of Medicine, University of Maribor, Maribor, Slovenia

[8]Division of Pediatric Respiratory Medicine, Hospital Universitario Donostia, San Sebastián, Spain

[9]Department of Pediatrics, University of the Basque Country (UPV/EHU), San Sebastián, Spain

[10]Genomics and Health Group, Department of Biochemistry, Microbiology, Cell Biology and Genetics, Universidad de La Laguna (ULL), San Cristóbal de La Laguna, Spain

[11]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

[12]Instituto de Tecnologías Biomédicas (ITB), Universidad de La Laguna, La Laguna, Spain

[13]Laboratory for Biochemistry, Molecular Biology and Genomics, Faculty of Chemistry and Chemical Engineering, University of Maribor, Maribor, Slovenia

[14]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

[15]Science and Development Campus Regensburg (WECARE), University Children's Hospital Regensburg (KUNO) at the Hospital St. Hedwig of the Order of St. John, University of Regensburg, Regensburg, Germany

[16]Department of Pediatric Pneumology and Allergy, University Children's Hospital Regensburg (KUNO) at the Hospital St. Hedwig of the Order of St. John, University of Regensburg, Regensburg, Germany

[17]Department of Clinical Pharmacy, Faculty of Pharmacy, Assiut University, Assiut, Egypt

Correspondence
Jelle M. Blankestijn, Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands.
Email: j.m.blankestijn@amsterdamumc.nl

Abstract

Background: Uncontrolled asthma can lead to severe exacerbations and reduced quality of life. Research has shown that the microbiome may be linked with asthma characteristics; however, its association with asthma control has not been explored. We

Jelle M. Blankestijn and Alejandro Lopez-Rincon shared first author.

Mahmoud I. Abdel-Aziz and Anke H. Maitland-van der Zee shared last author.

aimed to investigate whether the gastrointestinal microbiome can be used to discriminate between uncontrolled and controlled asthma in children.

Methods: 143 and 103 feces samples were obtained from 143 children with moderate-to-severe asthma aged 6 to 17 years from the SysPharmPediA study. Patients were classified as controlled or uncontrolled asthmatics, and their microbiome at species level was compared using global (alpha/beta) diversity, conventional differential abundance analysis (DAA, analysis of compositions of microbiomes with bias correction), and machine learning [Recursive Ensemble Feature Selection (REFS)].

Results: Global diversity and DAA did not find significant differences between controlled and uncontrolled pediatric asthmatics. REFS detected a set of taxa, including *Haemophilus* and *Veillonella*, differentiating uncontrolled and controlled asthma with an average classification accuracy of 81% (saliva) and 86% (feces). These taxa showed enrichment in taxa previously associated with inflammatory diseases for both sampling compartments, and with COPD for the saliva samples.

Conclusion: Controlled and uncontrolled children with asthma can be differentiated based on their gastrointestinal microbiome using machine learning, specifically REFS. Our results show an association between asthma control and the gastrointestinal microbiome. This suggests that the gastrointestinal microbiome may be a potential biomarker for treatment responsiveness and thereby help to improve asthma control in children.

KEYWORDS
asthma: disease management, asthma: treatment

1 | INTRODUCTION

Asthma is a common chronic airway disease that is characterized by nonspecific symptoms including shortness of breath, wheezing, chest tightness, and coughing, often occurring episodically.[1,2] Treatment of asthma comprises two aims: prevention of episodes and reducing symptoms during such episodes.[2] However, due to the heterogeneity of the disease, selecting the correct medication to treat asthma is difficult and often based on a trial and error approach in accordance with the Global Initiative for Asthma (GINA) guidelines.[2,3] Despite regular medication intake, some patients still experience severe symptoms.[4] In children, this can lead to a reduced quality of life and a decrease in lung function later in life.[5] Therefore, the Systems Pharmacology approach to uncontrolled Pediatric Asthma (SysPharmPediA) consortium aims to identify biomarkers and molecular mechanisms that underlie asthma control using a multi-omics systems medicine approach.[6]

The human microbiome is influenced by many factors; however, it is relatively stable and remains identifiable to a person over the span of multiple years.[7] Earlier research has shown that the composition of the microbiome is associated with multiple diseases including cancer, diabetes, cardiovascular disease, and asthma.[8–11] In asthma, the domination of the airway microbiome with specific taxa was found to be associated with neutrophil count and lung function, leading to a more severe asthma phenotype.[12] On the contrary, the microbiome may have a protective role in asthma. Germ-free mice models showed stronger hypersensitivity and airway inflammation after ovalbumin induction compared to mice with a commensal gastrointestinal microbiome.[13] Clustering of asthmatic patients solely according to their airway microbiome allowed for clusters with distinctions in several clinical and demographic features, further proving their relation to asthma severity.[14]

Microbiome data at a species level is sparse in nature, meaning that many taxa are only detected in a handful of samples. Because of this, conventional differential abundance and normalization methods used in transcriptomics and proteomics often result in a loss of power when used to analyze microbiome data.[15,16] Specialized methods for microbiome data have been created with varying rates of success.[15,16] However, these specialized methods fail to capture the interaction between microbes and perform their analysis per taxon, while these interactions have been shown repeatedly in previous

> **Key Message**
>
> Controlled and uncontrolled asthmatics can be distinguished based on their gastrointestinal microbiome using the machine learning technique, Recursive Ensemble Feature Selection, demonstrating an association between the microbiome and asthma control.

literature.[16] The decreasing cost of sequencing techniques has led to a general increase in sample sizes for microbiome research, allowing for the use of more complex machine learning techniques.[17] Machine learning has already been shown to successfully identify a set of taxa related to type 2 diabetes and future glucose increments.[18] However, machine learning techniques have hardly been employed in asthma microbiome research.

