

Supplementary Material

Supplementary Methods

Study design and procedures

The study design and procedures have been described in detail previously [1, 2]. For easier access we included this information here:

The ALLIANCE cohort is recruited at five pediatric centers (Cologne, Hannover, Lubeck, Marburg, Munich) and two adult centers (LungenClinic Grosshansdorf, Research Centre Borstel), all affiliated with the German Center for Lung Research (DZL). Recruitment started in 2013. Participants with preschool wheeze and asthma had annual study visits while healthy controls were assessed only once. Visits were postponed in case of upper respiratory tract infections, asthma exacerbations (adults) or fever $>38.5^{\circ}\text{C}$ in the past two weeks (children).

Caregivers and adult participants completed questionnaires covering demographic factors, early-life exposures, respiratory symptoms, medical history and medication use. Recorded variables included age, sex, BMI, mode of delivery, breastfeeding, siblings, pet contact (cat/dog), childcare, wheeze and cough during the previous 12 months, previous RSV infection, physician-diagnosed asthma, hay fever, eczema or food allergy, and use of hypoallergenic formula, antibiotics for recurrent cough, or inhaled/systemic corticosteroids. Definitions of clinical and lifestyle variables have been described previously [1]. Questionnaire-based outcomes were defined as follows: wheeze episode as salbutamol use due to wheeze on >2 of 7 days; exacerbation as a wheeze episode requiring systemic corticosteroids or hospitalization; and asthma control according to GINA guidelines.

Differential blood counts and total IgE were measured in routine on-site hospital laboratories. Allergen-specific IgE was assessed centrally using Euroline™ (Euroimmun, Germany) against a panel of common aeroallergens and food allergens [2]. Atopy was defined as ≥ 1 allergen-specific IgE ≥ 0.7 kU/L, based on previous findings from the ALLIANCE cohort [3]. Elevated blood eosinophil counts were defined as ≥ 470 cells/ μl , corresponding to the 90th percentile of healthy controls in the ALLIANCE cohort.

DNA extraction and library preparation

DNA of nasal swabs was extracted via freezing-heating cycles (5 min, 3 times) following ultrasonication (S220 Focused-ultrasonicator Covaris) with 6 cycles of 5 s at peak power=200/duty factor=2, 30 s peak power=275/duty factor=5 at 6°C [4]. DNA of stool samples was extracted from 0.4 g feces material using the phenol-chloroform-based protocol described by Lueders et al. [5] and the Precellys24 Instrument (PeqLab, Erlangen, Germany).

Amplicon sequencing of the V4 hypervariable region of the 16S rRNA gene was performed on a MiSeq Illumina instrument (MiSeq Reagent Kit v3 (600 Cycle); Illumina, San Diego, CA, USA) using the universal eubacterial primers 515F [6] and 806R [7]. PCR was done using NEBNext high fidelity polymerase (New England Biolabs, Ipswich, USA) in a total volume of 25 µl (10 ng DNA template (nose) / 5 ng DNA template (stool), 12.5 µl polymerase, 5 pmol of each primer, 2.5 µl 3% BSA) and the following PCR conditions: 5 min at 98°C; 30 cycles (nose) / 25 cycles (stool) of 10 s at 98°C, 30 s at 55 °C, 30 s at 72 °C; 5 min 72 °C. To identify potential contaminants deriving from DNA extraction and library preparation, both extraction and PCR no template control samples were processed (each 4). PCR products were purified using MagSi NGSprep Plus beads (Steinbrenner, Wiesenbach, Germany) and quantified via PicoGreen assay. Subsequently, indexing PCR was performed using the Nextera XT Index Kit v2 (Illumina, Inc. San Diego, CA, US) in a total volume of 25 µl (10 ng DNA template, 12.5 µl NEBNext high fidelity polymerase, 2.5 µl of each indexing primer) and the following PCR conditions: 30s at 98 °C; 8 cycles of 10 s at 98 °C, 30s at 55 °C, 30s at 72 °C; 5min 72 °C. Indexing PCR products were purified using MagSi NGSprep Plus beads, qualified and quantified via a Fragment Analyzer™ instrument (Advanced Analytical Technologies, Inc., Ankeny, USA) and pooled in an equimolar ratio of 4nM.

Sequence data generated in this study are available upon reasonable request from the authors.

Sequence processing

FASTQ files were trimmed with a minimum read length of 50 using Cutadapt [8] and quality control was performed via FastQC [9]. For subsequent data analysis, the DADA2 pipeline v 1.30 [10] was used with the following trimming and filtering parameters: 20 bp were removed n-terminally and reads were truncated at position 240 (forward) and 200 (reverse), respectively, with expected error of 3 (forward) and 4 (reverse). Taxonomic analysis was performed using SILVA v138.1. Reads were excluded if classified as mitochondria or chloroplast or if the phylum was missing. All blank samples were analyzed together with biological samples but received only 0-4 sequencing reads, thus all sample reads were

kept for subsequent analysis, resulting in a total amount of 3,744,773 reads (corresponding to an average of 72,014 reads per sample) assigned to 2,177 amplicon sequence variants (ASV).

