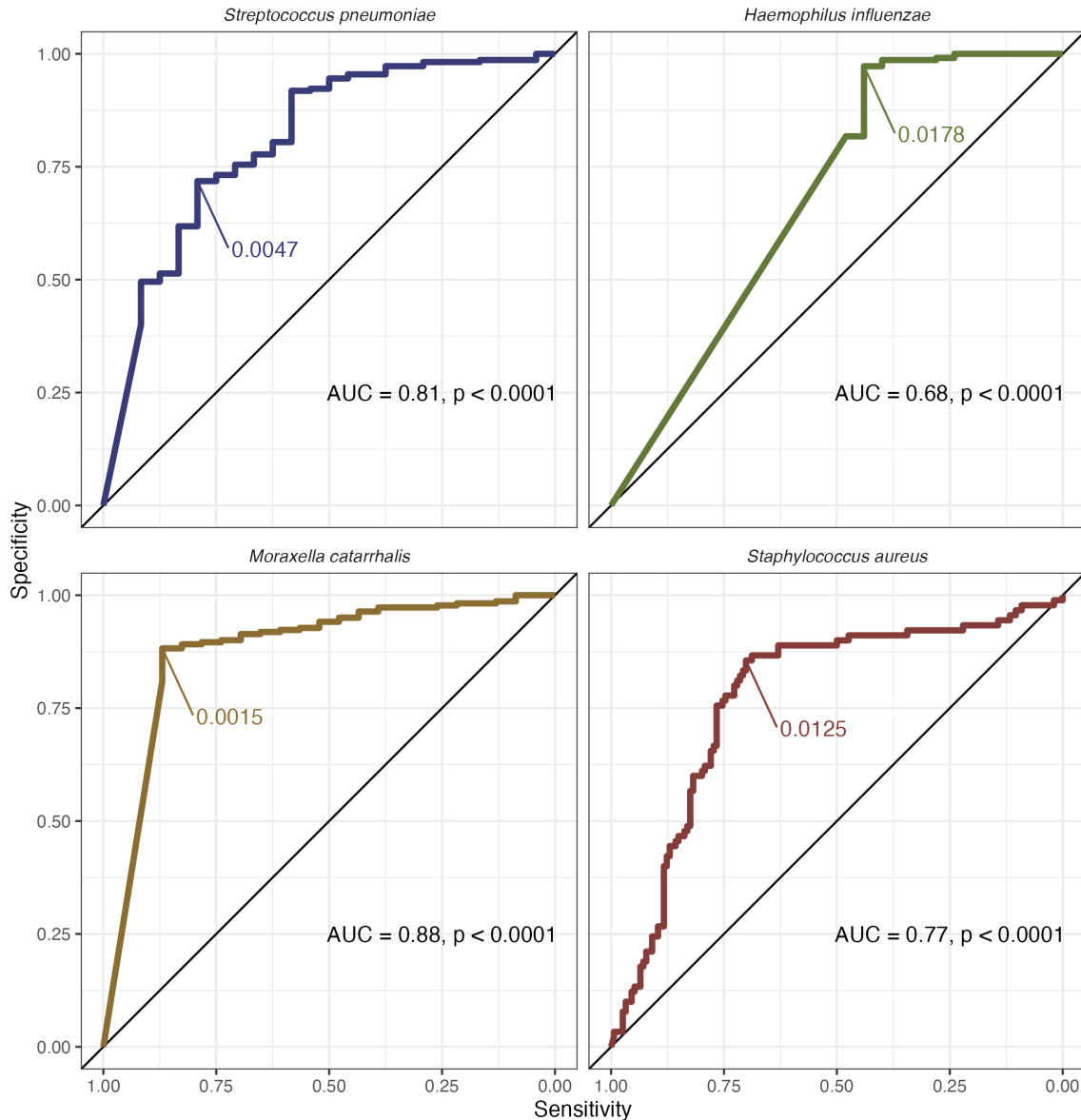


SUPPLEMENTARY INFORMATION

For the article Thorsen J et al, “The airway microbiota of neonates colonized with asthma-associated pathogenic bacteria”, *Nature Communications* 2023.