In this study, we performed an analysis on the salivary and fecal microbiome data of the SysPharmPediA study to find differences in the microbial composition of the gastrointestinal tract of controlled and uncontrolled pediatric asthmatics. This was performed using global diversity measures, a conventional differential abundance analysis technique, and an ensemble machine learning technique, Recursive Ensemble Feature Selection (REFS). Specifically, we aimed to discover a set of potential biomarkers that can predict treatment responsiveness to improve asthma control in children and show the benefits of utilizing machine learning compared with that of conventional techniques.

2 | METHODS

2.1 | Study design

SysPharmPediA is a multicenter, prospective, European study with the objective of identifying a set of biomarkers to classify phenotypes of pediatric uncontrolled asthma.[6] Asthmatic children were included in four centers located in the Netherlands, Germany, Spain, and Slovenia. Written informed consent was provided by either the parents/caretakers and/or the study participants themselves. Each study center obtained approval from the local ethics committee. More details can be found in.[6] This study is registered at ClinicalTrials.gov under the identifier NCT04865575.

2.2 | Participants

One hundred and forty-three out of 145 participants in the SysPharmPediA cohort provided either a saliva and/or feces sample. The inclusion criteria for the study population were as follows: (1) aged between 6 and 17 years, (2) diagnosed with asthma by a doctor, and (3) under treatment with at least step 3 according to the GINA guidelines.[2] The participants were classified as controlled or uncontrolled asthmatics based on their clinical manifestations. Controlled asthmatics did not have any severe exacerbations requiring oral cortical steroids (OCS), emergency room visits, or hospitalizations in the prior 12 months and a score >19 on the (childhood) asthma control test (ACT/cACT).[19,20] Uncontrolled asthmatics had at least 1 exacerbation requiring OCS, emergency room visits, or hospitalizations in the prior 12 months and a score ≤19 on the (c)ACT. In addition, patients treated under step 2 of the GINA guidelines that were hospitalized for a severe asthma exacerbation could also be included in the uncontrolled asthmatics.

2.3 | Sample preparation, sequencing, and read processing

143 saliva samples and 103 feces samples were collected from the study population. A detailed description of the sample preparation, sequencing, and read processing can be found in Appendix S1, Section 1. In short, V3-V4 hypervariable regions of the 16S rRNA gene were amplified and sequenced using the MiSeq v3 2x300 bp (Illumina). Reads were cleaned and grouped into amplicon sequence variants (ASVs). Finally, ASVs were assigned taxonomies using the Silva database.[21]

2.4 | Statistical analysis

The composition of the bacterial microbiome for uncontrolled and controlled children with asthma was compared using three levels of complexity (Appendix S1, Section 3). First, samples as a whole were compared using global diversity measures (alpha/beta diversity). Second, individual taxa were compared using a conventional microbiome analysis technique (ANCOM-BC). Finally, we used an advanced machine learning method, REFS to incorporate interactions between taxa in the models. If applicable, a significance threshold of .05 was applied after adjusting for multiple testing using the Benjamini–Hochberg correction.

3 | RESULTS

The baseline characteristics of the included patients are summarized in Table 1. Included patients had a median age of 11.93 years (IQR = 9.65–14.00), with a median age of 12.00 (IQR = 9.72–14.00) and 11.74 (IQR = 9.65–13.84) for uncontrolled and controlled asthmatics, respectively. The patients were predominantly male (59.4% of all patients, 57.3% and 63.0% in uncontrolled and controlled asthmatics) and predominantly European; however, the uncontrolled asthmatics contained a higher percentage of non-European participants. In addition, the center of inclusion also showed an imbalance between controlled and uncontrolled asthmatics. The median (childhood) Asthma Control Test ((c)ACT) score of the participants is 23.0 (IQR = 19.0–25.0), with a median score of 20.5 (IQR = 17.0–23.0) for uncontrolled asthmatics compared with 24.5 (IQR = 23.0–25.0) for controlled asthmatics. Finally, uncontrolled asthmatics showed a higher percentage of omalizumab medication use.

3.1 | Global diversity measures show no differences between controlled and uncontrolled asthmatics

The bioinformatics pipeline showed that negative control samples had no to minimal read counts, and bacterial taxa were correctly identified at the genus level with 100% accuracy in the mock

TABLE 1 Baseline characteristics of the SysPharmPediA cohort from which either a saliva or feces sample was obtained, classified into uncontrolled and controlled asthmatics.