Statistical analysis

All plots and statistics were performed in R version 4.5.1 (<https://www.R-project.org>). Prior to the main analysis, potential covariates (age, gender, cough antibiotics, siblings, daycare, and pet contact) were tested individually for association with microbiome outcomes due to the modest sample size (n = 25). Covariates improving model fit, as judged by metrics like delta AIC, beta coefficients and CI for mixed-effect models or explained variance for PERMANOVA, were included in the final analysis.

Alpha diversity was calculated using species richness, evenness as well as Shannon diversity index. Associations with covariates were assessed using generalized linear mixed-effects models (R package glmmTMB), including study subjects as a random effect. Covariates improving model fit in pre-tests (age and gender) were retained. Analyses were performed on TMM-normalized data [11] and raw counts, with log-transformed sequencing depth as covariate for raw counts when it was significantly correlated with the response variable (assessed using cor.test). Beta diversity was analyzed using unweighted and weighted UniFrac distances on raw counts, as well as Aitchison distance on CLR-transformed data. Associations with covariates were tested via PERMANOVA, with permutations restricted within study subjects using the strata option (R package vegan). Based on pre-tests, age and gender was included as covariate. For both alpha and beta diversity analysis, p value adjustment for multiple comparisons was performed with Benjamini-Hochberg correction.

To identify microbial taxa differing between healthy and wheezing children, ANCOM-BC2 (Analysis of Compositions of Microbiomes with Bias Correction 2) was used [12], with age and gender as covariates. Phylogenetic trees for the ASV of interest were constructed including SILVA-curated reference sequences (<https://www.arb-silva.de/>). Sequences were aligned with ClustalW following tree optimization via the Maximum Likelihood method (GTR model) and 1000 bootstrap iterations for reliability (R packages ape, phangorn, msa).

For analysis of within-group variation, the average weighted UniFrac distance between each subject and the group centroid was calculated (beta-dispersion). Statistical analysis was performed like for alpha diversity via generalized linear mixed-effects models (R package glmmTMB) and Benjamini-Hochberg p value correction for multiple comparisons, including study subjects as a random effect and age and

gender as covariates. In addition, Levene's Test was used to test for equality of variances within the groups.

To search for associations between individual variability (beta-dispersion), demographic/lifestyle variables, hematological immune markers (hemoglobin, thrombo-, leuko-, lymphocytes, neutro-, eosino-, basophils, immunoglobulin E), and microbial taxa, correlation analyses were performed, with Pearson correlations for numerical variables and Spearman correlations for the comparison of numeric and microbiome data after CLR-transformation. P value adjustment for multiple comparisons was performed with Benjamini-Hochberg correction.

The core microbiome per group was defined as ASV or genera present in at least 80 % of subjects without setting an additional abundance cutoff.

To explore heterogeneity among wheezing children, samples were classified as healthy-like or community-divergent based on their microbiome's distance to the centroid of healthy samples. First, Bray-Curtis distances were calculated on TSS-normalized data, and then Euclidean distances were computed from each sample to the healthy centroid in the ordination space. Wheezing samples within the 90th percentile of the healthy centroid distances were considered healthy-like, while those exceeding this threshold were classified as community-divergent. This resulted in body site-specific subgroup sizes of 9 and 8 for nasal samples and 10 and 7 for stool samples (healthy-like and community-divergent, respectively). Selected variables were fitted to the ordination space as vectors using the envfit function (R package vegan), with significance assessed via permutation tests. Further statistical analysis between those subgroups were performed as described above using generalized linear mixed-effects models (R package glmmTMB), ANCOM-BC2 and correlation analyses.

Microbial community assembly was assessed using the β -nearest taxon index (β NTI) and Raup-Crick-based Bray-Curtis (RCbray) approach following Stegen et al. [13]. Although β NTI was originally developed for macroecological systems, it is widely applied in microbial ecology because it quantifies phylogenetic turnover relative to null expectations, enabling inference of deterministic versus stochastic processes. Phylogenetic relatedness serves as a proxy for ecological similarity due to trait conservation in microbes. This framework was applied to evaluate how deterministic and stochastic processes shape microbial communities across healthy and wheezing children, where wheezing may impose strong selective pressures. For β NTI, thresholds of ± 2 were used to define deterministic selection, further categorized as heterogeneous/variable selection (β NTI > +2; divergent community configurations) and homogeneous selection (β NTI < -2; convergence toward similar composition). For pairwise

comparisons with $|\beta_{NTI}| < 2$, RCbray distinguished stochastic processes: RCbray $> +0.95$ indicated dispersal limitation, RCbray < -0.95 indicated homogenizing dispersal, and $|\text{RCbray}| \leq 0.95$ indicated undominated processes. Analyses were performed in R using the *picante* and *vegan* packages with null distributions generated from 999 randomizations. A phylogenetic tree was constructed prior to these analyses by aligning sequences with MAFFT (FFT-NS algorithm) and building the tree with FastTree under the GTR+CAT substitution model. The *ape* package was used to handle phylogenetic distances via a *phyloseq* object.