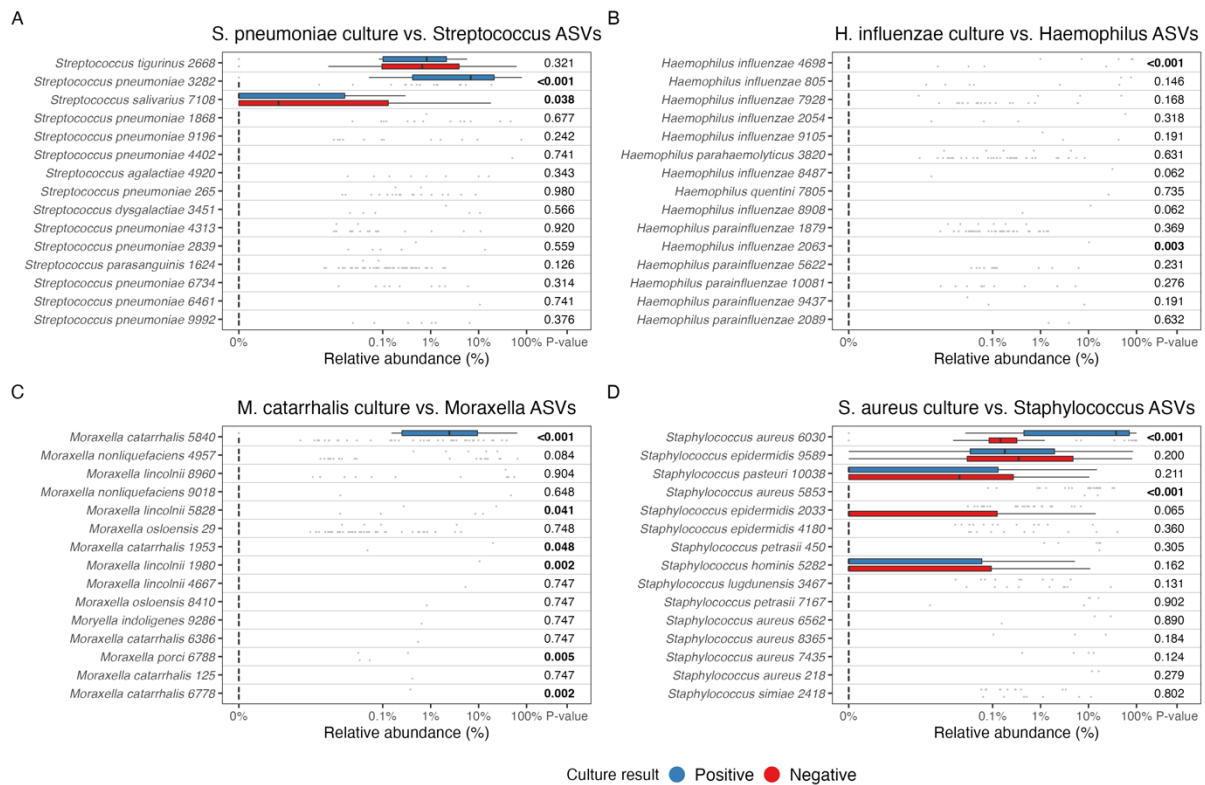
Table of Contents

Supplementary Fig. 1	2
Supplementary Fig. 2	3
Supplementary Fig. 3	4
Supplementary Fig. 4	5
Supplementary Fig. 5	6
Supplementary Fig. 6	8
Supplementary Table 1	9
Supplementary Table 2	10