	Total (N = 143)	Uncontrolled asthmatics (N = 89)	Controlled asthmatics (N = 54)	p-Value
Age in years				
Median (IQR)	11.9 (9.7, 14.0)	12.0 (9.7, 14.0)	11.7 (9.7, 13.8)	.744
Min	6.0	6.0	6.4	
Max	17.4	17.3	17.4	
Gender				
Male	85 (59.4%)	51 (57.3%)	34 (63.0%)	.622
Female	58 (40.6%)	38 (42.7%)	20 (37.0%)	
Population group				
European	107 (74.8%)	59 (66.3%)	48 (88.9%)	.018
Latino	10 (7.0%)	7 (7.9%)	3 (5.6%)	
African	7 (4.9%)	7 (7.9%)	0 (0.0%)	
Asian	2 (1.4%)	1 (1.1%)	1 (1.9%)	
Mixed/other	11 (7.7%)	11 (12.4%)	0 (0.0%)	
Living environment				
City	57/138 (41.3%)	42/85 (49.4%)	15/53 (28.3%)	.072
City center	10/138 (7.2%)	7/85 (8.2%)	3/53 (5.7%)	
Rural Area	18/138 (13.0%)	11/85 (12.9%)	7/53 (13.2%)	
Village	48/138 (34.8%)	23/85 (27.1%)	25/53 (47.2%)	
Village/rural area	5/138 (3.6%)	2/85 (2.4%)	3/53 (5.7%)	
Country of inclusion				
Germany	38 (26.6%)	19 (21.3%)	19 (35.2%)	.014
Netherlands	32 (22.4%)	25 (28.1%)	7 (13.0%)	
Slovenia	23 (16.1%)	10 (11.2%)	13 (24.1%)	
Spain	50 (35.0%)	35 (39.3%)	15 (27.8%)	
Start of life				
Vaginal birth	111/138 (80.4%)	70/85 (82.4%)	41/53 (77.4%)	.618
C-section	27/138 (19.6%)	15/85 (17.6%)	12/53 (22.6%)	
Breastfeeding	101/138 (73.2%)	60/85 (70.6%)	41/53 (77.4%)	.499
Breastfeeding duration (months)	n = 99 7.0 (4.0, 12.0)	n = 60 7.0 (4.0, 12.0)	n = 39 6.0 (4.0, 11.0)	.341
(Childhood) Asthma control test				
Median (IQR)	n = 138 23.0 (19.2, 25.0)	n = 86 20.5 (17.2, 23.0)	n = 52 24.5 (23.0, 25.0)	<.001
Allergy characteristics				
Atopic sensitization	120/136 (88.2%)	75/83 (90.4%)	45/53 (84.9%)	.490
Total IgE abnormal	98/114 (86.0%)	68/76 (89.5%)	30/38 (78.9%)	.215
Food allergy	29/128 (22.7%)	16/77 (20.8%)	13/51 (25.5%)	.683
Aeroallergens	120/138 (87.0%)	75/85 (88.2%)	45/53 (84.9%)	.760
Atopic dermatitis	53/131 (40.5%)	36/84 (42.9%)	17/47 (36.2%)	.574
Allergic rhinitis	101/135 (74.8%)	62/83 (74.7%)	39/52 (75.0%)	1.000
Eosinophil percent	n = 125 5.5 (3.4, 8.5)	n = 79 6.0 (3.8, 9.4)	n = 46 5.0 (2.8, 6.7)	.040
Current asthma medication				
ICS	143 (100.0%)	89 (100.0%)	54 (100.0%)	1.000
SABA	133 (93.0%)	84 (94.4%)	49 (90.7%)	.624
LABA	135 (94.4%)	83 (93.3%)	52 (96.3%)	.696

TABLE 1 (Continued)

	Total (N = 143)	Uncontrolled asthmatics (N = 89)	Controlled asthmatics (N = 54)	p-Value
OCS	3 (2.1%)	3 (3.4%)	0 (0.0%)	.446
LTRA	25 (17.5%)	18 (20.2%)	7 (13.0%)	.378
Omalizumab	14 (9.8%)	13 (14.6%)	1 (1.9%)	.028
Mepolizumab	2 (1.4%)	2 (2.2%)	0 (0.0%)	.708
Antibiotics use				
Prenatal	6/126 (4.8%)	2/76 (2.6%)	4/50 (8.0%)	.174
From 0 to 2 years old	53/138 (38.4%)	32/84 (38.1%)	21 (38.9%)	.186
Last 2 months	19/137 (13.9%)	14/83 (16.9%)	5 (9.3%)	.314
(Secondhand) smoking				
Prenatal	34/114 (29.8%)	19/67 (28.4%)	15/47 (31.9%)	.841
From 0 to 2 years old	30/109 (27.5%)	15/64 (23.4%)	15/45 (33.3%)	.234
Present	33/121 (27.3%)	18/72 (25.0%)	15/49 (30.6%)	.637
Spirometry % predicted, median (IQR)				
FEV1 pre-salbutamol	n = 140 94.1 (82.8, 103.3)	n = 87 95.4 (82.3, 103.3)	n = 53 92.6 (86.1, 103.3)	.887
FEV1 post-salbutamol	n = 138 99.7 (89.7, 109.1)	n = 86 100.4 (92.1, 108.1)	n = 52 97.6 (89.4, 109.4)	.350
FEV1/FVC pre-salbutamol	n = 140 95.6 (87.2, 100.3)	n = 87 94.0 (85.9, 99.1)	n = 53 97.2 (89.2, 102.8)	.051
FEV1/FVC post-salbutamol	n = 138 99.3 (93.1, 103.6)	n = 86 98.9 (90.8, 103.9)	n = 52 99.9 (94.6, 103.5)	.587