Microbial co-occurrence networks were constructed using the R package NetCoMi v1.1.0 [14], applying the “signed” transformation to convert estimated associations into dissimilarities and including only ASV present in $\geq 50\%$ of samples per group. Networks were inferred with the SpiecEasi framework on CLR-transformed data, with zero replacement via multiplicative imputation and the Meinshausen-Bühlmann neighborhood selection method for network inference. Network topology was characterized by modular structure (fast-greedy clustering) and identification of hub taxa based on degree and eigenvector centrality (top 10%). Centrality measures (degree, betweenness, closeness, eigenvector) were normalized, and weighted node degrees were calculated to assess connectivity patterns across microbial communities.

To evaluate cross-compartment associations between nasal and stool microbiomes, raw counts were transformed to relative abundances and CLR-transformed. Spearman correlations between all nose-stool pairs were calculated using the *rcorr* function (R package Hmisc), with p-values adjusted via Benjamini-Hochberg. Significant correlations ($p < 0.05$) were visualized as a bipartite network (R packages *igraph*, *ggraph*), with nodes representing taxa and edges indicating significant associations. If not stated otherwise, all plots were created in R using *ggplot2* [15], *ggpubr* [16], and *eulerr* [17].

Supplementary References

1. Fuchs O, Bahmer T, Weckmann M, et al. The all age asthma cohort (ALLIANCE) - from early beginnings to chronic disease: a longitudinal cohort study. *BMC Pulmonary Medicine*. 2018;18(1):140.
2. Maison N, Omony J, Illi S, et al. T2-high asthma phenotypes across lifespan. *Eur Respir J*. 2022;60(3):2102288.
3. Skevaki C, Tafo P, Eiringhaus K, et al. Allergen extract- and component-based diagnostics in children of the ALLIANCE asthma cohort. *Clinical & Experimental Allergy*. 2021;51(10):1331-45.
4. Pust M-M, Wiehlmann L, Davenport C, et al. The human respiratory tract microbial community structures in healthy and cystic fibrosis infants. *npj Biofilms and Microbiomes*. 2020;6(1):61.

5. Lueders T, Manefield M, Friedrich MW. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ Microbiol.* 2004;6(1):73-8.
6. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol.* 2016;18(5):1403-14.
7. Apprill A, McNally S, Parsons R, et al. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol.* 2015;75.
8. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal.* 2011;17:10-2.
9. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2010.
10. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581-3.
11. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26(1):139-40.
12. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nat Methods.* 2024;21(1):83-91.
13. Stegen JC, Lin X, Konopka AE, et al. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 2012;6(9):1653-64.
14. Peschel S. NetCoMi: Network Construction and Comparison for Microbiome Data. <https://netcomide>. 2025;R package version 1.2.0.
15. Wickham H. ggplot2: Elegant Graphics for Data Analysis. <https://ggplot2tidyverse.org>. 2016.
16. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0. <https://rpkgsdatanoviacom/ggpubr/>. 2023.
17. Larsson JG, Peter. A Case Study in Fitting Area-Proportional Euler Diagrams with Ellipses using eulerr. Proceedings of International Workshop on Set Visualization and Reasoning (SetVR 2018) co-located with 10th International Conference on the Theory and Application of Diagrams (Diagrams 2018), Edinburgh, UK, June 18, 2018 (pp 84-91). 2018;CEUR-WS.org. <https://ceur-ws.org/Vol-2116/paper7.pdf>.