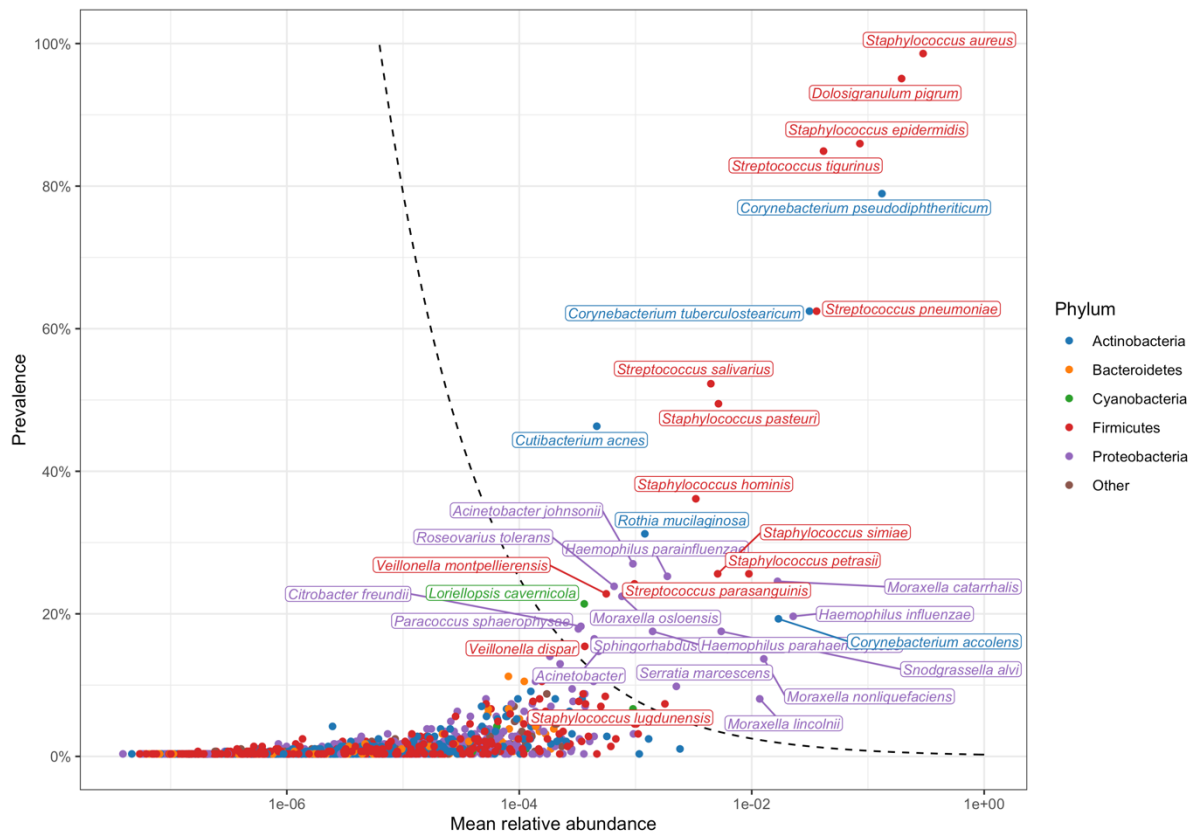
Supplementary Fig. 1

Receiver Operating Characteristic (ROC) curves for comparing relative abundances from nasopharynx swabs with culture results from hypopharynx aspirates, for the four cultured bacterial species. Area under the curve (AUC) values and p-values (Wilcoxon tests) are shown for each comparison. Cutoff values marked for each curve (relative abundance) corresponding to the maximum value of Youden's J index. N = 285 children.