Note: p-Values were calculated with a Wilcoxon rank-sum test for numerical variables and a χ^2 test for categorical variables.
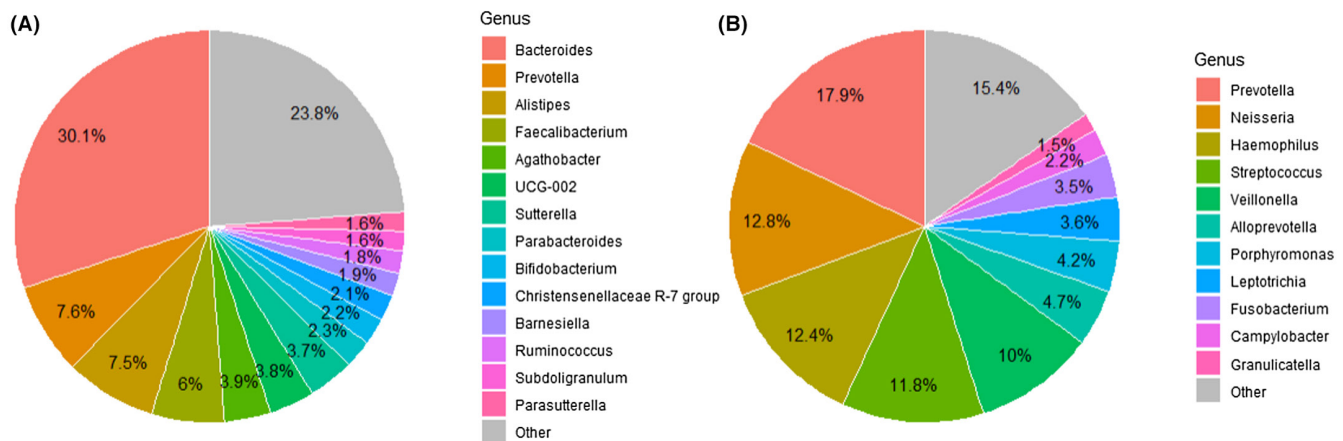

FIGURE 1 Mean relative abundance of the most abundant genera in the feces (A) and saliva (B) samples. Genera with a mean relative abundance of at least 1.5% in their compartment are shown individually. All other genera are grouped under "Other."

communities. A total of 5088 ASVs were identified including 2507 ASVs and 2684 ASVs in the feces and saliva samples, respectively. *Bacteroides*, *Prevotella*, and *Alistipes* were the highest abundant bacterial genera in the feces samples, and *Prevotella*, *Neisseria*, and *Haemophilus* were the highest abundant genera in the saliva samples (Figure 1). For the alpha diversity, there were no statistically significant differences between the controlled and uncontrolled asthmatics, based on the number of observed unique ASV per sample (feces: $p = .17$, saliva: $p = .061$) or the Shannon index (feces: $p = .51$, saliva: $p = .12$) (Figure S1). The beta diversity did also not find any

significant differences between the uncontrolled and controlled asthmatics (feces: $p = .586$, saliva: $p = .705$, Figure 2).

3.2 | Conventional differential abundance analysis shows no differentially abundant taxa

For the conventional differential abundance analysis, after considering ASVs only present in at least 5% of samples, 463 and 450 ASVs remained in the feces and saliva samples, respectively.

After correction for batch, age, sex, center of inclusion, living environment, and antibiotic use, the ANCOM-BC function revealed no statistically significant differentially abundant species between controlled and uncontrolled asthmatics (Figure 3). The volcano plots show that after correction for multiple testing, no taxa were close to reaching a significant adjusted *p*-value (red line, Figure 3).

3.3 | Taxa selected by REFS show enrichment in inflammation-related taxa

The same 5% criterion as for conventional differential abundance analysis was applied for REFS, with 2 datasets for both saliva and fecal samples. One including patients who used antibiotics in the last 2 months, and one with patients who used no antibiotics (NA)

in the last 2 months. For the feces samples, this resulted in 463 (NA: 451) ASVs being detected in 103 (NA: 87) patients. For the saliva samples, 450 (NA: 436) ASVs were detected in 143 (NA: 124) patients.

Through iterative removal of the least predictive taxa, the REFS machine learning algorithm selected a number of ASVs (features) for each of the datasets based on the highest accuracy over 10 separate runs (Figure S2). The highest discrimination accuracy by the REFS algorithm was 85.8% (NA: 86.6) for the feces samples and 80.9% (NA: 84.1%) for the saliva samples. Areas under the curve (AUC) for the individual feature reduction methods are shown in Table S1. These are considered as "very good".[22] The highest accuracies were achieved by using 30 (NA: 29) and 37 (NA: 56) taxa for the feces and saliva samples, respectively. These identified taxa include genera such as *Haemophilus*, *Veillonella*, and *Rothia*. However, there is only moderate overlap between the taxa found in the samples with
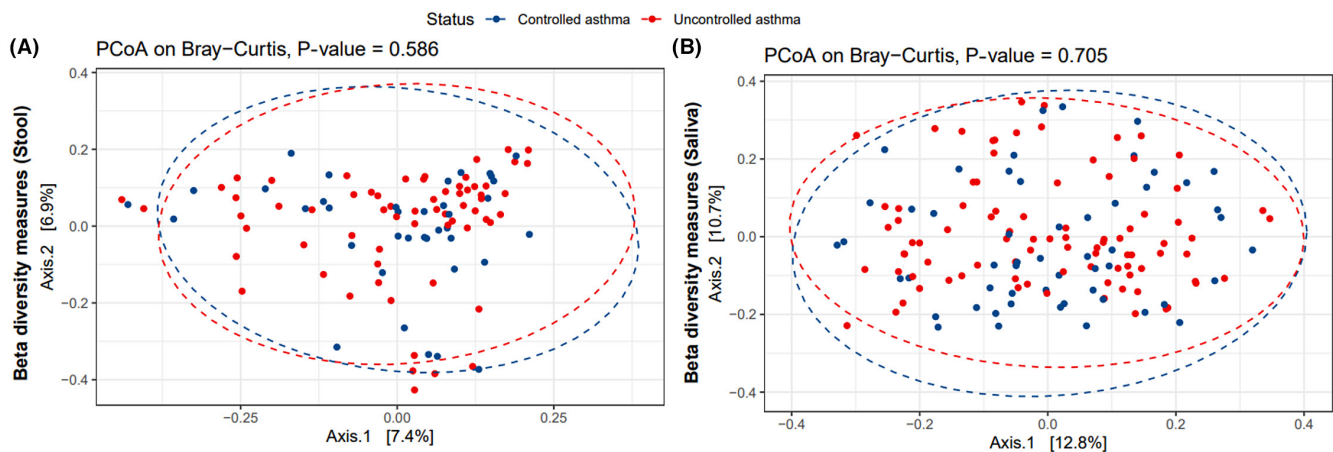