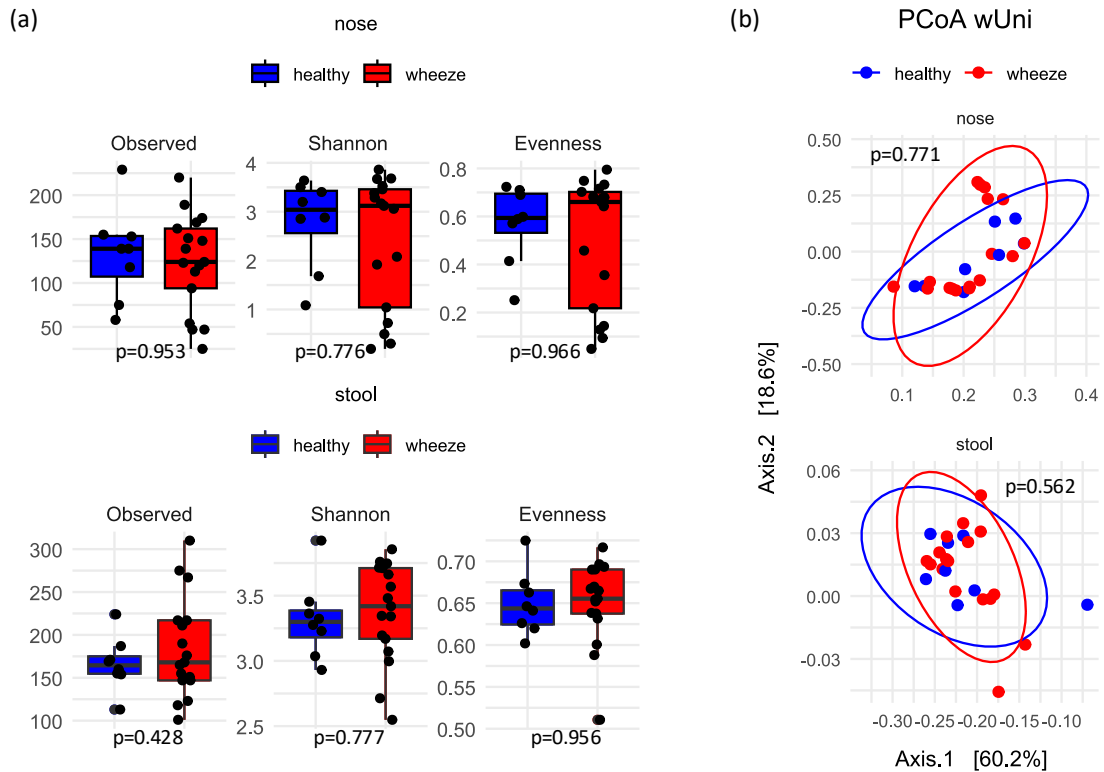


Figure S1. Microbial diversity and overall community composition in nose and stool of healthy (n=8) and wheezing (n=17) children.

(a) Boxplots of alpha diversity measures (observed ASV, Shannon index and evenness). Group differences were assessed using generalized linear mixed-effects models with Benjamini–Hochberg adjustment.

(b) Principal coordinates analysis (PCoA) of weighted UniFrac distances. Group differences were assessed using PERMANOVA with Benjamini–Hochberg correction.