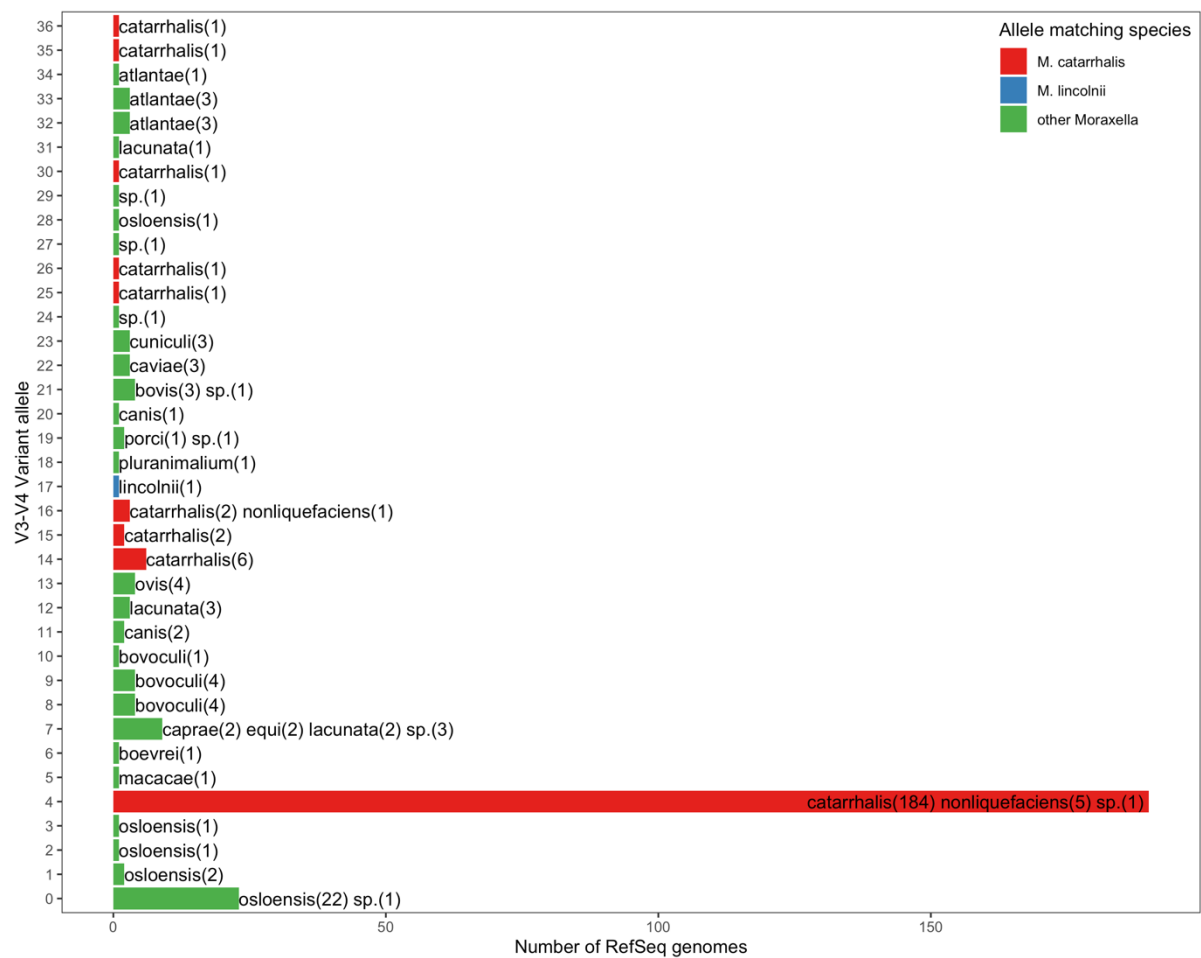
Supplementary Fig. 2

Comparison of the culture results (positive/negative) for each of the four species identified from the hypopharynx aspirates (Panel A – *Streptococcus pneumoniae*; B – *Haemophilus influenzae*; C – *Moraxella catarrhalis*; D – *Staphylococcus aureus*) and the relative abundances from 16S sequencing of nasopharyngeal swabs of each of the 15 most abundant Amplicon Sequence Variants (ASVs) within the same genus of each. The results show high specificity and agreement at species level. P-values are from Wilcoxon tests. Box plots show median (middle line), interquartile range (IQR, box), 1.5xIQR range (whiskers) and outliers (dots). A pseudocount of 1e-7 was added for visualization purposes, corresponding to “0%” on the x-axis. N=285.