FIGURE 2 Beta diversity measure comparing the controlled and uncontrolled asthmatics for the feces (A) and saliva (B) samples using PCoA on the Bray–Curtis differences. For both sampling compartments, no statistically significant differences were found between the two groups.
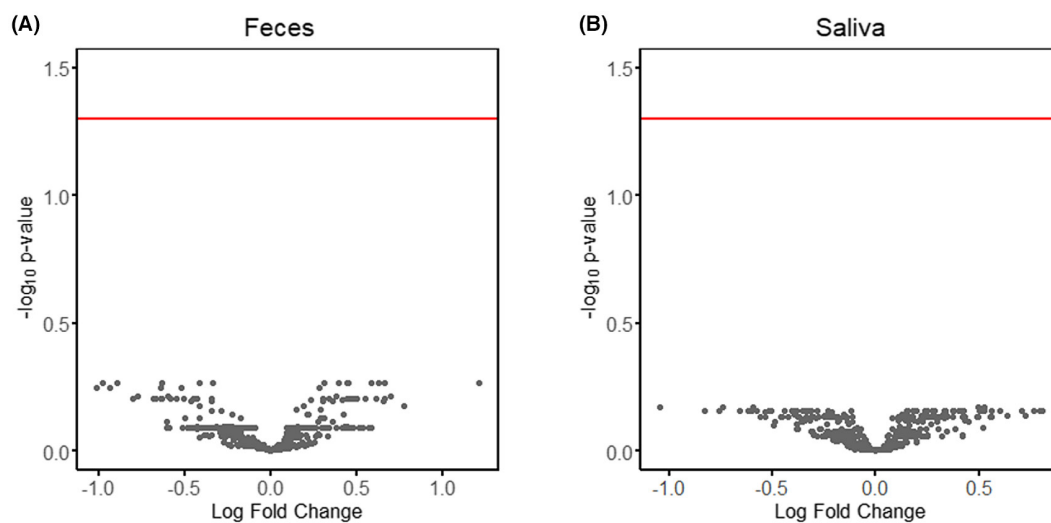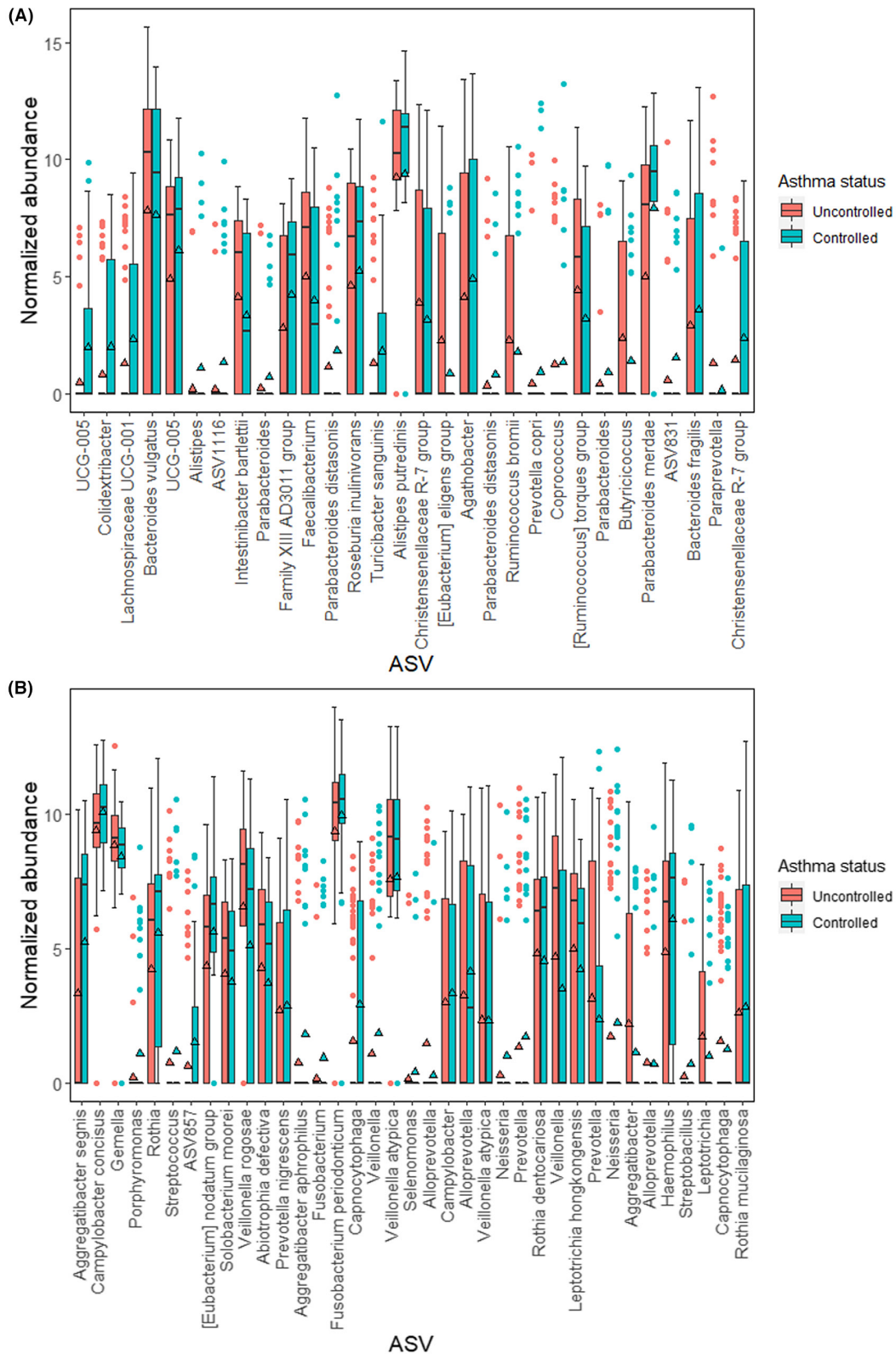