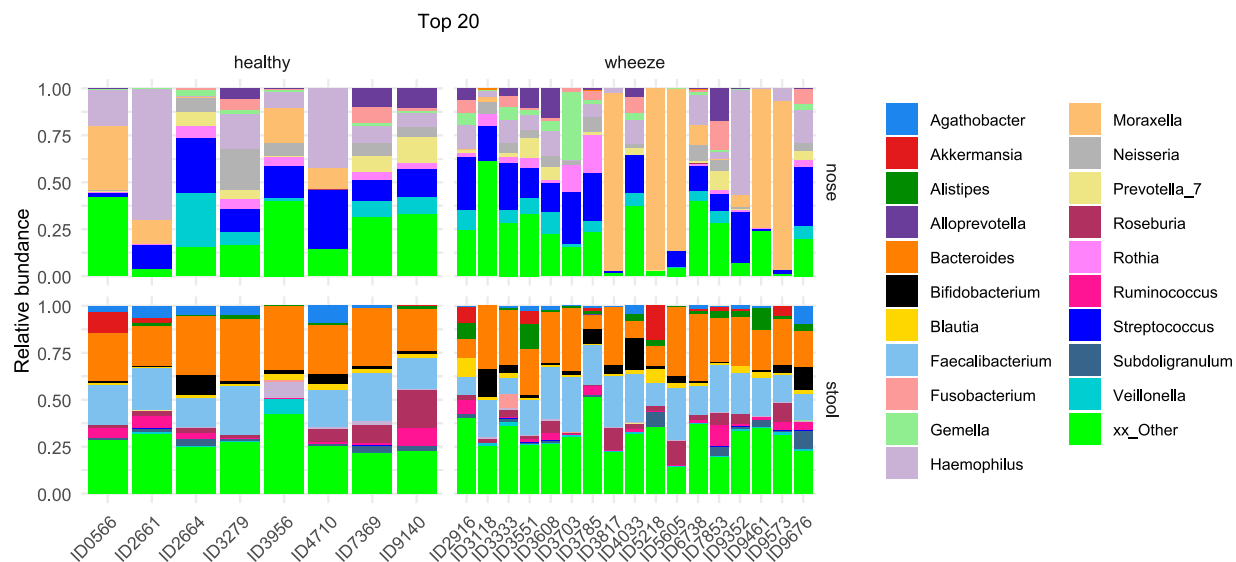


Figure S2. Relative abundance of the top 20 genera per subject

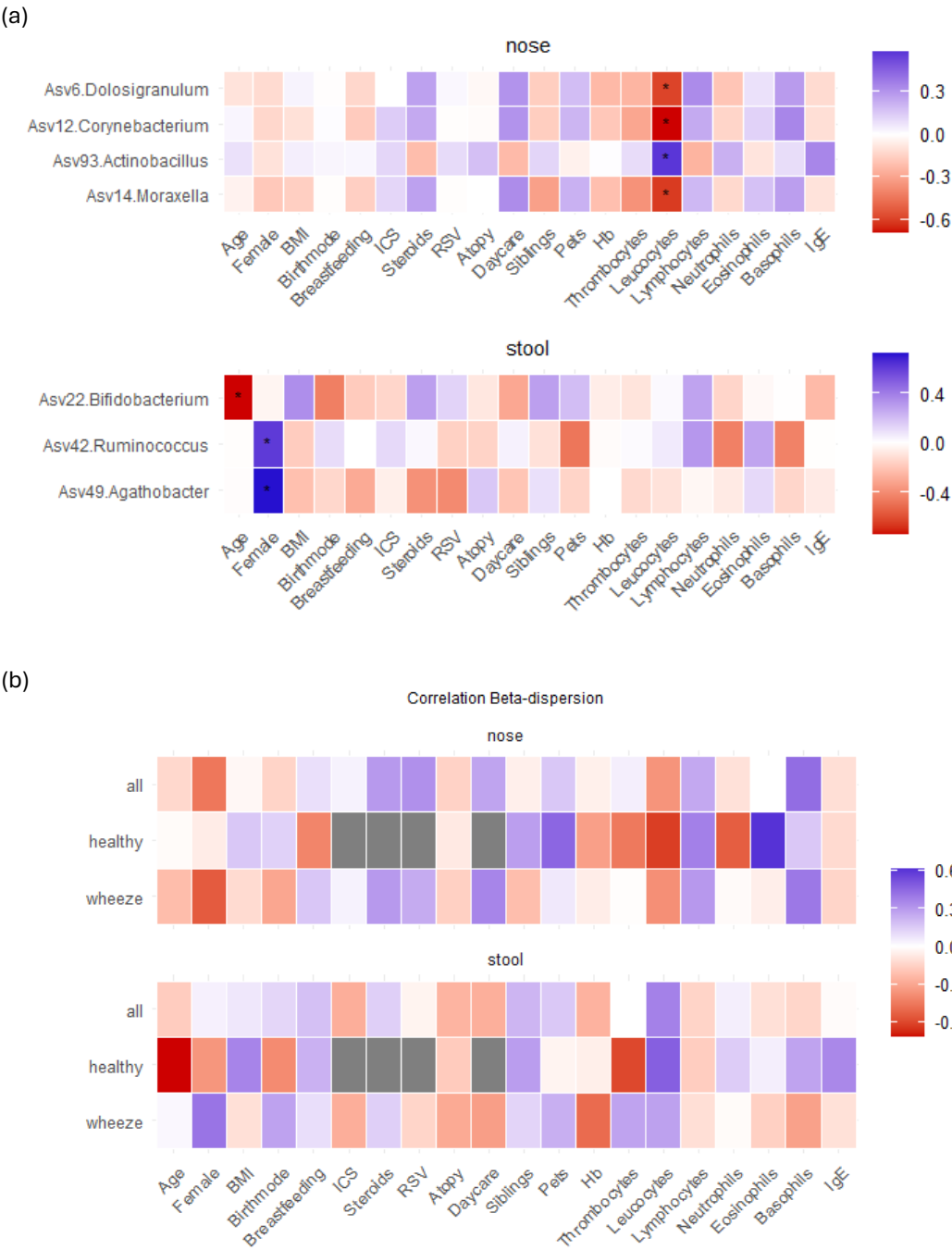


Figure S4. Associations between microbial community and demographic, lifestyle and immunological variables in nose and stool.

(a) Spearman correlations between individual ASV and selected demographic, lifestyle and immunological variables in nasal and stool samples. Asterisks indicate statistically significant correlations after Benjamini-Hochberg adjustment ($p < 0.05$), calculated using cor.test.

(b) Pearson correlations between beta-dispersion and selected demographic, lifestyle and immunological factors in nasal and stool samples. Asterisks indicate statistically significant correlations after Benjamini-Hochberg adjustment ($p < 0.05$), calculated using cor.test.