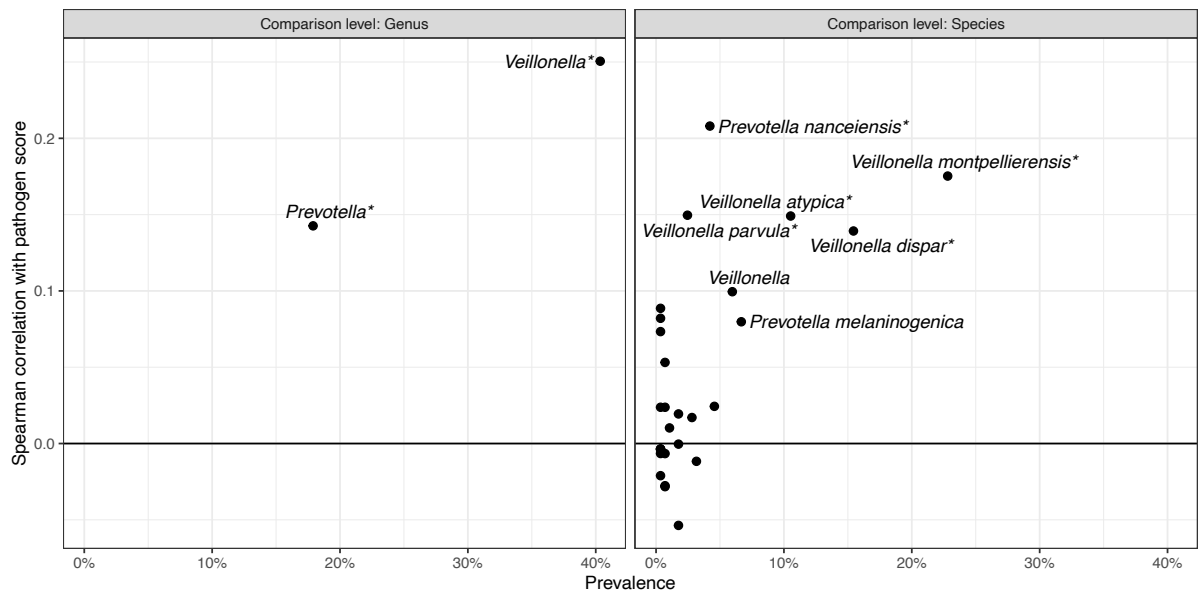
Supplementary Fig. 3

Comparison of mean relative abundance and prevalence for all species identified in the 16S sequencing data from nasopharyngeal swabs (n=285). The 35 most common species above the dashed line ($\text{Prevalence} = 25\% * (1e-4 / \text{Mean Relative Abundance})^{0.5}$) are annotated and were selected for downstream analysis.



Supplementary Fig. 4

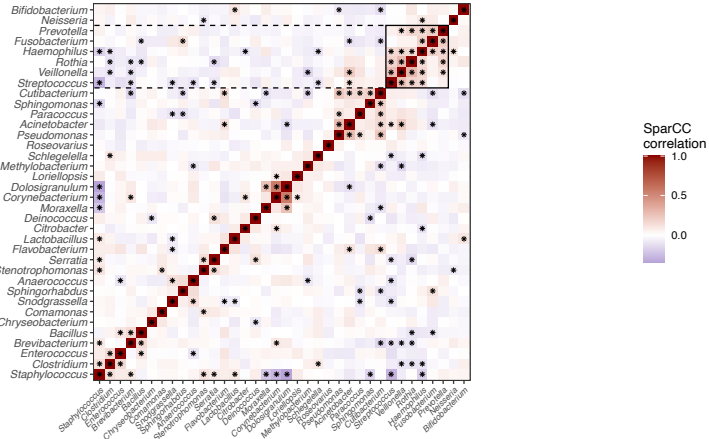
Comparison of simulated V3-V4 amplicons using the RibDif tool, based on RefSeq *Moraxella* genomes. While *M. catarrhalis* could be mistaken for *M. nonliquefaciens*, since their alleles are similar and cluster together, no other *Moraxella* species cluster with *M. lincolnii*.