FIGURE 3 Volcano plots of the conventional differential abundance analysis using the ANCOM-BC function. The gray dots represent nonsignificant taxa, while the red line represents an adjusted *p*-value of .05, the significance threshold. No significantly different taxa were found in either the feces or saliva samples.

FIGURE 4 Boxplots showing the abundances of the features selected by the REFS machine learning algorithm for the feces (A) and saliva (B) samples. Counts were normalized using cumulative sum scaling for visual purposes only, with the actual model using a Z-score transformation. The black horizontal line depicts the median, the triangle indicates the average, and the colored points show outliers. The taxa are ordered according to their predictive value.

and without the patients that used antibiotics, 50% and 57% for the feces and saliva samples respectively. The normalized abundance of these selected taxa can be found in Figure 4 (including antibiotic users) and Figure S3 (nonantibiotic users).

Enrichment analysis was performed with the MicrobiomeAnalyst tool[23] with the genus and species names of the identified taxa. Tables S2 and S3 show all significant taxon sets from the feces and saliva samples respectively. Interestingly, several taxon sets related to (liver) cirrhosis are prevalent in both the feces and saliva samples. Other significantly enriched taxon sets linked to inflammation include mucositis, periodontitis, ulcerative colitis, and chronic obstructive pulmonary disease (COPD).

4 | DISCUSSION

In this study, we aimed to show differences in the gastrointestinal bacterial microbiome of controlled and uncontrolled asthmatics in the pediatric SysPharmPediA cohort.[6] We showed that diversity measures and conventional differential abundance analyses were unable to discriminate between controlled and uncontrolled asthmatics, while the machine learning technique REFS was able to find sets of taxa with predictive power for asthma control status. The taxa identified showed enrichment in taxa involved in cirrhosis for both the feces and saliva samples and taxa involved in inflammatory diseases.

Among the predictive taxa in the saliva samples were the *Haemophilus* and *Streptococcus* genera. Earlier research has shown that the colonization of these genera in asthma is associated with longer disease duration, poorer lung function, and higher neutrophil counts.[12] However, in our dataset, the abundance of these genera is similar or slightly increased in the controlled asthmatics. Species of the *Veillonella* genus showed both increased and decreased abundances among the predictive taxa. This potentially reflects the conflicting conclusions about this genus in literature. Earlier research has shown that *Veillonella* is more abundant in a less severe asthma phenotype.[14] On the contrary, higher *Veillonella* abundance in children in the airway has been associated with an increased duration of asthma episodes.[24] The *Rothia* genus was also identified in the saliva samples, with an increased abundance in the controlled asthmatics. *Rothia* was previously found to have an anti-inflammatory effect by inhibiting the activation of the NF-κB pathway,[25] a major player in inflammation and asthma pathophysiology.[26]

Hu et al. identified 3 genera associated with allergic rhinitis in the fecal microbiome of children at 10 years old. Of these genera, one (*Agathobacter*) was associated with higher odds of having allergic rhinitis, and two (UCG-005 and *Christensenellaceae* R-7 group) were associated with a decrease in the odds of having allergic rhinitis.[27] All these genera were identified to have predictive value in our study, with two species of UCG-005 being the best and fifth-best predictors of asthma control in the feces samples.

The processing of the input by machine learning methods is often seen as a black box. It is difficult for humans to understand how certain methods make their classification and decisions, with fields attempting to uncover this black box growing bigger in recent years.[28,29] Interestingly, in our study, the machine learning ensemble selected several features that did not show any distinction on their own between controlled and uncontrolled asthmatics. This can be seen in two ASVs attributed to *Veillonella atypica* in the saliva samples. It is possible that the predictive power of these taxa lies in the interaction with other taxa, only detected by machine learning. For most of the appointed taxa, the differences between uncontrolled and controlled asthmatics are too small for conventional methods to detect. The strength of the machine learning methods and REFS lies in the detection of patterns and interactions that arise from changing asthma status.