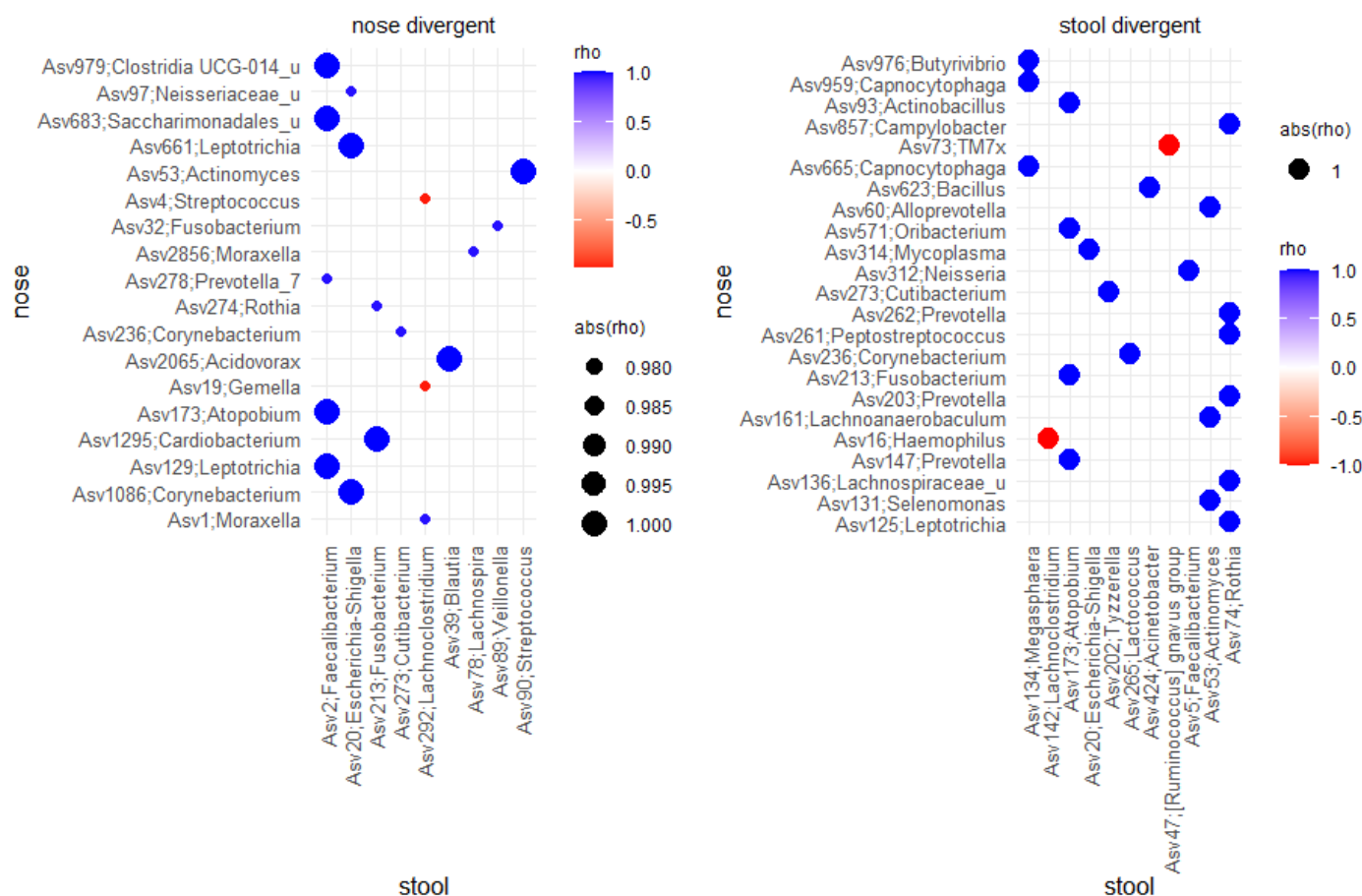


Figure S5. Pairwise cross-compartment correlations in community divergent subgroups. Spearman correlations were calculated between individual ASV in nasal (n=8) and stool (n=7) samples. Only statistically significant correlations after Benjamini–Hochberg adjustment ($p < 0.05$) are shown.

Supplementary Table S1. Core ASV ($\geq 80\%$ prevalence within group) in nasal and stool samples, shown as binary presence/absence (1/0). Nose: 9 shared, 4 healthy-specific, 9 wheezer-specific ASV; stool: 30 shared, 11 healthy-specific, 13 wheezer-specific ASV.