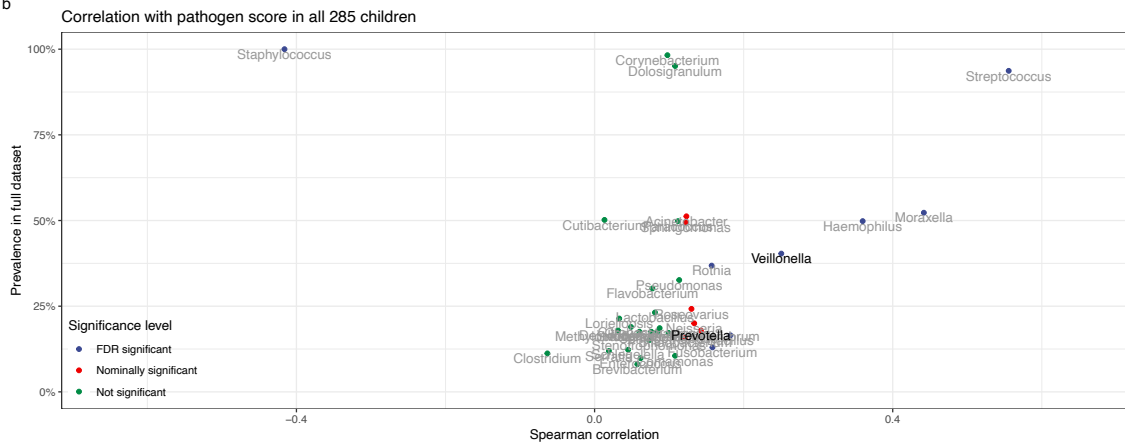
Supplementary Fig. 5

Spearman correlations between the pathogen score (derived from the a priori summed relative abundance of *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*) and the genera *Veillonella* and *Prevotella*. Significant correlations ($p < 0.05$) are marked with *. Both *Veillonella*, *Prevotella* and several of their species constituents are significantly associated with the pathogen score. N=285.