The strength of this study lies in the multi-analytical approach. We approached the microbiome using diversity measures, conventional differential abundance analysis, and advanced machine learning, performing analysis on the microbiome obtained from multiple sources. We showed the differences in results between each approach. While the first two approaches were not able to detect differences, REFS identified taxa that were previously discovered to have been implicated in asthma severity or asthma mechanisms. Another strength of this study is the multinational European scope of the SysPharmPediA cohort. Including patients from four countries allows the results to be more generalizable over Europe compared to patient inclusions from only one country. This is of particular relevance, as the microbiome is influenced by environmental exposures and cultural differences such as dietary intake and lifestyle.

This study is, however, not without its limitations. Firstly, while the microbiome remains sufficiently stable overtime without changes in disease status,[7] evidence shows that changes in microbiome are already detectable before the onset of asthma symptoms. As a result, the microbiome of some children in the controlled group could have similar microbial composition to uncontrolled asthmatics, if they might develop exacerbation. Secondly, for machine learning techniques the number of samples is considered relatively low.[30] This puts the results at risk of overfitting. We attempted to prevent overfitting as much as possible by applying an ensemble of classifier methods, 10-fold cross-validation, and using different classifier methods for training and testing. This might not, however, have led to full removal of overfitting, as the overlap between taxa before and after removal of the patients with antibiotics was only around 55%. Therefore, validating the results in external cohorts is needed to reach definitive conclusions.

In conclusion, we have shown that the gastrointestinal microbiome can be used to discriminate between controlled and uncontrolled asthmatics using machine learning. This suggests that machine learning techniques can provide complementary insights into the link between the community of gastrointestinal bacteria and asthma control, where conventional techniques fail. This study suggests that the gastrointestinal microbiome can be a potential biomarker for treatment responsiveness and thereby help to improve asthma control in children. However, refining of the taxa sets and careful validation are needed before this can be applied in clinical diagnostics and treatment.

CONFLICT OF INTEREST STATEMENT

J.M.B., A.L.-R., S.H. (Simone Hashimoto), O.S.P., P.C.E., J.K.M., C.B., A.F., C.W., S.B., A.A.T, P.K., and A.D.K. have no conflicts of interest to disclose. A.H.N. and S.J.H.V. were funded by a grant paid to the institution from ERANET Systems Medicine and ZonMW [project number: 9003035001]. M.G. and U.P. were funded by a SysPharmPediA grant, co-financed by the Ministry of Education, Science and Sport Slovenia (MIZS) (contract number C3330-16-500106), and funded by Slovenian Research Agency (research core funding No. P3-0067). M.P.-Y. was funded by Instituto de Salud Carlos III (ISCIII) through AES and EC within AAL framework, and the SysPharmPedia grant from the ERACoSysMed 1st Joint Transnational Call from the European Union under the Horizon 2020 (AC15/00015), and by the Ramón y Cajal Program [RYC-2015-17205] by MCIN/AEI/10.13039/501100011033 and by the European Social Fund "ESF Investing in your future," and by the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033 and the European Regional Development Fund "ERDF A way of making Europe" by the European Union grant [SAF2017-83417R], and by MCIN/AEI/10.13039/501100011033 [grant PID2020-116274RB-I00] (outside of the submitted work), and by the Allergopharma Award 2021 (outside of the submitted work), and by ISCIII [CB06/06/1088] (outside of the submitted work), and received grant support from GlaxoSmithKline S.A. (outside of the submitted work). S.H. (Sussane Harner) received payment for lectures by Nutricia. M.K. was funded for this project by the German Ministry of Education and Research (BMFM) [grant 031 L0088] and previously received funding in the last 36 months from the European Union, BMFM, German Research Foundation, and Infectopharm, received consulting fees from Bionorica, Sanofi, Novartis, and Bencard, received payment for lectures from ERS, EAACI, ATS, Novartis, Glaxo, Chiesi, Sanofi, Nutricia, Hipp, and Allergopharma, and is part of a patent for "METHOD FOR TESTING A SUBJECT THOUGHT TO HAVE OR TO BE PREDISPOSED TO ASTHMA," under European patent application 5 EP07301135.5. M.I.A.-A. was funded by a full PhD scholarship from the Ministry of Higher Education of the Arab Republic of Egypt during the conduct of the study and received a grant from Stichting Astma Bestrijding. A.H.M.-v.d.Z. was funded by a grant paid to the institution from ERANET Systems Medicine and ZonMW [project number: 9003035001], received unrestricted research grants from Vertex and Boehringer Ingelheim, received consulting fees (paid to institution) from Astra Zeneca and Boehringer Ingelheim, and received honoraria (paid to institution) for lectures by GSK. A.H.M.-v.d.Z. is the PI of P4O2 (Precision Medicine for more Oxygen) public–private partnership sponsored by Health Holland involving many private partners who contribute in cash and/or in kind. Partners in the Precision Medicine for more Oxygen (P4O2) consortium are the Amsterdam UMC, Leiden University Medical Center, Maastricht UMC+, Maastricht University, UMC Groningen, UMC Utrecht, Utrecht University, TNO, Aparito, Boehringer Ingelheim, Breathomix, Clear, Danone Nutricia Research, Fluidda, MonitAir, Ncardia, Ortec Logiqcare, Philips, Proefdiervrij, Quantib-U, RespiQ, Roche, Smartfish, SODAQ, Thirona, TopMD, the Lung Alliance Netherlands (LAN), and the Lung Foundation Netherlands (Longfonds). The consortium is additionally funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health (LSHM20104; LSHM20068), to stimulate public–private partnerships and by Novartis. A.H.M.-v.d.Z. is the (unpaid) chair of the DSMB SOS BPD study, an advisory board member of the CHAMP study, president of the federation of innovative drug research in the Netherlands (FIGON) and the European Association of systems medicine (EASYM). The funders had no role in the design, collection, analyses, or interpretation of the study. Neither were they involved in the writing of the manuscript or the decision to publish the results.