Taxa	nose		stool		Prev
	Healthy	Wheeze	Healthy	Wheeze	
Asv1;Moraxella	1	0	0	0	prev0.8
Asv10;Veillonella	0	1	1	1	prev0.8
Asv101;Clostridium sensu stricto 1	0	0	1	0	prev0.8
Asv105;Monoglobus	0	0	1	1	prev0.8
Asv106;Clostridium sensu stricto 1	0	0	1	1	prev0.8
Asv110;Neisseria	0	1	0	0	prev0.8
Asv111;Anaerostipes	0	0	1	1	prev0.8
Asv114;Porphyromonas	1	1	0	0	prev0.8
Asv116;Veillonella	0	0	1	1	prev0.8
Asv13;Roseburia	0	0	1	0	prev0.8
Asv132;Bifidobacterium	0	0	0	1	prev0.8
Asv142;Lachnoclostridium	0	0	1	1	prev0.8
Asv15;Bacteroides	0	0	1	0	prev0.8
Asv152;Oscillibacter	0	0	0	1	prev0.8
Asv158;Intestinibacter	0	0	1	1	prev0.8
Asv16;Haemophilus	0	1	1	1	prev0.8
Asv166;Clostridium sensu stricto 1	0	0	1	1	prev0.8
Asv167;Romboutsia	0	0	1	1	prev0.8
Asv17;Prevotella_7	0	1	0	0	prev0.8
Asv172;Anaerostipes	0	0	0	1	prev0.8
Asv18;Dialister	0	0	1	0	prev0.8
Asv180;Porphyromonas	0	1	0	0	prev0.8
Asv19;Gemella	1	1	0	0	prev0.8
Asv2;Faecalibacterium	0	0	1	1	prev0.8
Asv20;Escherichia-Shigella	0	0	1	1	prev0.8
Asv209;Actinomyces	0	1	0	0	prev0.8
Asv215;Erysipelatoclostridium	0	0	1	1	prev0.8
Asv22;Bifidobacterium	0	0	1	0	prev0.8
Asv23;Roseburia	0	0	1	1	prev0.8
Asv234;Ruminococcaceae_u	0	0	1	1	prev0.8
Asv24;Alistipes	0	0	0	1	prev0.8
Asv25;Streptococcus	1	1	1	1	prev0.8
Asv250;Hungatella	0	0	1	1	prev0.8
Asv252;Lautropia	1	0	0	0	prev0.8
Asv257;[Clostridium] innocuum group	0	0	1	1	prev0.8
Asv267;[Eubacterium] hallii group	0	0	0	1	prev0.8
Asv268;UBA1819	0	0	0	1	prev0.8
Asv27;Bifidobacterium	0	0	1	1	prev0.8
Asv274;Rothia	1	0	0	0	prev0.8
Asv28;Faecalibacterium	0	0	1	0	prev0.8
Asv287;Eggerthella	0	0	1	1	prev0.8
Asv29;Alloprevotella	1	1	0	0	prev0.8
Asv292;Lachnoclostridium	0	0	0	1	prev0.8
Asv3;Bacteroides	0	0	0	1	prev0.8
Asv311;Lachnospiraceae NK4A136 group	0	0	1	1	prev0.8
Asv33;Agathobacter	0	0	1	1	prev0.8
Asv37;Neisseria	0	1	0	0	prev0.8
Asv39;Blautia	0	0	1	1	prev0.8
Asv4;Streptococcus	1	1	1	0	prev0.8
Asv41;Bacteroides	0	0	1	0	prev0.8
Asv42;Ruminococcus	0	0	1	1	prev0.8
Asv45;Ruminococcus	0	0	0	1	prev0.8
Asv47;[Ruminococcus] gnavus group	0	0	1	1	prev0.8
Asv49;Agathobacter	0	0	1	1	prev0.8
Asv5;Faecalibacterium	0	0	1	1	prev0.8
Asv506;Anaerotruncus	0	0	1	0	prev0.8
Asv51;Granulicatella	1	1	0	0	prev0.8
Asv53;Actinomyces	1	0	0	0	prev0.8
Asv61;Granulicatella	1	1	0	0	prev0.8
Asv62;Actinomyces	0	1	0	0	prev0.8
Asv64;Streptococcus	1	1	1	0	prev0.8
Asv657;Terrisporobacter	0	0	1	0	prev0.8
Asv67;Blautia	0	0	0	1	prev0.8
Asv71;Blautia	0	0	0	1	prev0.8
Asv77;Flavonifractor	0	0	1	1	prev0.8
Asv8;Haemophilus	1	1	0	0	prev0.8
Asv85;Veillonella	0	1	0	0	prev0.8
Asv86;Lachnospiraceae NK4A136 group	0	0	1	1	prev0.8
Asv88;Fusicatenibacter	0	0	1	1	prev0.8
Asv89;Veillonella	0	0	0	1	prev0.8
Asv95;Lachnospiraceae UCG-004	0	0	0	1	prev0.8

Supplementary Table S2. Summary of baseline characteristics of wheezing children further stratified by microbial community composition distances.