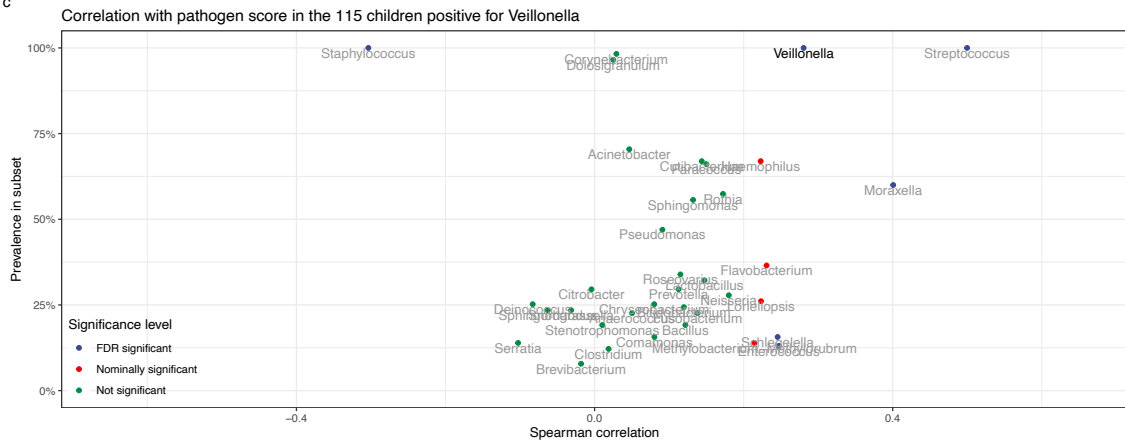
a



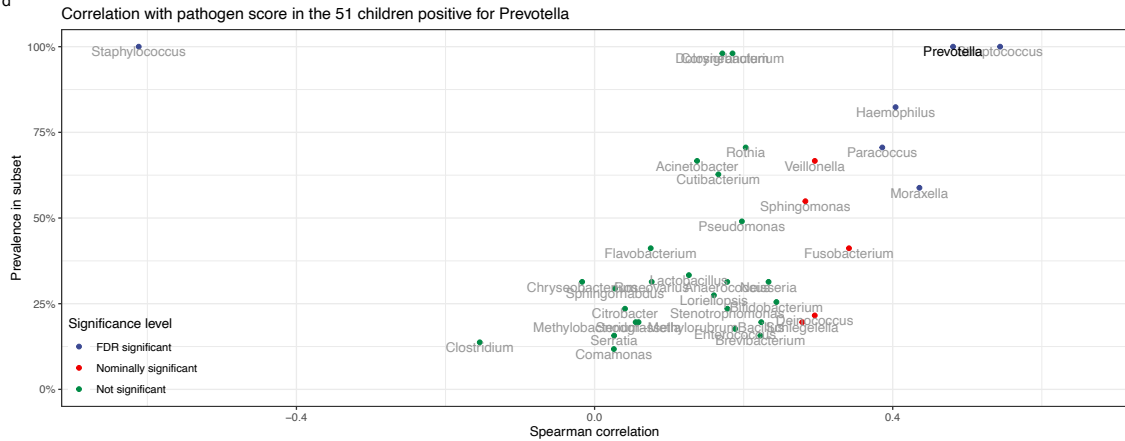
b



c



d



Supplementary Fig. 6

a) Correlation heatmaps between most common genera, arranged by hierarchical clustering, using the SparCC compositional correlation metric. Similar to fig 5, we observe that *Veillonella* and *Prevotella* form a cluster with two of the three pathogen genera, *Streptococcus* and *Haemophilus*, which also includes a few other taxa. Highlighting was performed manually. Stars indicate significant correlations (bootstrap p-values < 0.05). b-d) Correlations between the pathogen score and the most common genera in b) all children, c) children positive for *Veillonella* and d) children positive for *Prevotella*. In panel b), we observe that *Veillonella* has the fourth highest correlation, after the genera of the three pathogenic species on which the score was built. *Prevotella* is significant, but not after FDR correction. However, due to low prevalence and excess zeroes, the power to detect this correlation is low, which can be seen as a V-shape limit in the distribution of panel b. When subsetting the data to only those children positive for each genus, *Veillonella* remains strongly correlated and advances slightly to become the third most correlated, while *Prevotella* increases immensely to become the second most correlated, almost at the level of *Streptococcus*. N=285.

Predictor	N	Beta	95% CI ¹	p-value
Sex	285			0.2
Female		—	—	
Male		-0.16	-0.39, 0.07	
Ethnicity	285			0.5
Northern European		—	—	
Other		0.25	-0.42, 0.92	
Delivery mode	285			0.009
Vaginal		—	—	
Cesarean Section		-0.37	-0.65, -0.09	
Gestational age (weeks)	285	0.01	-0.06, 0.09	0.7
Birth weight (kg)	285	0.09	-0.13, 0.31	0.4
Birth length (cm)	285	0.02	-0.03, 0.07	0.3
Living environment	285			0.6
Rural		—	—	
Urban		-0.06	-0.29, 0.18	
Older siblings at home	267			<0.001
No siblings		—	—	
Siblings		0.63	0.39, 0.86	
Furred pet at home	271			0.6
No		—	—	
Yes		0.07	-0.19, 0.33	
Maternal smoking at 1 month	249			0.2
No		—	—	
Yes		0.21	-0.13, 0.55	
Breastfeeding at 1 month	257			0.091
No		—	—	
Yes		0.27	-0.04, 0.59	
Social circumstances PCA score	265	-0.02	-0.15, 0.10	0.7
Antibiotics during pregnancy	285			0.2
No antibiotics		—	—	
Antibiotics		0.16	-0.09, 0.42	

¹CI = Confidence Interval

Supplementary Table 1

Key environmental factors and their association with the pathogen score, which was defined as the log-transformed and z-scaled combined relative abundances of *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*. P-values are two-sided and not adjusted for multiple comparisons.