DATA AVAILABILITY STATEMENT

The 16S rRNA microbiome generated in this study is available on the BioProject database under BioProject ID PRJNA867125 (http://www.ncbi.nlm.nih.gov/bioproject/867125). This will become available upon publication. Additional data, such as clinical characteristics, from the SysPharmPediA study can be made available upon specific requests subject to the requestor obtaining ethical, research, data access, and collaboration approvals from the SysPharmPediA study management board. Requests can be sent to a.h.maitland@amsterdamumc.nl.

ORCID

Susanne J. H. Vijverberg https://orcid.org/0000-0002-4579-4081
Antoaneta A. Toncheva https://orcid.org/0000-0001-5019-5174
Susanne Harner https://orcid.org/0000-0001-6733-5823
Michael Kabesch https://orcid.org/0000-0003-0697-1871

REFERENCES

1. Papi A, Brightling C, Pedersen SE, Reddel HK. Asthma. *Lancet*. 2018;391(10122):783-800.
2. Mauer Y, Taliercio RM. Managing adult asthma: the 2019 GINA guidelines. *Cleve Clin J Med*. 2020;87(9):569-575.
3. Falk NP, Hughes SW, Rodgers BC. Medications for chronic asthma. *Am Fam Physician*. 2016;94(6):454-462.
4. Nordlund B, Melen E, Schultz ES, et al. Prevalence of severe childhood asthma according to the WHO. *Respir Med*. 2014;108(8):1234-1237.
5. Pijnenburg MW, Fleming L. Advances in understanding and reducing the burden of severe asthma in children. *Lancet Respir Med*. 2020;8(10):1032-1044.
6. Abdel-Aziz MI, Neerincx AH, Vijverberg SJH, et al. A system pharmacology multi-omics approach toward uncontrolled pediatric asthma. *J Pers Med*. 2021;11(6):484.
7. Chen L, Wang D, Garmaeva S, et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell*. 2021;184(9):2302-2315.e12.
8. Garrett WS. Cancer and the microbiota. *Science*. 2015;348(6230):80-86.
9. Karlsson FH, Tremaroli V, Nookaew I, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*. 2013;498(7452):99-103.
10. Karlsson FH, Fak F, Nookaew I, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun*. 2012;3:1245.
11. Huang YJ, Boushey HA. The microbiome in asthma. *J Allergy Clin Immunol*. 2015;135(1):25-30.
12. Green BJ, Wiriyachaiporn S, Grainge C, et al. Potentially pathogenic airway bacteria and neutrophilic inflammation in treatment resistant severe asthma. *PLoS One*. 2014;9(6):e100645.
13. Herbst T, Sichelstiel A, Schar C, et al. Dysregulation of allergic airway inflammation in the absence of microbial colonization. *Am J Respir Crit Care Med*. 2011;184(2):198-205.
14. Abdel-Aziz MI, Brinkman P, Vijverberg SJH, et al. Sputum microbiome profiles identify severe asthma phenotypes of relative stability at 12 to 18 months. *J Allergy Clin Immunol*. 2021;147(1):123-134.
15. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes*. 2020;6(1):60.
16. Weiss S, Xu ZZ, Peddada S, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*. 2017;5(1):27.
17. Namkung J. Machine learning methods for microbiome studies. *J Microbiol*. 2020;58(3):206-216.
18. Gou W, Ling CW, He Y, et al. Interpretable machine learning framework reveals robust gut microbiome features associated with type 2 diabetes. *Diabetes Care*. 2021;44(2):358-366.
19. Nathan RA, Sorkness CA, Kosinski M, et al. Development of the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol*. 2004;113(1):59-65.
20. Liu AH, Zeiger R, Sorkness C, et al. Development and cross-sectional validation of the childhood asthma control test. *J Allergy Clin Immunol*. 2007;119(4):817-825.
21. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590-D596.
22. Simundic AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC*. 2009;19(4):203-211.
23. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat Protoc*. 2020;15(3):799-821.
24. Thorsen J, Stokholm J, Rasmussen MA, et al. The airway microbiota modulates effect of azithromycin treatment for episodes of recurrent asthma-like symptoms in preschool children: a randomized clinical trial. *Am J Respir Crit Care Med*. 2021;204(2):149-158.
25. Rigauts C, Aizawa J, Taylor SL, et al. Rothia mucilaginosa is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease. *Eur Respir J*. 2022;59(5):2101293.
26. Mishra V, Banga J, Silveyra P. Oxidative stress and cellular pathways of asthma and inflammation: therapeutic strategies and pharmacological targets. *Pharmacol Ther*. 2018;181:169-182.
27. Hu C, van Meel ER, Medina-Gomez C, et al. A population-based study on associations of stool microbiota with atopic diseases in school-age children. *J Allergy Clin Immunol*. 2021;148(2):612-620.
28. Azodi CB, Tang J, Shiu SH. Opening the black box: interpretable machine learning for geneticists. *Trends Genet*. 2020;36(6):442-455.
29. Bodini M, Rivolta MW, Sassi R. Opening the black box: interpretability of machine learning algorithms in electrocardiography. *Philos Trans A Math Phys Eng Sci*. 2021;379(2212):20200253.
30. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One*. 2019;14(11):e0224365.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.