Characteristic	Overall N = 17	nose			stool		
		healthy-like N = 9	divergent N = 8	p value ¹	healthy-like N = 10	divergent N = 7	p value ¹
Age (year)	3.00 (1.90-3.40)	3.00 (2.60-3.30)	2.55 (1.50-3.70)	0.772	3.00 (1.90-3.40)	3.20 (1.50-3.70)	0.807
Female	4 (24%)	4 (44%)	0 (0%)	0.082	3 (30%)	1 (14%)	0.603
BMI (kg/m ³)	16.98 (15.50-18.20)	17.40 (15.50-18.20)	16.67 (15.76-18.08)	0.501	16.84 (15.30-18.20)	17.90 (15.50-18.26)	0.696
Any sibling	12 (71%)	7 (78%)	5 (63%)	0.620	7 (70%)	5 (71%)	>0.999
Daycare	15 (88%)	7 (78%)	8 (100%)	0.471	9 (90%)	6 (86%)	>0.999
Pet in household (cat/dog)	7 (41%)	4 (44%)	3 (38%)	>0.999	2 (20%)	5 (71%)	0.058
Asthma or recurrent spastic/obstructive bronchitis - DD	2 (13%)	2 (22%)	0 (0%)	0.475	0 (0%)	2 (33%)	0.125
Missing	1	0	1		0	1	
Hayfever - DD	1 (5.9%)	1 (11%)	0 (0%)	>0.999	1 (10%)	0 (0%)	>0.999
Eczema - DD	2 (13%)	1 (11%)	1 (14%)	>0.999	1 (11%)	1 (14%)	>0.999
Missing	1	0	1		1	0	
Food allergy - DD	0 (0%)	0 (0%)	0 (0%)		0 (0%)	0 (0%)	-- ²
Birthmode				0.091			0.449
Vaginal delivery	8 (47%)	5 (56%)	3 (38%)		6 (60%)	2 (29%)	
Primary cesarean section	6 (35%)	1 (11%)	5 (63%)		2 (20%)	4 (57%)	
Secondary cesarean section	1 (5.9%)	1 (11%)	0 (0%)		1 (10%)	0 (0%)	
Operative vaginal delivery	2 (12%)	2 (22%)	0 (0%)		1 (10%)	1 (14%)	
Any exclusive breastfeeding ever	14 (82%)	7 (78%)	7 (88%)	>0.999	8 (80%)	6 (86%)	>0.999
Any hypoallergenic formula ever	8 (47%)	5 (56%)	3 (38%)	0.637	5 (50%)	3 (43%)	>0.999
Prior RSV infection ever	7 (47%)	3 (38%)	4 (57%)	0.619	3 (38%)	4 (57%)	0.619
Missing	2	1	1		2	0	
Antibiotics given for recurrent cough ever	1 (33%)	1 (50%)	0 (0%)	>0.999	1 (33%)	NA	-- ²
Missing	14	7	7		7	7	
Systemic steroids ever	9 (53%)	4 (44%)	5 (63%)	0.637	4 (40%)	5 (71%)	0.335
ICS ever	13 (76%)	6 (67%)	7 (88%)	0.576	8 (80%)	5 (71%)	>0.999
Hb (g/dl)	11.90 (11.45-12.25)	11.90 (11.50-12.25)	11.75 (11.45-12.40)	0.958	11.90 (11.50-12.30)	11.90 (11.30-12.20)	0.558
Missing	1	1	0		1	0	
Thrombocytes (10 ³ /μl)	403 (372-443)	403 (374-430)	418 (318-449)	0.958	397 (370-443)	408 (374-455)	0.491
Missing	1	1	0		1	0	
Leucocytes (10 ³ /μl)	9.27 (8.24-11.65)	10.80 (9.27-11.95)	8.52 (7.94-9.91)	0.074	9.40 (8.25-10.78)	9.03 (8.23-12.20)	0.874
Missing	1	1	0		1	0	
Neutrophils (%)	42 (36-55)	48 (36-55)	39 (33-51)	0.563	40 (37-47)	55 (21-60)	0.427
Missing	1	1	0		1	0	
Lymphocytes (%)	40 (36-47)	38 (33-42)	47 (38-59)	0.115	41 (39-48)	38 (30-45)	0.427
Missing	1	1	0		1	0	
Eosinophils (%)	1.50 (0.78-2.40)	1.14 (0.47-2.20)	2.00 (1.00-3.90)	0.295	2.00 (1.00-2.40)	0.90 (0.50-2.00)	0.375
Missing	2	1	1		1	1	
Basophils (%)	0.55 (0.08-1.00)	0.25 (0.04-0.75)	0.85 (0.50-1.00)	0.194	0.70 (0.08-1.00)	0.50 (0.10-0.60)	0.841
Missing	3	1	2		1	2	
IgE (IU/ml)	19 (9-82)	19 (9-93)	26 (9-43)	0.724	19 (9-43)	37 (6-82)	0.724
Missing	2	0	2		1	1	
Atopy (spec. IgE to any allergen ≥0.7 kU/L)	5 (29%)	3 (33%)	2 (25%)	>0.999	4 (40%)	1 (14%)	0.338

Values are summarized as median and interquartile range (Q1-Q3) in brackets, or as total numbers per condition level and percentage in brackets.

DD = parental report of doctor's diagnosis

¹Kruskal-Wallis rank sum test; Fisher's exact test

²p value not calculated due to insufficient variability (single-level data) and/or missing values

Supplementary Table S3. Alpha diversity indices and betadispersion of wheezing children further stratified by microbial community composition distances

	nose			stool		
	healthy-like	divergent	p	healthy-like	divergent	p
Observed ASV	149.00	94.75	<0.001	180.40	190.57	0.835
Shannon index	3.46	1.23	<0.001	3.32	3.44	0.321
Evenness	0.69	0.28	<0.001	0.64	0.66	0.236
Betadispersion	0.17	0.30	<0.001	0.11	0.15	0.001