Genus	Species	Prevalence	N	Mean relative abundance	Genus HR for asthma COPSAC2010
<i>Veillonella</i>	<i>Veillonella montpellierensis</i>	22.81%	65	0.00056	1.45
<i>Veillonella</i>	<i>Veillonella dispar</i>	15.44%	44	0.000366	1.45
<i>Veillonella</i>	<i>Veillonella atypica</i>	10.53%	30	0.000157	1.45
<i>Veillonella</i>	<i>Veillonella sp.</i>	5.96%	17	0.000195	1.45
<i>Veillonella</i>	<i>Veillonella tobetsuensis</i>	4.56%	13	3.77E-05	1.45
<i>Prevotella</i>	<i>Prevotella melaninogenica</i>	6.67%	19	8.09E-05	1.32
<i>Prevotella</i>	<i>Prevotella nanceiensis</i>	4.21%	12	0.000139	1.32
<i>Prevotella</i>	<i>Prevotella veroralis</i>	3.16%	9	2.18E-05	1.32
<i>Prevotella</i>	<i>Prevotella jejuni</i>	1.75%	5	4.72E-06	1.32
<i>Prevotella</i>	<i>Prevotella histicola</i>	1.75%	5	2.81E-05	1.32
<i>Gemella</i>	<i>Gemella taiwanensis</i>	4.91%	14	6.56E-05	1.28
<i>Gemella</i>	<i>Gemella sanguinis</i>	2.46%	7	4.39E-05	1.28
<i>Gemella</i>	<i>Gemella parahaemolysans</i>	2.46%	7	4.14E-05	1.28
<i>Gemella</i>	<i>Gemella morbillorum</i>	2.11%	6	7.88E-06	1.28
<i>Streptococcus</i>	<i>Streptococcus tigurinus</i>	84.91%	242	0.041463	1.22
<i>Streptococcus</i>	<i>Streptococcus pneumoniae</i>	62.46%	178	0.036148	1.22
<i>Streptococcus</i>	<i>Streptococcus salivarius</i>	52.28%	149	0.00445	1.22
<i>Streptococcus</i>	<i>Streptococcus parasanguinis</i>	24.21%	69	0.000979	1.22
<i>Streptococcus</i>	<i>Streptococcus sanguinis</i>	5.96%	17	8.36E-05	1.22
<i>Lactobacillus</i>	<i>Lactobacillus curvatus</i>	7.37%	21	0.000371	1.21
<i>Lactobacillus</i>	<i>Lactobacillus gasseri</i>	6.32%	18	0.000168	1.21
<i>Lactobacillus</i>	<i>Lactobacillus iners</i>	3.16%	9	2.28E-05	1.21
<i>Lactobacillus</i>	<i>Lactobacillus sakei</i>	2.46%	7	9.83E-05	1.21
<i>Lactobacillus</i>	<i>Lactobacillus reuteri</i>	1.40%	4	3.89E-06	1.21
<i>Neisseria</i>	<i>Neisseria subflava</i>	7.37%	21	0.000129	1.18
<i>Neisseria</i>	<i>Neisseria cinerea</i>	7.02%	20	0.000187	1.18
<i>Neisseria</i>	<i>Neisseria polysaccharea</i>	2.46%	7	2.59E-05	1.18
<i>Neisseria</i>	<i>Neisseria sicca</i>	2.11%	6	3.95E-06	1.18
<i>Neisseria</i>	<i>Neisseria perflava</i>	2.11%	6	2.23E-05	1.18

Supplementary Table 2

Top five species in COPSAC₂₀₀₀ from each of the genera found to be associated with asthma in the COPSAC₂₀₁₀ cohort (Ref DOI: 10.1038/s41467-019-12989-7).